

SAYANDEEP PURKAYASTH

A HIERARCHICAL IMAGE PROCESSING
APPROACH TO ANALYSIS OF EARLY MEDIEVAL
MANUSCRIPT ART

A HIERARCHICAL IMAGE PROCESSING
APPROACH TO ANALYSIS OF EARLY
MEDIEVAL MANUSCRIPT ART

SAYANDEEP PURKAYASTH

A thesis submitted to Trinity College Dublin, the University of
Dublin in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

Department of History of Art and Architecture
School of Histories & Humanities

FACULTY OF ARTS, HUMANITIES AND SOCIAL SCIENCES



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

SEPTEMBER 2015

ABSTRACT

The discipline of art history is one in which digital technologies have traditionally played a minor role. Recently with advances in computing methods, software and the capabilities of hardware, art analysis has begun to see the benefits of automated image understanding and pattern recognition techniques. The specific domain of historical manuscript analysis aims to answer questions regarding the visual design, structure, execution, comparison, meaning and context of these historical works of art. Compared to other forms of art, such as paintings, architecture, and abstract art, investigations into manuscripts deal with the construction of design at the much more primitive level of textual and illustration content and styles. Such a bottom-up approach to analysis is encouraging for digital image processing studies into art. The range of questions that are of interest to art historians extends from the elicitation of design and layout, the accuracy of execution and the restoration of degraded art works to the understanding of the aesthetic effect of these manuscripts. Recent work in the field of digital art history has seen application of a various pattern recognition and image statistical techniques to specific art historical analyses. Most of these applications are based on techniques that are irreconcilable with each other methodologically. The subject of this thesis is the problem of developing a unified framework for such studies that allows the historian to answer a wide range of art historical problems. The rudiments of such a framework are developed herein and its applicability is demonstrated via approaching some questions regarding the Book of Kells (as an exemplar), specifically the analysis of calligraphy, interlace, detection of various design elements and degradation removal. The framework improves on biologically inspired hierarchical models of vision and makes it feasible to develop art historical applications in a consistent fashion. This consistency also affords the possibility of approaching problems of a highly abstract nature, viz. aesthetics, using information extracted from other (less abstract) applications, such as object detection. As a particular application, the identification of areas of manuscripts that might draw more attention than others is demonstrated.

*Art thou not, fatal vision, sensible
To feeling as to sight? Or art thou but
A dagger of the mind, a false creation
Proceeding from the heat-oppressed brain?*

— William Shakespeare, *Macbeth* Act II. Sc. 1.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the guidance and support of my supervisors Dr. Laura Cleaver, Dr. John Dingliana and Prof. Roger Stalley. I sincerely thank them for their guidance, continuous encouragement, constructive criticism, generous help, endless patience and enormous freedom given me during my study. My colleagues in the Graphics Visualisation and Vision Group and the Trinity Irish Art Research Centre over the past four years played a vital role in this work. I am grateful to them all for their support, advice, proof reading and general good company. My thanks also go out to the University College Dublin - Trinity College Dublin Innovation Alliance for sponsoring this research.

Sayandeep Purkayasth

Dedicated to my loving grandmother
Arundhati Manna

DECLARATION

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Dublin, September 2015

Sayandeep Purkayasth

PERMISSION TO LEND OR COPY

In submitting this thesis to the University of Dublin, I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Trinity College Library conditions of use and acknowledgement.

Dublin, September 2015

Sayandeep Purkayasth

CONTENTS

I	INTRODUCTION	3
1	INTRODUCTION	5
1.1	Art history	5
1.2	Manuscript analysis	6
1.3	Scope and contributions	7
1.4	Outline of thesis	9
II	LITERATURE REVIEW	11
2	ART HISTORY	13
2.1	Introduction	13
2.2	Digital art history	16
2.3	Manuscripts	17
2.4	Questions about the Book of Kells	18
2.4.1	Design	18
2.4.2	Execution	19
2.4.3	Other questions	20
3	IMAGE PROCESSING	23
3.1	Image processing methods	25
3.2	Hierarchical architectures	27
3.2.1	NeoCognitron	30
3.2.2	Occlusion reconstruction network	31
3.2.3	HMAX model for object recognition	32
3.2.4	Convolutional Networks	34
3.3	Manuscripts and image processing	35
3.4	Discussion	37
III	APPLICATIONS	41
4	CHARACTER ANALYSIS	43
4.1	Introduction	45
4.1.1	Questions	48
4.1.2	Problem definition and scope	49
4.1.3	Possible solutions	49
4.2	Hand classification	50
4.2.1	Character recognition	50
4.2.2	Object recognition model	51
4.2.3	Feature extraction and hand classification	54
4.3	Conclusion and future work	55
5	INTERLACE	57
5.1	Introduction	59
5.1.1	Questions	61
5.1.2	Nomenclature	64
5.1.3	Problem definition and scope	64

5.1.4	Possible solutions	65
5.2	Localisation of interlace crossings	66
5.3	Metrics for Interlace analysis	66
5.3.1	Grid fitting	66
5.3.2	Mechanical or optical aids	67
5.3.3	Scale	68
5.3.4	Similar knotwork detection	68
5.4	Results	68
5.4.1	Extraction of crossings	68
5.4.2	Statistical measures	70
5.5	Conclusion and future work	73
6	LEARNED DEGRADATION REMOVAL	81
6.1	Introduction	83
6.1.1	Questions	85
6.1.2	Problem definition and scope	86
6.1.3	Possible solutions	86
6.2	Degradation removal for interlace	87
6.2.1	Implementation	88
6.2.2	Training and testing	89
6.2.3	Results and discussion	89
6.3	Conclusion and future work	90
7	MISCELLANEOUS APPLICATIONS	93
7.1	Accuracy of spiral designs	95
7.1.1	Rotational symmetry in images	95
7.1.2	Symmetry axes extraction network	96
7.1.3	Results	98
7.2	Localisation of faces	99
7.3	Layout elicitation	100
7.3.1	Detecting lines and paragraphs	102
7.4	Conclusion and future work	102
8	AESTHETICS	105
8.1	Introduction	107
8.1.1	Questions	108
8.1.2	Problem definition, scope and possible solutions	108
8.2	Eye tracking experimental design	109
8.3	Gaze prediction	110
8.3.1	Extended hierarchy for gaze prediction	113
8.3.2	Results	114
8.4	Conclusion and future work	114
9	CONCLUSION AND FUTURE WORK	119
9.1	Conclusion	119
9.2	Future work	122
A	APPENDICES	127
A.1	Object recognition and localisation model	127
A.2	Degradation reconstruction model	134
A.3	Hand classification metrics	136

A.4	Spiral metrics	137
A.4.1	Rotational symmetry extraction model	137
A.5	Gaze prediction by boosted learning	137
A.6	Implementation of models	138
A.6.1	Graphics Processing Units: memory and processor	138
A.6.2	Frameworks	139
A.7	The Book of Kells images	139
A.8	Design elements in Insular manuscripts	139
BIBLIOGRAPHY		145

LIST OF FIGURES

Figure 1	A bottom-up image processing scheme and a specific application	26
Figure 2	Integration of bottom-up, lateral, and top-down processing in a hierarchical architecture. Images are represented at different levels of abstraction. As the spatial resolution decreases, feature diversity and invariance to transformations increase.	28
Figure 3	Neural abstraction pyramid	29
Figure 4	The Neocognitron model. Digit features of increasing complexity are extracted in a hierarchical feed-forward neural network.	30
Figure 5	Functional primitives of a theory of object recognition [118]. The network consists of alternating S-layers and C-layers that extract features of increasing complexity, size, and invariance.	33
Figure 6	Convolutional neural network LeNet-5 [86]	34
Figure 7	Some letters used for training and testing character recognition	54
Figure 8	Constructing knotwork in stages by George Bain [4]	62
Figure 9	Some interlace designs	62
Figure 10	Graph based construction of interlace by Kaplan and Cohen [74]	63
Figure 11	Parallelogram model for grid	67
Figure 12	Candidates for similarity computation	68
Figure 13	Results for mean squared errors over dataset	71
Figure 14	Registration	72
Figure 15	Scale of interlace design	73
Figure 16	Extraction of crossings. Extracted points are marked in blue.	74
Figure 17	Extraction of crossings (contd.) Extracted points are marked in blue.	75
Figure 18	Extraction of crossings. Crossing points are marked in red.	75
Figure 19	Extraction of crossings (contd.). Crossing points are marked in red.	76
Figure 20	Some grid fitting results. Corrected extracted grid points are marked in blue, and fitted grid is depicted red.	77
Figure 21	Proposed degradation removal network	88

Figure 22	Degradation removal results	91
Figure 23	Symmetry axis detection model by Fukushima and Kikuchi [49]	96
Figure 24	Rotational symmetry axes model	98
Figure 25	Feature extraction in the elements with circular symmetry. The image domain features selected by the user are marked in red. The lines in blue represent relations between recurring features or between the output and feature.	99
Figure 26	Faces detected in MS 58 folio 32v	101
Figure 27	Layout analysis of textual content	103
Figure 28	Raw gaze data for folio 18 of the Lindisfarne Gospels for one observer	111
Figure 29	Raw gaze data for folio 114r of <i>the Book</i> (Trinity College Library MS 58) for one observer	112
Figure 30	Gaze prediction for Lindisfarne Gospels (folio 18)	115
Figure 31	Gaze prediction for 114r of <i>the Book</i> (Trinity College Library MS 58)	116
Figure 32	Modular comparison of (flattened) models for applications examined in this study. The layers and connections marked in red collectively represent the core object recognition and localisation model. Extensions and other operations developed for the applications covered in the previous chapters are marked in black.	121
Figure 33	Overview of the base model (from Mutch and Lowe [98]). Each layer has units covering three spatial dimensions (X/Y/scale), and at each 3D location, an additional dimension of feature type. The image layer has only one type (pixels), layers S1 and C1 have 4 types, and the upper layers have d (many) types per location. Each layer is computed from the previous via convolution with template matching or max pooling filters. Image size can vary and is shown for illustration.	131
Figure 34	Dense vs. sparse S2 features (from Mutch and Lowe [98]). Dense S2 features in the base model are sensitive to all orientations of C1 units at each position. Sparse features are sensitive only to a particular orientation at each position. A 4×4 S2 feature for a 4-orientation model is shown here. Stronger C1 unit responses are shown as darker.	132

Figure 35	NeoCognitron: an occlusion reconstruction network model [48]	135
Figure 36	Design elements used of Insular art and sculpture (collected from [70] and other sources)	141

LIST OF TABLES

Table 1	Similarity scores for interlace designs	73
Table 2	Comparison of layer sizes (in position)	135

Part I

INTRODUCTION

Introduction to the fields of art history, manuscript analysis, digital art history, and the scope and outline of the thesis.

INTRODUCTION

1.1 ART HISTORY

The academic discipline of Art History is the specialised analysis of the historical development and stylistic contexts of the various forms of human artistic expression. It aims to describe art and devotes specific attention to the classification of artistic traditions, periodisation, comparison and analysis of distinctive and influential characteristics of works of art, often in interaction with other faculties in the arts and sciences.

While the history of art deals with the principal art forms of architecture, dance, sculpture, music, painting, poetry, film, photography and graphic arts, the core focus of art history is the visual arts. Studies in the discipline can be classified into several broad areas based on the time period of focus, starting with Antiquity and the Middle ages and so on till current time. They may also be categorised based on geographical, anthropological, cultural or ethnic focus.

With the coming of the era of digitisation and information technology, the various disciplines in the arts have also entered an exciting age. The information technological boom has allowed artists and historians to share digital representations of artworks. Archival of artworks has taken huge strides with large scale digitisation efforts in major museums and academic institutions. Image archival for paintings, photographs and manuscripts, that traditionally used analogue media such as film negatives, has transitioned into the age of digital image archival with the commercial availability of very high quality digital image capture systems. With recent advances in imaging sciences, such historical artefacts are not only available at increasing quality, but also at diminishing costs. Simultaneously, the latest developments in statistical and image processing technology have opened up a new niche for automated analysis of such digitised artworks. The end goal of such automated techniques is to mirror the way humans understand images and structures, and allow for a data-driven foundation for any scientific art analysis.

In the next section, we briefly introduce the area of manuscript analysis, which is the focus of this thesis, and subsequently detail the scope of this study.

1.2 MANUSCRIPT ANALYSIS

A manuscript is any document written by hand or manually type-written, as opposed to being printed in an automated way, i.e. mechanically or digitally. Before the advent of the printing press, all documents and books produced were manuscripts. Manuscripts may contain writing, mathematical calculations, maps, explanatory figures or illustrations, and may be in the form of a book or scrolls. The tradition of illumination in manuscripts is one of considerable historical and cultural significance. Illuminated manuscripts are characterised by the decoration of text with elaborate pictures, borders, embellished initial letters and full page illustrations. This tradition was prevalent circa 400-1200 AD, although there are instances of illumination well into the sixteenth century. It represents a well defined art: a combination of the written word and graphic art. The major lines of inquiry into manuscripts are — understanding their historical and cultural background, painting technique, pigments used and physical properties of the parchment, characterising the calligraphy, palaeography, drawing correlations with other forms of art, the structural and mathematical analysis of elements designs, the examination of layout, page design, textual structure, exegesis and aesthetics and estimation the time period of creation, of manuscripts. In comparison, the research on illuminated manuscripts also includes the characterisation of the style of illumination used and determining the procedures and practices followed during their creation. The topic of stylistic investigations into illuminated manuscripts is an excellent area for the application of digital techniques to art history, as will be discussed in section 3.4. Over the course of this thesis, we explore the area of manuscript analysis using the Book of Kells as an exemplar and follow some lines of inquiry that would benefit from the field of digital image processing.

The Book of Kells, sometimes referred to as the Book of Columba, is an illuminated manuscript written in Latin, and is considered by many as the pinnacle of Insular illumination, as well as an extraordinary product of western calligraphy. The illustrations and ornamentation of the Book of Kells surpass that of other Insular Gospel books in extravagance and complexity. The decoration combines traditional Christian iconography with the ornate motifs typical of Insular art. Figures of man, animals and mythical beasts, together with interlace and key patterns, are found in abundance in the manuscript's pages. Many of these minor decorative elements are imbued with Christian symbolism and so further emphasise the themes of the major illustrations. Manuscripts such as *the Book* typically contain a number of stylistic elements¹, including text, illuminated text, knotwork, spirals, key patterns, plaitwork, etc. These elements are usually skilfully com-

¹ The major design elements of the Insular tradition are illustrated in appendix A.8.

bined by the artists with the aim to accompany the actual text. Visual inspection of *the Book* suggests that the scribes and illustrators prioritised visual appeal of the design over practicality and readability of its contents. The variation in calligraphy, the accuracy and complexity of design, usage of particular colours, mechanical tools and the ultimate effect of the design on its enduring visual appeal are topics of active art historical debate. Drawing on the basic themes of manuscript art analysis mentioned earlier, most prominent art historical questions pertaining to illuminated manuscripts such as the Book of Kells (referred to as *the Book* here onward) that would benefit from the methods in the fields of computer vision, pattern recognition and image processing are detailed in section 2.4. Most of these questions have a fundamental basis in the deconstructing of the design of the Book using a structural approach to identification of design motifs, their layout, etc. With the aim of developing automated digital techniques for such art historical analysis of manuscripts, we present an overview of the emerging field of digital art history, the art historical questions that we consider for this study in the following section.

1.3 SCOPE AND CONTRIBUTIONS

There has been digital art, i. e. computer generated art since the inception of the computing industry. In the context of art history, several computer algorithms have been developed to generate some of the key design elements referred to earlier. Structural analysis of the content of manuscripts, such as *the Book*, has been attempted manually to date. The aim of the study detailed in this thesis aims to provide objectively measurable information to the domain of art history. Image processing is a research field in the computer sciences and engineering which specialises in understanding digital images or videos by allowing information to be extracted from them. Typical tasks that may be expected of an image processing (also known as computer vision) system are recognition of objects, image restoration, motion analysis (from videos). In recent years, there has been a spurt of interest in the quantitative measurement of various properties of artwork, photographs and other visual information, even to the extent of estimating aesthetic appeal. Computational techniques have been used in providing valuable evidence in identifying some of the methods and devices used by a few master painters: David G. Stork [124] in collaboration with Johnson [126] and Duarte [125] discuss some such techniques in considerable detail, focusing on perspective analysis (that has some applications in understanding and de-warping anamorphic art), analysis of directions of shadows, which is used in identifying inconsistencies in lighting and forgery detection. Digital restoration of paintings has been a specialised branch of this emerging field with Berns [12] and Shahram et al. [120] discussing colour

restoration and an algorithm for removal of brush strokes from paintings, essentially reversing the process of composition. While there has been an immense interest in curation and digital collections (see [11]) for manuscripts, artwork and architectural models, the development of digital techniques to robustly analyse such data has yet to catch up with the proliferation of easily accessible manuscript images.

The advances in digital art history discussed above demonstrate the immense potential of computer vision and image processing to aid the art historian in a number of ways. While the goal of a digital analysis toolbox for historical manuscripts may be approached in various ways, the most obvious approach is to create a toolbox with a battery of techniques that can perform measurements and other operations on images toward answering art historical questions. The limited number of these questions mentioned in the previous sections exhibit a large diversity in nature and complexity from a computational stand point. Solving such a variety of problems using traditional computer vision methods would require building a separate solution for each question. As the main contribution of this thesis, in section 3.4, we argue that an alternative approach of developing a more unified framework for image analysis is not only theoretically possible, but also has many properties that would be beneficial to future research in this area. This approach is based on a family of image processing techniques, members of which are designed for specific tasks in computer vision, such as object recognition, localisation, image reconstruction, and other more complex classification operations. All these related image processing techniques share a hierarchical organisation of processing units that is modelled to emulate the neurons in the human visual cortex. These techniques can be adapted to perform different tasks, and this extensibility is developed in this thesis into a common framework via applications to some specific art historical questions:

CALLIGRAPHY Classification of hands in *the Book* using a hierarchical object recognition model to recognise characters and extraction of some stylistic features of characters

INTERLACE Extraction of the structure of interlace designs by localising interlace crossings, assessing the accuracy, similarity and scale of designs using novel metrics

DEGRADATION REMOVAL Improving the accuracy of interlace structure detection by developing a degradation removal model that learns to reconstruct damaged parts of designs

LAYOUT Extraction of layout for textual content of folios of *the Book* by extending the existing object (character) localisation model

SPIRALS Analysis of accuracy of rotational symmetry of spiral designs by localising repeated image features and by performing angular measurements

AESTHETICS Development and validation of an object-aware visual attention metric for manuscript images that depends on localisation of faces, text and interlace

1.4 OUTLINE OF THESIS

In chapters 2 and 3, we present the state of the art from two viewpoints: art history and computer vision (respectively). The existing techniques that might be used to solve each problem individually are discussed in the latter. Therein we discuss a family of hierarchical image processing models that we consolidate into a *unified* approach to answering the art historical questions. It will be argued that this approach is more favourable for the quickly evolving field of digital art history. Specific applications of the framework that aim at answering some of the art historical questions outlined earlier: chapters 4, 5 and 6 focus on the problems of hand classification, interlace analysis and degradation removal respectively. Chapter 7 covers layout elicitation and some introspection into spiral designs. Finally, in chapter 8, we predict attention behaviour for manuscripts as a specific indicator of aesthetics. It is noteworthy that chapters 6, 7 and 8 address art historical problems more abstract in nature than those in earlier chapters, and clearly demonstrate the benefit of the *unified* approach over traditional image processing techniques, which would necessitate multiple independent solutions. Each of the application chapters introduces some literature, presents a hierarchical image processing model and applies it to the art historical problem being discussed therein. Implementation specific details for some of these applications and other miscellaneous notes are collected in the appendices A.1 through A.8.

Part II

LITERATURE REVIEW

A history of art history and image processing, and arguments for a unified approach to digital analysis in art history.

ART HISTORY

2.1 INTRODUCTION

WHAT IS ART?

There really is no such thing as Art. There are only artists.

Ernst Gombrich [54]

Gombrich's Delphic observation theorises that art is something that is produced by an artist. A multitude of forms of human expression — architecture, dance, sculpture, music, painting, poetry, film, photography, graphic arts, ceramics, constructions, land art, installations and performance art — evoke emotional response from the audience. The label 'art' connects these very disparate objects, practices and processes. Recognising this diversity, various categorisations have been made within definitions of visual art (including that by Fernie [41]). Building on these one might propose a general set of guidelines for understanding what art is thought to be. Fine art has traditionally been used to distinguish arts such as painting, drawing and sculpture, from craft-based arts. The latter typically refers to those works created for a function — such as ceramics, jewellery, textiles, needlework and glass which are still termed decorative arts. This distinction does not apply so strongly in contemporary art making where a wide variety of media are used including, for example, ceramics and embroidery. There is, however, still a loose boundary between objects made with a specific function in mind and those which are made primarily for display. A broader definition of art encompasses those activities which produce works with aesthetic value, including film making, performance and architecture. For example, architecture has always had a close connection to painting, drawing and sculpture, two instances being the classical revival in the eighteenth and early nineteenth centuries and the Bauhaus aesthetic of the 1930s which frequently integrated fine art with design, craft and architecture. To make a broad generalisation, all forms of art are affected by each other. Contemporary definitions of art are not medium specific or particularly restrictive about the nature of aesthetic value. These ideas are associated with the Institutional Theory of Art which is probably the most widely used definition. It recognises that art can be a term designated by the artist and by the institutions of the art world, rather than by any external process of validation. On the one hand it provides an expansive framework for understanding diverse art practices, but on the other, it is so broad as to be virtually meaningless.

ART THEORIES One of the oldest schools of art history views art as an imitation of nature and the world around us. This can be traced back to Plato's idea of *mimesis* [121]. However this viewpoint does not allow for the artistic creativity in music, abstract art, avant-garde art, architecture and imaginative fiction or poetry. Clive Bell in *Art* [1914] outlined formalist aesthetics that emphasises the appearance and composition of the art work (its form) rather than its narrative content. He asserted that our response to a work of art arose from particular properties of its form: lines, colours, shapes and tones. He claimed that such an aesthetic response was also intuitive and involuntary. Although Bell's theory was simplistic, it underlined a shift of sensibility away from making evaluative judgements about art arising from criteria of resemblance or academic naturalism which had previously defined mainstream British, European and American painting. Bell's proposition that the narrative content of art was at best irrelevant, and at worst actually negated an object's aesthetic status, is important because it anticipated some of the critical responses to avant-garde abstraction which subsequently became widespread. In *The Principles of Art* [1943], Robin G. Collingwood argued that art is distinguished by a particular and unique emotion, not possessed by either craft or art as amusement, which he describes as lesser forms of technical art. As he puts it, art's place is to tell the audience '*the secrets of their own hearts*'. This point of view advocates that art communicates an authentic thought or a state of mind and consequently enables both the artist and viewer to gain self-knowledge. Collingwood's theory does not afford recognition of authentic pleasure as a product of look at or engaging with art. It also discounts representational art and craft arts. The idea of 'family resemblance' (see Warburton [133, pp. 149–50]) draws a parallel between aesthetic traits in artworks and inheritance of physical traits in families, and offers the possibility that new and innovative artistic forms can be incorporated into existing art 'families'. The Institutional Theory of Art leaves the burden of acceptance of an art work to the artist and art world. There are also Aesthetic theories of art that try and define beauty instead of art. One of the major proponents of such an approach, Kant claimed that aesthetic judgements claim universal validity, but remained vague on what the particular criteria of such aesthetic judgements might be, or how differing aesthetic judgements are possible [73].

ART HISTORIES Art history spans the entire history of humankind, from prehistoric times to the twenty-first century: from the paintings of cavemen to Botticelli angels and more recently, digital art. Art history emerged as a discipline that specializes in teaching people how to evaluate and interpret works of art based on their own perspective. This academic study of art, has been described as the 'discipline that

examines the history of art and artefacts' by Pointon and Peltz [107]. Like many definitions, it verges on the tautological, but Pointon succinctly registers the expansive scope of study which typifies the discipline — those objects and practices arising from human agency which are judged to have aesthetic value. The field has frequently been criticized for its subjectivity because the definition of what is beautiful varies from individual to individual, across cultures and time.

Many art historians and historiographers date the inception of the field of art history to the appearance of Vasari's *The Lives of the Artists* (1550 and 1568) [130]. Giorgio Vasari (1511–74), painter, architect and courtier, chronicled the work of Italian artists, describing the styles of painting, and of their personalities. However, this is by no means the earliest of such essays on the artist and his art. Cennino Cennini's (c.1370–c.1440) instruction manual *The Craftman's Handbook* (c.1390) [24] looked at the physical process of actually making a fresco or panel painting, including mixing the pigments and preparing the ground. A twelfth century manual *On Divers Arts* [56] explains pigments and their combination, and details the crafts of painting, metalwork and glasswork. With time, the craft of the art historian became more well defined and modern art history is generally regarded to have been founded by Johann Joachim Winckelmann in recognition of his attempts to apply scientific methodology to the study of sculpture and architecture. His first book, *Reflections on the Painting and Sculpture of the Greeks* (1755) [135] established the Greek cultural ideal, and his study *The History of the Art of Antiquity* (1764) [136] became central to the neo-classical movement. Later movements include aesthetic comparisons where Heinrich Wölfflin (1864–1944). In *Principles of Art History* [1915], he expounded a basis for comparative visual analysis which followed the opposition of the 'linear' and the 'painterly'. More recently, some of the most well known figures included Walter Friedländer (1873–1966), Sir Ernst Gombrich (1909–2001), Erwin Panofsky and Karl Popper (1902–94). The effect of critical theories since the 1970s, from structuralism to anthropology, psychoanalysis, feminism, Marxism and post-colonial critiques, has been collectively called *new art history*.

DISCUSSION In *Neurovisuality* [35], Whitney Davis discusses why vision science should take cues from recursions of neurovisuality, which he defines as the neural correlate of how human visual perception is shaped through interaction with items of visual and material culture. He suggests the possibility of neurovisuality giving way to a general theory of visual culture that is congruent with the science of vision. This is particularly encouraging for the fields of image processing and artificial intelligence that often apply recursive methods to problem solving. The grand goal of art history is a consistent and replicable definition of 'style', 'art' and 'aesthetics'. Art historical de-

bates are centred on the meaning and status of what artists actually do, and the elicitation of such meaning involves a deep understanding of the artworks themselves as well as the historical context in which the objects are created. The theory of mimesis approaches this process of understanding by treating art as part of an evolutionary tree, where each node would be an assemblage of artistic features: structural and socio-historical. Such a view does have parallels in the digital sciences as we shall see in section 3.3. The question that has been left unanswered to date, is the precise definition of these 'artistic features', if such a definition is at all possible. Even Kant steps away from this incontrovertibly massive task of deconstructing aesthetic judgement in universally measurable terms. Keeping in mind that there is a growing contention among the art historical community that the field is not self-critical (see [109]), it is the opinion of the author that the discursive nature of the field does have a lot to gain by stepping back to the analysis of form and structure of the image, the focus of Bell's theory. Gombrich [53] recognises the importance of 'the Beholder's Share' and it follows that any attempt at defining 'beauty' would rely on a comprehensive understanding of the mental processes of the audience of a piece of art. By his reasoning, an understanding of form and structure alone would not suffice in reaching the sought definition. As will be discussed in chapter 3, some mental processes have been approximated in various ways in the computer sciences. However, only a few of these approaches are generic enough to fit any art historical question. In this thesis, we limit ourselves to artistic properties of a structural or morphological nature, and propose an approach to reconcile art historical analysis with these properties. In the next section, the emerging field of digital art history is discussed, which raises several possibilities that could bring a degree of quantifiability to the task of describing art.

From the brief history of art history and theory presented in the earlier sections, the common thread that runs through the various schools and theories is that a work of art may only be understood in entirety, only in relationship with other objects and contexts. It follows that the definition of its beauty and aesthetics too requires a comprehensive understanding of such relationships. This line of reasoning has parallels in the Gestalt theory, and the hierarchical image processing approach defended in the next chapter also aims to bring a similar comprehensiveness to the task of image understanding.

2.2 DIGITAL ART HISTORY

In *History of Art in the Digital Age: Problems and Possibilities* [11, pp. 3–26], William Vaughan discusses the impact of computers on the field of art history. Drawing a parallel with the 'Gutenberg' revolution, he impresses the immense potential of exploration and analysis that is af-

forded by Information Technology, and discusses new and innovative methods of teaching that engages students in learning and interactive discussion. Other contributions to Bentkowska-Kafel et al. [11] look at the impact of digital collections and archives, showing exciting ways of using computers to make available information about collections and to provide a new accessibility to archives. The way such material can now be accessed via the Internet has revolutionised the search methods of scholars; but it has also made information available to all. It also offers the opportunity for new ways of exploring the structure and design of images.

Some of the active research areas in this field are: media archival (digital collections), retrieval, analysis and metadata management. Recently, there has been growing interest in the field of digitisation of sculpture and architecture with advances in three dimensional scanning technologies. A large part of the effort has been concentrated in the domain of visual material: paintings, books, photographs, manuscripts, scrolls and similar two dimensional media. Digitisation of artworks offers fascinating possibilities in exploring the structures of images: visual analysis, searching within images, and reconstruction. In the specific context of image analysis, Vaughan warns that the possibilities offered by artificial intelligence, expert and automated systems are vague at best. However some recent advances in the digital domain are aimed at aiding art historians in performing quantifications and measurements on images of paintings and manuscripts. For instance, Criminisi et al. [29] discusses computer vision and graphics techniques to analyse perspective in paintings. Even the highly subjective matter of aesthetics has received attention from a computational perspective [34]. Some of these advances will be described in detail in section 3.3 and in the relevant application chapters. Most of these advances are driven by very specific questions that are of interest to the art historian. This very specificity inherently makes it cumbersome to employ the same technique in combination with others, or to attack other problem domains. As will be argued in sections 3.3 and 3.4, a more generic and modular approach is necessary to answer the large number of questions that can be asked about a work of art.

2.3 MANUSCRIPTS

Manuscripts represent important artefacts of cultural heritage. The term ‘manuscript’ originates from the Latin *manu scriptus*, literally, ‘written by hand’. Some of the common genres of manuscripts includes the Bibles, liturgical books and calendars, among which the Bible was the most studied book of the Middle Ages. Illuminated manuscripts, briefly introduced in the previous chapter, are classified by art historians into their historic periods or types, including (but not limited to) late Antiquity, Insular, Carolingian, Ottonian, Ro-

manesque, Gothic and Renaissance manuscripts. The type of book that was most often heavily and richly illuminated, sometimes known as a 'display book', varied between periods. In the first millennium, these were most likely to be the Gospel Books, such as the Lindisfarne Gospels and the Book of Kells.

THE BOOK OF KELLS (Irish: Leabhar Cheanannais), Trinity College Dublin MS 58 was probably created ca. 800 AD. It contains the four Gospels of the New Testament along with other prefatory texts and canons. The text of the Gospels is largely drawn from the Vulgate, although it includes several passages from earlier versions of the Bible (known as the *Vetus Latina*). The manuscript today comprises 340 folios and, since 1953, has been bound in four volumes. The leaves are on high-quality calf vellum, and the exceedingly elaborate ornamentation that covers them includes ten full-page illustrations and text pages that are vibrant with decorated initials and interlinear miniatures and mark the furthest extension of the anti-classical and energetic qualities of Insular art. The Insular script of the text appears to be the work of at least three different scribes, although the exact number is a matter of ongoing debate [50, 59, 103]. The lettering is done in iron gall ink, while the pigments used throughout the manuscript were derived from a wide range of substances [14].

The Book has often been associated with the pinnacle of the Insular scribal tradition. The manuscript has been the subject of a extensive studies, from a wide range of perspectives. A conference at Trinity College, Dublin on *the Book* from 6–9 September 1992 [103] brought together experts from various fields, including physical sciences, mathematics and art history. At the conference Bernard Meehan presented a detailed analysis of the stylistic idiosyncrasies of the scribes of *the Book* in *The division of hands in the Book of Kells* [103, pp. 183–195]. Robert D. Stevick proposed some rules that the artists might have used to lay out content in the folios in *Page design of some illuminations in the Book of Kells* [103, pp. 243–256]. Also, Michael Ryan discussed the immense similarities between designs in *the Book* and metalwork of the same era in *The Book of Kells and metalwork* [103, pp. 270–279]. During the proceedings, apart from the aforementioned articles, a majority of the topics of manuscript research listed in section 1.2 were discussed in some form. These aspects are discussed in some detail in the next section.

2.4 QUESTIONS ABOUT THE BOOK OF KELLS

2.4.1 *Design*

2.4.1.1 *Forms*

The construction of interlace, key patterns and spirals has been studied using rule-based and mathematical approaches. The identification

of use of templates and copies, rule breaking, symmetry and asymmetry, etc. are often of interest to the art historian, being vital clues to the intentional use of such techniques to lend *the Book* an impressive aesthetic appeal and dating of stylistic trends.

2.4.1.2 *Scale of ornament*

One of the unique features of *the Book*, when compared with other manuscripts of the pre-medieval and medieval periods, is that design elements occur in a wider range of scales. For example letters appear at the smallest scales in regions of running text, and the largest letters appear on the *Chi Rho folio* (34r) where the χ initial is expanded to fill a majority of the page. Similarly, in the case of interlace, strands vary in scale ranging from very tiny to reasonably large, some of which have even more knotwork inlaid. While considering the question of the manual dexterity of the scribes, a quantitative measurement of the consistency in the thinnest brush strokes would provide an empirical insight into the level of skill and training of the scribes.

2.4.1.3 *Use of mechanical aids or optical tools*

The accuracy in execution of the design and layout, especially for the most intricate arrangements, is an indicator of the proficiency of the artists, and can be used to infer the extent to which mechanical or optical aids were used in the production of manuscripts.

2.4.1.4 *Repetition, symmetry, rotation and other transforms*

There is abundant use of these elements of composition in the designs of folios in *the Book*. The detection of such instances and their effect on perception of folios is another interesting avenue of exploration into the intentional use of such devices to enhance the readability or visual appeal.

2.4.2 *Execution*

2.4.2.1 *The number of scribes*

The calligraphy associated with text in *the Book* has a small range of stylistic variation, and some art historians agree that text can be classified into three major hands [59, 103], while other studies propose differing assessments of the matter [20, 50]. The Insular tradition of apprenticeship in the art of calligraphy, known to be stylistically very strict, ensured that the stylistic variation within each of the hands is minimal. A quantification of calligraphic styles could provide a more statistical basis for such studies and potentially improve the soundness of such studies.

2.4.2.2 *Comparison with other manuscripts and art forms*

The dating and periodisation of art works often relies on the determination of similarity in design motifs across art forms: manuscripts, metalwork, sculpture, etc. The identification of instances of a design being used as a prototype or exemplar for others allows historians to draw correlations and even causal links between the manuscripts and thus map the evolution of the particular styles such as the Insular tradition of illumination.

2.4.3 *Other questions*

2.4.3.1 *What did the Book look like originally?*

The Book has been subject to degradation over the centuries, intentional and otherwise. Surprisingly, it is still able to awe the naïve viewer and the art historian alike. A digitally restored rendition of *the Book*, apart from being a spectacular motivation in itself, also has the appeal from the point of view of preservation.

2.4.3.2 *Aesthetics and style of the Book of Kells*

Identification of some of the stylistic features that affect the enduring aesthetic appeal of *the Book* which make it stand apart from other manuscripts is another major motivation behind this study. Quantifying the various features listed above would provide an insight into this aspect of its history. Some of these features include the characteristic shape of faces and eyes, proportions of human figures. Apart from the human and human-like figures, the level of detail, minuteness of decorative elements, and levels of embedding such elements within larger elements are potential areas that can be used to characterise and even define the unique style of *the Book*.

2.4.3.3 *Colour*

The usage of colours on the various folios has not been analysed in much detail. A recent study by Bioletti et al. [14] of the pigments used in *the Book*, examines a limited number of sites for the identification of pigments used via Raman spectroscopy. While pigments that remain unidentified by such efforts may be inferred from a broader analysis of colours, how they correlate with content and aesthetics of the pages is also of interest to historians.

In the next chapter we introduce the fields of artificial intelligence and image processing, discuss the state of the art and recent forays into art historical applications. After revisiting the art historical questions listed above, we list those that might benefit from these fields and discuss possible approaches and argue in favour of a more comprehensive approach than is currently available.

IMAGE PROCESSING

ABSTRACT This chapter introduces the field of image processing in the context of artificial intelligence. A comparison of various methodologies in the subject are presented, with specific focus on bottom-up and top-down methods and a family of hierarchical image processing models. We review potential applications to art history and present the central argument of the thesis: the variety of problems that are of interest to the art historian requires an extensible unified image processing approach. It is argued that the family of hierarchical image processing models provide an optimal starting point for development of such a unified approach.

The academic field of image processing is a specialised branch of the computer sciences that focuses on development of methods for acquiring, processing, analysing, understanding and interpreting images and, in general, high-dimensional digital data from the real world. This area has close ties to the areas of artificial intelligence, machine learning and pattern recognition. The field of image processing has applications in a wide range of fields, from medical imaging to astronomy, satellite imaging, microscopy, biology, law enforcement and industrial applications. A brief overview of the basic concepts and methods are explained in the next sections.

3.1 IMAGE PROCESSING METHODS

A digital image is defined as a spatially ordered collection of pixels, each of which represents the smallest unit of pictorial information that can be digitally stored, retrieved and processed. Basic image processing operations include transformations applied to individual pixel intensities or groups of pixels. The two major components of computer vision systems are image capture and interpretation of captured images. Since the invention of charge-coupled devices, image sensors with millions of pixels have been commercially available. With improvements in imaging technologies, image acquisition has become easier and improved dramatically in terms of resolution in the last few decades. Cameras and other image capture devices produce tremendous amounts of data, and although the processing speed and storage capabilities of computers has increased almost exponentially over the same period, processing high-resolution images and video is still a challenging task for today's general-purpose computers. Continuing advances in processor speed and in the area of graphics processing units (GPU) promise higher public accessibility to faster and robust applications. While dedicated computer vision systems employed in very limited domains already outperform humans in terms of processing speed, the quality of interpretation has not reached any level comparable to that of the human vision system. Depending on the application, computer vision algorithms try to extract different aspects of the information contained in an image (or a video). Applications such as automated product quality control, license plate identification, ZIP code reading for mail sorting, and image registration in medical applications are some typical examples of such systems.

Two main approaches exist for the interpretation of images: bottom-up and top-down. Figure 1a shows the feed-forward image processing chain of bottom-up analysis. It consists of a sequence of steps that transform one image representation into another. Examples for such transformations are edge detection, feature extraction, segmentation, template matching, and classification. Through these transformations,

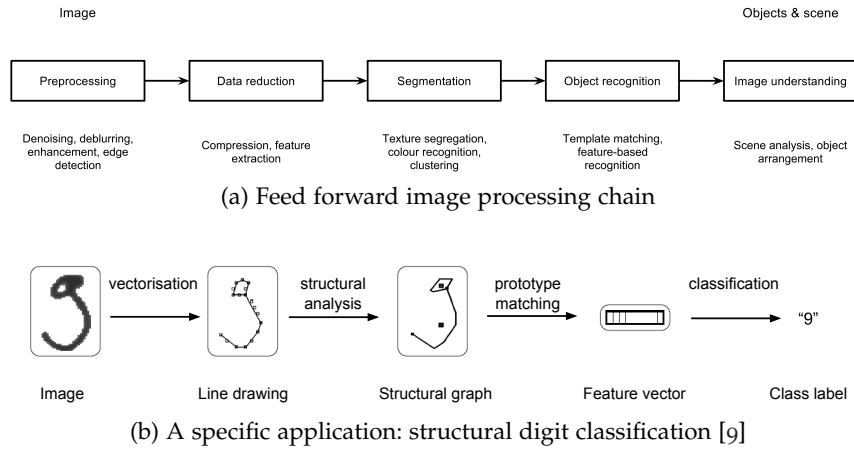


Figure 1: A bottom-up image processing scheme and a specific application

the representations become more compact, more abstract, and more symbolic. The transformations resulting from each individual step are relatively small, but the nature of the representation changes completely from one end of the chain, where images are represented as two-dimensional signals to the other, where symbolic scene descriptions are used. One example of a bottom-up system for image analysis is the structural digit recognition system [9], illustrated in Figure 1b. It transforms the pixel-image of an isolated handwritten digit into a line-drawing, using a vectorization method. This discards information about image contrast and the width of the lines. Using structural analysis, the line-drawing is transformed into an attributed structural graph that represents the digit using components like curves and loops and their spatial relations. In the application, the generated graph is matched against a database of structural prototypes. Finally, the attributes of the graph are compiled into a 'feature vector', i.e. an ordered collection of attributes, that is classified by a neural network.

The top-down approach to image analysis works in the opposite direction. It starts with a collection of abstract representations of the objects that are of interest. Each such model is associated with a set of parameters, which are estimated from features extracted from the image. Several hypotheses are formed that correspond to the presence of the object in the image. These hypotheses are checked for matches, on the basis of which the hypothesis is accepted or rejected. This method is successful if good models of the objects potentially present in the images are available and verification can be done reliably. Furthermore, one must ensure that the correct hypothesis is among the first ones that are generated. Top-down techniques are often used for image registration and for tracking of objects in image sequences.

While both top-down and bottom-up methods have their merits, the image interpretation problem is far from being solved. One of the major issues not addressed by either type of method is the segmentation-recognition dilemma. Frequently, it is not possible to segment,

i.e. separate objects from the background without recognizing them. Conversely, many recognition approaches require object segmentation prior to feature extraction and classification. Other limitations include invariance to object transformations, and poor performance in case of noisy images. In comparison, the ability of the human vision to predictively recognise objects well even in the presence of obstructions in the field of vision suggests a different approach is needed. Since the performance of the human visual system by far exceeds that of current computer vision systems, it may prove fruitful to draw inspiration from its architecture when designing computer vision systems. Although the human visual system is far from being understood, some design patterns that may account for parts of its performance have been revealed by researchers from neurobiology and psychophysics.

In the next section, we see how these limitations of following either a top-down or bottom-up approach are overcome by a family of image processing methods that exhibit the following characteristics:

- hierarchical architecture with increasingly abstract representations
- iterative refinement of interpretation through integration of bottom up, top down, and lateral processing
- adaptability and learning for specific applications

3.2 HIERARCHICAL ARCHITECTURES

As discussed earlier, most image processing applications attempt to bridge the immense gap between the objects in an image and image pixels using either one of the top-down or bottom-up approaches. As a result, extraction of any information of abstraction beyond the level of objects is cumbersome, without even considering the immense task of formulating of models that can represent elements of the complex designs present in sections of *the Book*. The deterioration of pigments in *the Book* over the centuries along with other types of degradation would further inhibit the performance and reliability of such methods.

It is important to note that the human visual cortex does not solve the problem of pattern understanding in this way. Vision models like the Memory Prediction Framework [55] break down input signals into a hierarchy of computing elements arranged in layers, that slowly reduce dimensionality. While most computer vision systems maintain multiple representations of an image with different degrees of abstraction, these representations usually differ in the data structures and the algorithms employed. While low-level image processing operators, like convolution, are applied to matrices representing discretised signals, high-level computer vision usually manipulates sym-

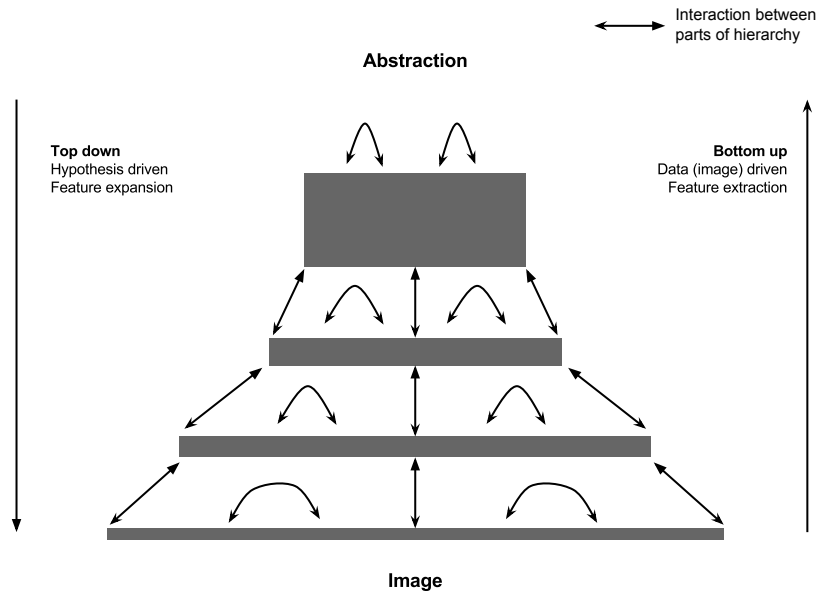


Figure 2: Integration of bottom-up, lateral, and top-down processing in a hierarchical architecture. Images are represented at different levels of abstraction. As the spatial resolution decreases, feature diversity and invariance to transformations increase.

bols in data structures such as graphs and collections. This leads to the difficulty of establishing a correspondence between the symbols and the signals. Furthermore, although the problems in high-level vision and low-level vision are similar, techniques developed for the one cannot be applied for the other. What is needed is a common representation that treats low-level vision and high-level vision in the same way. An abstract hierarchical architecture with local recurrent connectivity, that attempts to overcome the limitations of image processing approaches discussed earlier is discussed in the following sections and illustrated in figure 2. Images are transformed into a sequence of analog representations with an increasing degree of abstraction. As one ascends the hierarchy, the spatial resolution of these representations decreases, while the diversity of features and their invariance to transformations increase. It is noteworthy that since the computational elements in each layer function independent of each other, this model is amenable to parallelisation via community graphics processing units, as we shall see in later chapters.

ITERATIVE REFINEMENT The architecture consists of simple processing elements that interact with their neighbours. These interactions implement bottom-up operations (feature extraction), top-down operations (feature expansion), and lateral operations (feature grouping). The main idea is to interpret images iteratively. While images frequently contain parts that are ambiguous, most image parts can be interpreted relatively easily in a bottom-up manner. This produces

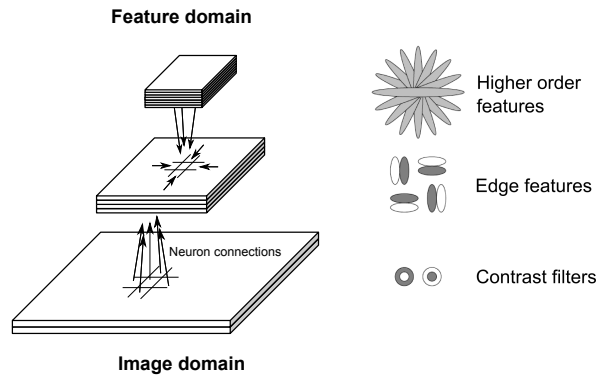


Figure 3: Neural abstraction pyramid

partial representations in higher layers that can be completed using lateral interactions. Top-down expansion can now influence the interpretation of the ambiguous stimuli in the lower layers. This iterative refinement scheme is a flexible way to incorporate contextual information. When the interpretation cannot be decided locally, the decision is deferred, until further evidence arrives from the context.

ADAPTABILITY AND LEARNING While current computer vision systems usually contain adaptable components, such as trainable classifiers, most steps of the processing chain are designed manually. Depending on the application, different preprocessing steps are applied and different features are extracted. This makes it difficult to adapt a computer vision system for a new task. Neural networks are a tool that has been successfully applied to a variety of machine learning tasks, and be organised into a hierarchical network with local recurrent connectivity. While the architecture is biased for image interpretation tasks, e.g. by utilizing the two dimensional nature and hierarchical structure of images, it is general enough to be adapted for different tasks. In the following sections, some hierarchical image representation approaches are discussed.

NEURAL NETWORKS (HIERARCHICAL) Such networks consist of multiple neural networks connected in the form of a cyclic/acyclic graph. Tree structured neural architectures are a special type of hierarchical neural network. The networks within the graph can be single neurons or more complex neural architectures such as multilayer perceptrons or radial basis function networks. Decision trees, hierarchical self-organizing maps, hierarchies of experts, hierarchical or tree-based classifiers are typical applications for hierarchical neural networks.

Some well known members of this family with this hierarchical organisation of processing units, with applications to computer vision, are the NeoCognitron [45], LeNet-5 [86], the HMAX model. Most of these models are inspired by the organisation of neurons in the pri-

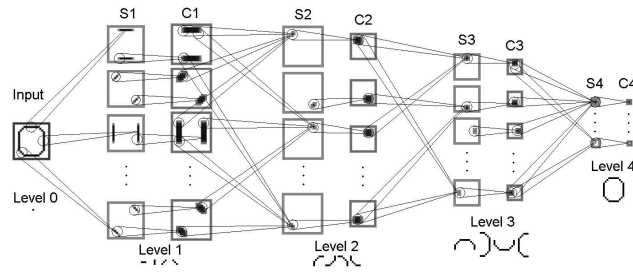


Figure 4: The Neocognitron model. Digit features of increasing complexity are extracted in a hierarchical feed-forward neural network.

mate visual cortex. Below we present a few members of this family which have differing applications, all of which coincide with different aspects of our study. A generic model of a neural abstraction architecture from [8] is shown in Figure 3. The order of layers after the image input layer is contrast extraction, edge extraction, followed by higher order feature extraction. A large number of studies suggest existence of recurrent connections from higher to lower layers in the visual cortex; this implies the ability to recreate parts of scenes at different levels in the hierarchy and has implications in image generation and image reconstruction, as discussed below.

3.2.1 *NeoCognitron*

One classical example of such adaptable hierarchical image representations is the NeoCognitron, proposed by Fukushima [45] for the application to digit recognition, a typical image classification problem. The architecture of this network is illustrated in Figure 4. It consists of several levels, each containing multiple cell planes. The resolution of the planes decreases from the input towards the upper levels of the hierarchy. The cell planes consist of identical feature detectors that analyse a receptive field located in the input. The size of the receptive fields increases with height, as do the invariance to small translations and the complexity of the features. The cells in the first level of the network analyse only a small input region and extract edge features. Cells located at the second level receive input from the edge features and extract lines and corners. Increasingly complex features, such as parts of digits, are extracted at the third and fourth levels. Feature detectors at the topmost level react to the entire image and represent digit classes.

Each level consists of three layers that contain different cell types. The S-layer is the first layer of a level. It contains S-cells that receive excitatory input via adjustable weights from small windows centred at the corresponding position in all C-planes of the layer below. S-cells in Level 0 access the input image directly. Not shown in the figure are V-cells that provide inhibitory input to the S-cells. V-cells are ex-

cited by all C-cells of the corresponding position in the lower level and compute a smoothed activity sum to control the gain of S-cells. The output of an S-cell depends on the total excitation, the total inhibition and a selectivity parameter. An S-cell reacts to features that resemble its specific excitatory combinations of its input ('afferent') cells. All S-cells of a plane share the same weights and thus extract the same feature at different locations. Invariance is produced in the network by the connections from the S-cells to the C-cells, which reside in the second layer of a level. These excitatory weights are not adjustable. They are pre-wired in such a way that a C-cell responds if any of the S-cells from a small window in the associated S-plane at the corresponding position is active. Hence, C-representations are blurred copies of S-activities that are less variant to input distortions. The NeoCognitron is trained level by level, starting at the bottom of the hierarchy. The adaptable excitatory weights of the S-cells can be trained either in a unsupervised mode or with supervision. For unsupervised training, the S-cells of a layer that correspond to similar positions first compete to react to an input pattern. The winning cell is then updated, such that it will react more strongly the next time the same pattern appears. In the supervised training mode, a human operator selects the features that a cell should respond to and the weights are updated according to a Hebbian rule [57] that is multiplied with a Gaussian window to give the features in the centre of the receptive field an advantage. Inhibition and excitation are increased simultaneously to make the cells more and more specific. Although the network is able to learn to recognize distorted patterns from relatively few training examples, training has been reported to be rather difficult due to the sensitivity of the network's performance to the choice of network selectivity parameters.

3.2.2 Occlusion reconstruction network

Fukushima also developed a neural network model that has an ability to restore missing portions of partly occluded patterns [46, 48]. It is based on the earlier NeoCognitron [45] and it mirrors the bottom-up pyramid used in the NeoCognitron model in the bottom half, and this allows a neural pathway from higher level abstractions in the feature domain to influence the image layer through the intermediate network connectivity. Thus, this network is capable of generating images using the learned features and fill in occlusions or degradation in an image. We note that such learned features are stored in the weights as strength of connections between neurons of two consecutive layers. The reconstruction property of this network has implications in the restoration of image degradation as discussed in considerable detail in chapter 6.

3.2.3 HMAX model for object recognition

This proposes a unifying theory for the task of immediate object recognition in primate visual cortex, which occurs in the first 150 milliseconds or so. This model integrates evidence from neuroscience into a feed-forward computational object recognition model. Unlike the occlusion reconstruction network, it does not consider backward connections. An improvement on this model is discussed in chapter 4, and an overview follows.

The HMAX hierarchical feature extraction model was proposed by Riesenhuber and Poggio [111]. The architecture of the network is shown in figure 5. Similar to the NeoCognitron, it consists of alternating S-layers and C-layers. The S-layers contain feature extracting cells that compute a weighted sum of their inputs, followed by a rectifying transfer function. S-cells receive their inputs from C-cells at corresponding positions in the next lower layer. C-cells are used to pool a group of S-cells that share some parameters, but differ in one or more other parameters. They compute the maximum of the activities of these S-cells. Hence, C-cell responses are invariant to the parameters spanned by their associated S-cells.

Again, when going up the hierarchy, the receptive field size, i.e. the effective part of the image layer that is connected to a particular cell via intermediate layers, of the feature detectors is enlarged, the feature complexity rises, and the responses become more and more invariant to input transformations, such as shifts or rotations. Cells in layer S1 correspond to V1 simple cells. They analyse the input image and extract oriented features at different positions, scales, and orientations. At the S1 layer, the image space is sampled at every pixel, and four orientations are extracted. The number of S1-cells is reduced in layer C1 by pooling cells with the same orientation, similar position, and similar scale. C1 cells correspond to V1 complex cells that detect oriented image structure. S2 cells receive input from a neighbourhood of C1 units of arbitrary orientation, yielding S2 cells of 256 different types. They detect composite features, such as corners and line crossings. All cells of a certain type are pooled to a single C2 cell that is now totally invariant to stimulus position. At the top of the hierarchy reside view-tuned cells that have Gaussian transfer functions. They receive input from a subset of the C2 cells. Almost all weights in the network are pre-wired. Only the weights of the view-tuned cells can be adapted to a dataset. They are chosen such that a view-tuned unit receives inputs from the C2 cells most active when the associated object view is presented at the input of the network. Riesenhuber and Poggio showed that these view-tuned cells have properties similar to the cells found in the inferotemporal cortex. They also demonstrated that view-invariant recognition of 3D paper clips is possible by combining the outputs of units tuned to

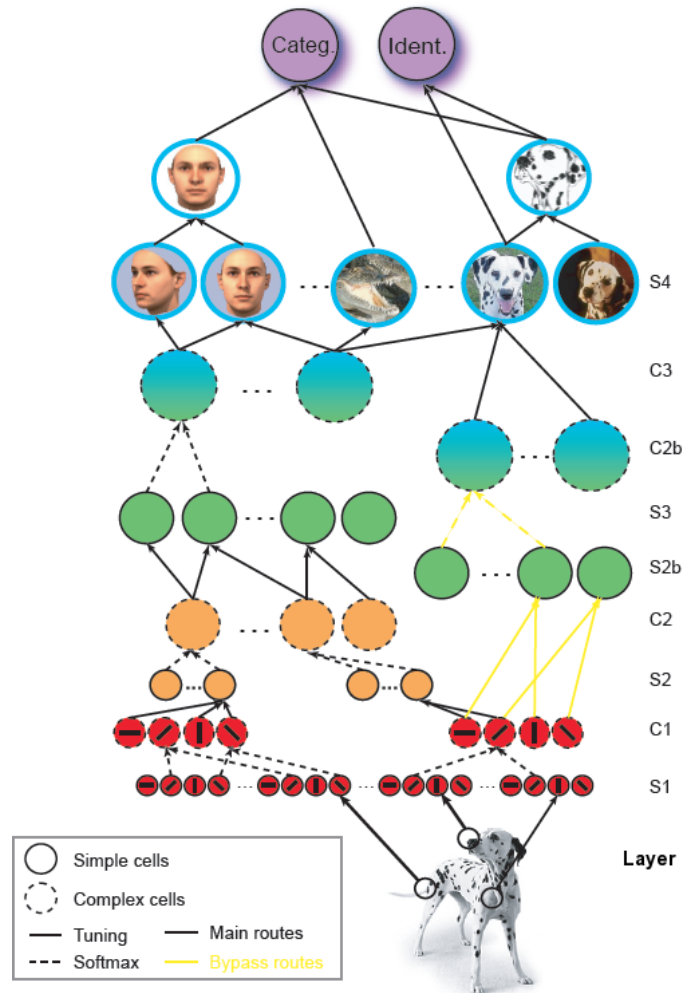


Figure 5: Functional primitives of a theory of object recognition [118]. The network consists of alternating S-layers and C-layers that extract features of increasing complexity, size, and invariance.

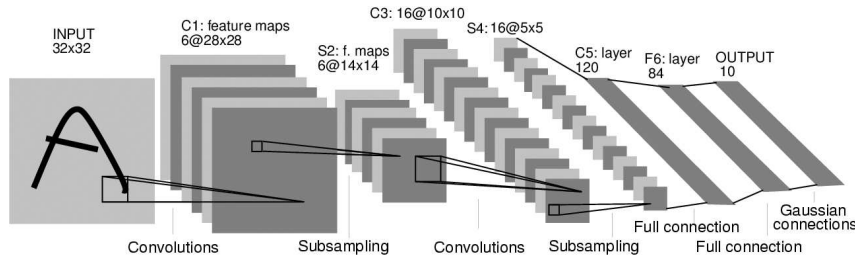


Figure 6: Convolutional neural network LeNet-5 [86]

different views of an object. Other applications of the model include categorisation of images showing various objects. Since the lower levels of the hierarchy contain retinotopic representations, features of spatially separated objects do not interact and hence are bound by spatial proximity. Features in the higher levels are complex combinations of simple features. Since there are many such combinations, it is unlikely that the features of two objects can be combined to a valid third object. The use of the maximum operation for pooling makes the cell responses invariant to input transformations and also suppresses noise. Thus such a model is inherently resistant to low amounts of noise in images.

3.2.4 Convolutional Networks

The creation of features by enumeration of all possible sub-feature combinations is easy, but computationally inefficient. For practical applications, such as optical character recognition and the interpretation of handwritten text, where often a high speed of operation is a requirement, the network size plays an important role since real-time conditions must be met for the network recall. If more of the network parameters can be adapted to a specific task, smaller networks suffice to extract the relevant features. One example of a fully adaptable hierarchical neural network is the convolutional network proposed by Cun et al. [32] for the recognition of isolated normalized digits. An improvement on the basic architecture proposed therein called LeNet-5 [86] is illustrated in Figure 6. The network consists of seven layers and an input plane that contains a digit. It has been normalized to 20×20 pixels and centred in the 32×32 frame. The input intensities are scaled such that the white background becomes -0.1 and the black foreground becomes 1.175 to obtain inputs with approximately zero mean and unit variance. The first five network layers are alternating convolutional (C) and sub-sampling (S) layers that contain an increasing number of feature maps.

A convolutional layer computes local image features by convolving the previous representation with 5×5 kernels. These layers decrease in size since only such pixels for which the receptive field lies entirely in the previous layer are computed. If the previous represen-

tation consists of multiple feature maps, multiple feature windows describing the same image location are combined to compute a more complex feature. For C1 and C5 all S-features of the previous layer are used, while C3 features access different subsets of at least three S2 features. The size of the feature maps is further reduced by the sub-sampling layers that compute the average of 2×2 windows of an associated feature map in the next lower C-layer.

The upper two layers of the network have full connectivity with the previous layer. Layer F6 has a size of 7×12 and represents the desired output in a distributed code which is an icon that looks like an idealized digit. This has the advantage that similar patterns are represented by similar icons, facilitating post-processing if these patterns are confused. The neurons in the first six layers of the network pass their activations through a sigmoidal transfer function $f(\alpha) = \alpha \tanh(\beta \alpha)$ that limits the output values to $[-\alpha, +\alpha]$ ($\alpha = 1.7159$, $\beta = \frac{2}{3}$, such that $f(1) = 1$, $f(-1) = -1$ and $|f''(\alpha)|$ is maximal at 1 and -1). In contrast, the 10 output units in the topmost layer compute the difference between their weight vector and the F6 activity and pass it through a Gaussian transfer function. Hence, they are radial basis function (RBF) units that signal the class of the digits in a 1-out-of-10 code. While the weights of the RBF units are fixed to represent the icon associated with the class, all other weights are trained by gradient descent. The gradients of the weights with respect to a loss-function are computed by the back-propagation method [114]. Since shared weights are used, the gradients of the weight instances must be averaged when updating a weight. The degree of weight sharing is high in the lower levels of the network. This also allows for sharing of examples since many small windows are contained in a single digit.

So far, we have introduced the field of image processing and the basic types of approaches: bottom-up, top-down, and lateral. Following this, we introduced a class of image processing techniques that share organisational similarities, and have a wide variety of applications. In the next section, we attempt to tie image processing techniques to specific problems in the study of manuscripts.

3.3 MANUSCRIPTS AND IMAGE PROCESSING

Let us investigate the various aspects of study of manuscripts as covered in section 2.4 and later generalised and consolidated into broader areas of inquiry in section 2.4.

At the most basic level, a manuscript is a tangible object. The physical and chemical properties of the parchment and inks and the interaction between the two have been studied by a variety of techniques, the most popular of which is Raman spectroscopy [87]. *The Book* has recently been the subject of a study [14] on the identification of pigments used in it, and while such expensive and invasive approaches

are extremely precise, they are severely limited by the percentage of manuscripts that can be feasibly analysed. Image processing provides an interesting alternative for the analysis of pigments from the visual appearance alone. A reconstruction of the palette of documents can be easily accomplished using methods such as [76] and more recently [84]. Digitally modelling brushes and brush strokes [140] and interaction of paint [123] or pigment and the substrate have allowed virtual recovery of brush strokes [120]. Such techniques may provide valuable clues to understanding the techniques of painting and penmanship used during the creation of manuscripts. Wolf et al. [137] very recently developed a series of digital tools and aids for palaeographic analysis of the Cairo Genizah manuscripts. Older approaches such as [65] took a more structural and mathematical approach to covert characters in printed documents back into smooth curvilinear representations, which if applied to the text found in manuscripts could provide interesting insights in characterising calligraphic trends.

Moving from calligraphy to textual content, studies such as Carol Farr's *Textual structure, decoration and interpretive images in the Book of Kells* [103, pp. 437–449] that look into organisation and stylistic devices used in the manuscript text could benefit from the computer science discipline of natural language processing for, say stylistic comparisons across a large variety of digitised manuscripts. The art theory of mimesis, introduced in the previous chapter, has a parallel in artificial intelligence technique of evolutionary or genetic algorithms, and this speaks to the possibility of modelling artistic evolution using such methods.

Coming back to the image domain, structural measurements of and deconstruction of the complex designs (see appendix A.8) found in illuminated manuscripts could provide clues to the use of mechanical or optical tools in the design or execution processes. Inspired by the work on Celtic patterns by Bain [4], a computational technique for the generation of Celtic interlace [74] has been recently developed. However, to unravel the structure of interlace in an automated fashion, we could develop an image processing solution following a bottom-up or top-down methodology. Under a model based approach, reconstructing the existing designs in the Book requires such an approach to be merged with segmentation or image tracing techniques such as [116], [2] or [142]. Some existing well defined mathematical or procedural models include those for interlace [30] and key patterns [37]. Such models tend to be defined after making some heuristic assumptions. The artists of manuscripts such as *the Book* appear to embrace a much freer form of design than other pre-medieval and medieval manuscripts. Even if all the elements used in Celtic design were successfully modelled, the rules of combined usage of these elements a composition would be prohibitively complex.

While all forms of artistic expression are continually being inspired by other other works and forms of art, the comparison of designs across manuscripts, and art forms, would need a common way of representation of designs. Also the problem of deterioration of manuscripts over the ages via various physical and chemical processes would pose serious issues for a majority of image processing solutions. Change or loss in pigments has been virtually reversed in paintings [12, 105] and the problem of inks bleeding through the parchment to the other side has been tackled Tonazzini et al. [129].

The aspect of aesthetics of manuscripts has not been seriously explored in the computer sciences to date, though recent neuroscientific and experimental evidence from paintings, etc. has helped enumerate several preferences in the visual aesthetics [104]. While capturing the social political and cultural context of manuscripts, and more generally art, is beyond the scope of applicability of image processing, a lot of the problems can be addressed by a hierarchical image processing model as introduced in section 3.2 as argued below.

3.4 DISCUSSION

The image processing models discussed in the section 3.2 have properties which are directly applicable to a number of art historical concerns. One of the most important features of a majority of the members of this family is object recognition; from the previous paragraphs, the problem of detection of design elements is a potential application of this property. The analysis of calligraphic styles would require extraction of specific morphological features for differentiating between the stylistic identities of the scribes involved in the production of a manuscript. The extraction of heuristic or semi-heuristic features from the hierarchy, and building task/application specific layers that operate on such calligraphic features is made possible by the abstraction provided by the hierarchy.

The model also demonstrates the property of localisation, that allows us to search for specific objects or structures in images. As long as it is possible to define the design elements used in manuscripts as collections of uniquely recognisable shapes or features, localisation can be used for the problem of deconstruction of designs. The occlusion reconstruction network [48] with its ability to reconstruct missing parts of images raises the possibility of use as a degradation removal technique. Given that learning is required for the applicability of such a degradation reversal technique, the reconstruction would consequently be limited by the learnt image features.

The simple recursive nature of the hierarchy also suggests the potential of processing increasingly complicated questions about the image input to such a network. While the introduction of new layers that draw conclusions based on some logical combination of previous con-

clusions sounds promising, it might provide a new way to ask and answer questions closer to image aesthetics than currently possible.

The Book is one of the most prominent manuscripts famous for its employment of complicated designs, illuminated letters, human and animal figures. With a limited set of pigments used in the production, its style may be characterised as cartoon like: distinct areas of the design are coloured in one pigment, and there are a very few regions where one colour is intended to fade into another creating a gradient. Such images do reduce the burden on the complexity of an image processing solution.

As such the hierarchical image processing models discussed in section 3.2 provide a starting point in developing a digital analysis framework for manuscripts, that can be customised and extended to accommodate the wide variety of questions art historians are interested in. In the following part of the thesis, we discuss some of these questions and apply such models to answering them.

Part III

APPLICATIONS

Development of an object recognition and localisation computer vision framework and applications to the art historical analysis of manuscripts.

CHARACTER ANALYSIS

ABSTRACT A significant portion of most manuscripts is devoted to textual content; consequently understanding the textual information present in these historical artefacts has received considerable attention, specifically the characterisation of calligraphic styles. This chapter discusses the topic of character and hand analysis. A novel way of hand classification is proposed that can potentially be used to distinguish, compare and define individual scribal styles.

4.1 INTRODUCTION

In art history, pre-medieval and medieval manuscripts have been seen as the vehicle of the spread of religion across Europe. Camille [22] discusses some of the implications of medieval literacy and illiteracy and suggests that the relationship between text and image in manuscripts of the age is a complementary one. Illustration work often aided prose by increasing retention among the masses, and mistrust of the materiality of images was balanced by the written word. Larson and Picard [85] have measured the perceived aesthetic and cognitive effects produced by elevated mood. In particular, they demonstrated the benefits of good typography, concluding that good typography induces a good mood. The study makes certain assumptions about the aesthetics of typography and it is conceivable that the notion of good typography can be extended to good calligraphy in the case of manuscripts, allowing for the change in concept of aesthetics over time. It is generally assumed that manuscripts such as *the Book* were completed by a school of monks, who actively practised and honed the arts of calligraphy and illumination, and the importance of these arts in the perception of *the Book* can be inferred from such studies.

A detailed examination of the half-uncial *a* and uncial *a* throughout *the Book* by Gabhann [50] shows a remarkable consistency and suggests that these letters are the work of a single scribe, in agreement with the conclusions of Brown and Verey [20], who described the script as the work of ‘one great scribe’. However, an earlier study by Henry and Kennedy [62] on the division of hands proposed that there were three distinct scribes. *The Division of Hands in the Book of Kells* by Bernard Meehan [103, pp. 183–195], however, analysed the stylistic variations within Henry’s hand ‘C’ and concluded that there are four scribes: hands ‘A’, ‘B’, ‘C’ and ‘D’ instead of just three. All of the aforementioned studies are based on a meticulous (human) expert analysis of various structural characteristics of the various letters in the folios. Most of these characteristics are shared with studies in type and typography, that have been studied as an art since the invention of the movable type and more recently as a science. Renowned typographer and poet Robert Bringhurst provides a comprehensive overview to the art of typography with his masterful style guide on the subject [19]. It covers the practical, theoretical, and historical aspects of the field. *Type and Typography* [6] takes the reader through every aspect of typography, from the history of language and writing systems to the invention of movable type and the evolution of the digital systems of today. It discusses the various characteristics features that uniquely define a typeface. The chapter on ‘Form’ illustrates these characteristics by examples of ‘part-alphabets’ for 60 typefaces and detailed explanations of their visual features. Subsequent chapters on ‘Manufacture & Design’, ‘Conventions’ and ‘Structure’ delve

into the importance of layout, specifically grids, and spacing. *Design Elements, Typography Fundamentals: A Graphic Style Manual for Understanding How Typography Affects Design* [31], aimed at type designers, explores the fundamentals of typography in a more instructional fashion rather than historical survey.

In an effort to standardise the classification of type, Kindel [77] aimed to catalogue specimens of fonts and typefaces from the last 10-15 years in a descriptive way. These were catalogued alongside the older historical typeface material and qualified by descriptive fields based upon their visual appearance, or, as Sutton and Bartram describe, as ‘typeforms’ [127]. Dixon [36] addresses the limited scope of existing typeface classificatory systems to adequately describe the diversity of forms represented within current type design practice and proposed a new framework for the description of typeforms.

Since the invention of graphical displays in computing, digital typography has been evolved into computerised typeface (font) design systems such as developed by Knuth [80, 81]. Such systems model the motion and physical characteristics of a pen. Interaction of brushes with parchment or paper with various media (water, oil, etc.) has also been pursued. For instance, Wong et al. [140] proposed a parametrised brush model called the Virtual Brush, which enables efficient representation and generation of Chinese calligraphic writings such that the rendering is scalable in resolution and allows high quality publishing mirroring the aesthetic appeal of traditional calligraphy.

While digital design and rendering of fonts has progressed smoothly over the last four decades, digital analysis of calligraphic images was initially posed as the problem of separating text from the background in images. This problem is a part of the segmentation problem we discussed earlier in section 3.2. A typical advance in this space by Leydier et al. [89] proposes an adaptive algorithm for the segmentation of colour images suited for document image analysis. The algorithm is based on a serialisation of the k-means algorithm [91] that is applied sequentially by using a sliding window over the image. The algorithm reuses information about the clusters computed by the previous classification and automatically adjusts the clusters during the windows displacement in order to better adapt the classifier to any new local modification of the colours. For digitised documents, we propose to define several different clusters in the colour feature space for the same logical class.

Leedham and Sagar [88] reported the need for fast and accurate identification of individuals via verifiable biometric identification such as handwriting and signature analysis. Although image processing was quite nascent at this stage, the possibility of using forensic handwriting analysis was advocated using handwriting experts. Since then, computer-aided forensic handwriting analysis has developed into a

separate research area. Franke and Köppen [42], Franke et al. [43] detail an layered assistance framework for forensic handwriting experts that builds on fundamental document-independent filters for textured, homogeneous background removal and for foreground removal. The automation of handwriting and signature analysis is not as mature as the field of character recognition, which simply aims to accurately identify characters from text in images.

One of the most popular character recognition approaches [86] uses multilayer neural networks trained with the back-propagation algorithm constitute the best example of a successful gradient-based learning technique. Given an appropriate network architecture, gradient-based learning algorithms can be used to synthesise a complex decision that can classify high-dimensional patterns, such as handwritten characters, with minimal pre-processing. Convolutional neural networks, which are specifically designed to deal with the variability of two dimensional shapes, are shown to outperform all other techniques of the time. Real-life document recognition systems are composed of multiple modules including field extraction, segmentation, recognition, and language modelling. LeCun also describes a new learning paradigm, called graph transformer networks (GTN), which allows such multi module systems to be trained globally using gradient-based methods so as to minimise an overall performance measure. [9] evaluates various classifiers to the recognition of handwritten digits. Four different classification schemes are compared in parallel and sequential combination. Furthermore, it is described how to tune the sequential combination using a boosting technique.

But for the purpose of understanding the scribe's individual styles, a quantification of stylistic features of the calligraphy. Itoh and Ohno [65] present an algorithm that automatically generates outline fonts from a grey-level image of a character obtained by a scanner. This algorithm begins by extracting contour points from the image and dividing the points into a number of segments at the corner points, and attempts fitting a piecewise cubic Bézier curve to each segment, using least-squares fitting. Wong et al. [140] also develop an automated approach to estimate the Virtual Brush parameters from images of Chinese calligraphy, such as the properties of brush hair and variations of ink deposition along a stroke trajectory, directly from an image of calligraphic writing.

Word spotting was developed by Rath and Manmatha [110] as an alternative to automatic handwriting recognizers, which at the time, did not perform well in classifying calligraphic styles. The words in a collection are matched as images and grouped into clusters which contain all instances of the same word. By annotating clusters, an index that links words to the locations where they occur can be built automatically. Rath analysed a range of features suitable for matching words using dynamic time warping, which aligns and compares

sets of features extracted from images. Most recently, automatic handwriting matching has been explored in the context of the Cairo Genizah, a collection of hand-written documents containing approximately 350,000 fragments of Jewish texts discovered in the late 19th century [137]. With a focus on palaeography, script style, and more specifically, handwriting, are being used to identify fragments that might originate from the same original work. However, such matched fragments, are currently identified manually by experts.

DISCUSSION The most reliable work on identifying the distinguishing features of individual artists in the case of works which are produced by groups has been carried out by expert handwriting analysts and art historians specialising in calligraphy. Although forensic handwriting analysis has been aided by image processing techniques, as pre-processing and feature extraction routines, this area largely continues to depend on human experts. Clearly, different studies have used differing image processing techniques to serve the particular art historical problem they are concerned with. Moreover, there is a lack of approaches that may be also used directly (or with minor modifications) for other art historical applications, as has been previously discussed in section 3.4. Therein, we discussed the possibility of using a hierarchical image processing framework for art analysis applications, such as that of hand classification, as we outline in the following sections.

4.1.1 Questions

The text in *the Book* has been the subject of scrutiny in terms of the content as well as style. Some of key areas of research in relation to content analysis are the textual structure, literary devices used, proximity based associations with illustrations and exegesis [103]. Stylistic analysis of the calligraphy in *the Book* [20, 62, 50, 103, pp. 183–195] and other manuscripts has focused on the questions discussed below.

STRUCTURAL CHARACTERISTICS OF THE SCRIPT At times, historians such as Françoise Henry describe the script in terms of the morphology of the letters: particularly, their dimensions, relative spacing and the tendency to use majuscule over minuscule forms among others.

DESCRIPTIVE AND EMOTIONAL CHARACTERISTICS OF THE SCRIPT While some qualities associated with script by such studies, such as ‘cramped’ and ‘compact’, are very specific, other attributes are more abstract, viz. ‘extrovert’, ‘imitation ... to the point of caricature’ and ‘archaic’.

IDIOSYNCRASIES Some letters are stretched horizontally within the line, others are displaced vertically to signal line endings, even accompanied by animal figures. Some interesting combinations of colour and letter forms are also found.

DEFINITION OF THE STYLE OF SCRIBAL ‘HANDS’ Ultimately such analysis are focused on the problem of identifying the distinguishing stylistic features and characteristics of the scribes involved in the production of manuscripts. This is very similar to the identification of ‘hands’ in paintings by Giovanni Morrelli as described by Ginzburg and Davin [52] in comparison with Sigmund Freud and Sherlock Holmes.

4.1.2 *Problem definition and scope*

In this chapter, the problem of definition of ‘hand’ styles for calligraphy in manuscripts is treated as a complex classification problem. An automated digital system for hand classification must address the issues of — recognition of a part of an image as a letter of the alphabet, the computation of various stylistic features and finally, the combination of these metrics of style into a classification scheme to identify a particular (text) region of a folio as being written by a particular scribe. The recognition of letters from an image has been well researched under the image processing fields of character recognition and, more generally, object recognition. The problem of extraction of stylistic features from these identified letters would provide valuable information for the differentiating ‘hand’ styles. The dimensions of letters and the thickness of the pen [62, 36] have been identified as key structural features that may be used to describe letters. More importantly, these features are present for all letters of the alphabet, as opposed to features which may or may not be present for particular letter, such as heights of ascenders, descenders, and bars, and slants of axes [6]. The classification of a letter into a particular scribal hand would definitely be influenced by such letter-independent features. We discuss some approaches that may be taken to integrate character recognition and structural features of characters into a hand classification scheme and argue in favour of a hierarchical image processing approach in the next subsection.

4.1.3 *Possible solutions*

In developing a hand classification image processing application for manuscripts, the most naïve solution would be to combine an optical character recognition method such as LeNet-5 [86] with calligraphic feature extractors, and finally use these features in a classifier that is trained to recognise hand styles. Such a modular approach would

assume that the character recognition, calligraphic feature extraction and the hand classifier function independently of each other. An alternative approach would be to allow these separate modules to interact with each other. Although such an approach would require introspection into the working of each of the three stages, it has the benefit of being extensible to incorporate other tasks, such as degradation removal and reconstruction. If we follow this line of reasoning, such a system could use a hierarchical network architecture such as those discussed in section 3.2 to tackle all three aspects of the problem: character recognition, feature extraction and ultimately hand classification. It is important to note that such a biologically motivated approach also allows us to perform computations of a more abstract nature, such as layout analysis, and even answer abstract questions about the aesthetics of text, if we consistently use such a layered architecture.

In the next section, we delve into the problem in more detail, and describe an image processing model that is hierarchically structured to address the ‘hand’ classification problem.

4.2 HAND CLASSIFICATION

A calligraphic ‘hand’ may be defined as a collection of letters, which considered together significantly differ from other such collections. The problem of classification of hands in a manuscript such as *the Book* can thus be divided into two well defined stages. The first stage is character recognition, followed by the second stage where we group letters into a hand. For the first stage, we use an object recognition framework [98] as explained shortly. Following this, we quantify a few morphological characteristics of each character, and use these in a classification scheme to demonstrate the viability of this approach.

4.2.1 Character recognition

The specific goal of character recognition may be treated as a particular application of the image processing problem of object recognition. The object recognition model we use to reach this goal was proposed by Mutch and Lowe [99]. It is structurally similar to the convolution network proposed by LeCun et al. [86] that is considered a major advance for the application of recognition of handwritten characters. Both these approaches start with an image layer of (grayscale) pixels, and successively compute features in higher layers in the model, alternating between S- and C- layers (named by analogy with the simple and complex cells discovered by Hubel and Wiesel [63]). Simple (S) layers use convolution with local filters to compute higher-order features by combining different types of units in the previous layer, whereas complex (C) layers increase invariance by pooling units of

the same type in the previous layer over limited ranges, at the same time, reducing the number of units by sub-sampling. The model is illustrated in figure 33. Such recent object recognition models have improved in terms of biological plausibility with increased physiological data about the ventral stream in the primate and human visual systems (see [118]). The application of the model proposed by Mutch and Lowe [99] to the slightly wider domain of character recognition in manuscripts allows us to consider future possibilities such as improved resilience to degradation (as will be demonstrated in chapter 6). The customisability of this network for particular classification tasks provides us the framework for the contribution of this chapter with regard to classification of hands.

4.2.2 *Object recognition model*

The image processing model we use for the problem of character recognition was by Mutch and Lowe [99] as an improvement on Mutch and Lowe [98], which builds up on a series of improving visual models by Serre et al. [117] based on the earlier HMAX model by Riesenhuber and Poggio [111]. HMAX was designed to account for the tuning and invariance properties of neurons in Inferior temporal cortex. Rather than attempting to learn its bottom-level (S1) features, HMAX uses hard-wired filters designed to emulate simple cells. Subsequent C layers are computed using a hard MAX — a C unit's output is the maximum value of S units of the preceding layer. Serre et al. introduced learning of intermediate-level shared features, made additional quantitative adjustments, and added a final (non-biologically motivated) SVM classifier to make the model useful for classification. The model consists of an initial image layer and four subsequent layers, each layer built from the previous by alternating template matching and max-pooling operations in the following order: image layer, Gabor filter (S1), local invariance (C1), an intermediate feature (S2) and global invariance (C2). Finally we use an all-pairs linear Support Vector Machine (SVM) classifier after C2 vectors are normalised. These layers are discussed in some detail below and again in appendix A.1.

IMAGE LAYER This layer contains the image that is input to the model. It iteratively creates scaled copies of the image over a range of scales, such that the copies at each scale is $2^{1/4}$ smaller than the last.

GABOR FILTER (S1) LAYER The S1 layer is computed from the image layer by centring two dimensional Gabor filters with a full range of orientations at each possible position and scale. The base model follows [117] and uses 4 orientations. Where the image layer is a two dimensional pyramid of pixels, the structure of the S1 layer is four

dimensional, having the same three dimensional pyramid shape, but with multiple oriented units at each position and scale (see figure 33). Each unit represents the activation of a particular Gabor filter centred at that position and scale. This layer approximates the behaviour of simple cells found in the visual cortex that function as edge detectors.

LOCAL INVARIANCE (C1) LAYER Each unit in this layer is connected to several S1 units (of the same orientation) and combines their output to create position and scale invariance over larger local regions. Consequently, the layer has a reduced number of units compared with the preceding S1 layer. For each orientation, the S1 pyramid is convolved with a three dimensional max-filter. A C1 unit's value is simply the value of the maximum S1 unit (of that orientation) that falls within the input of the max-filter. This layer also achieves sub-sampling of the previous layer and the resulting C1 layer is smaller in spatial extent but has the same number of feature types (orientations) as S1. This layer implements a biologically plausible model for complex cells found in the visual cortex of the mammalian brain.

INTERMEDIATE FEATURE (S2) LAYER Each unit in this layer is connected to several C1 units centred at the its position and scale and matches their output with a small number of prototype C1 responses. These prototype responses are randomly sampled from C1 layers during the training phase, and behave as fuzzy templates consisting of a grid of simpler features that are all slightly position and scale invariant. The prototype responses or patches represent the intermediate features of the model. It may be noted that these patches vary in size to account for learning shapes (at smaller scales) and textures (larger scales). Across the collection of these intermediate model features, some may represent objects and textures of interest, while others may be disregarded for particular recognition tasks. The selection of features important to a particular object recognition task is performed by the Support Vector Machine layer described later.

GLOBAL INVARIANCE (C2) LAYER The function of this layer is to remove all position and scale information from the intermediate S2 features computed in the previous layer. This is achieved by selecting the maximum response (anywhere in the image) to each of the model's prototype patches.

SVM CLASSIFIER The position and scale independent C2 vectors collected from the previous layer are classified using an all-pairs linear Support Vector Machine.

Mutch and Lowe [99] suggest several performance improvements over Serre et al. [118] that aim at reduction in the number of features that are finally combined in the classification layer as listed below.

1. Sparsify S2 input
The number of inputs to an S2 feature is reduced to one per C1 position by keeping track of the identity and magnitude of the dominant orientation (maximally responding C1 unit) at each position in the patch. This is illustrated in figure 34. This however does introduce some added invariance in orientation which is offset by increasing the number of orientations used in S1 and C1.
2. Inhibit S1/C1 outputs
Similarly, non-dominant orientations are ignored by allowing S1 and C1 units to compete in describing the dominant orientation at their location. Such lateral interaction is often seen in the cortex, where lateral inhibition refers to units suppressing their less-active neighbours.
3. Limit position/scale invariance of S2 features
To account for the fact that neurons in the visual cortex above levels corresponding to the S2 layer are not fully invariant to position and scale, finding features from this layer is restricted to a fraction of the image size.
4. Select features that are highly weighted by the SVM
Feature selection, the process of elimination of unnecessary features, is performed by dropping features with low weights in the classification layer.

MODIFICATIONS In applying the above model to character recognition, it was noted that the inhibition of S1/C1 non-dominant orientation features did improve speed but had an adverse impact on recognition accuracy. This is expected as the scale of letters in our application is higher than that of objects present in images used in the validation of the aforementioned studies. This issue will be revisited in chapter 6 where lateral interactions are studied in detail. All the other modifications suggested before were incorporated into the model and tested as described below.

TRAINING, TESTING AND RESULTS The image dataset used for character recognition contains 180 images comprising 18 of the 26 letters of the half-uncial script used in *the Book*. These images were cropped letters from various folios of *the Book*. Some sample images are shown in figure 7. First we ran the model (described in section 4.2.2) on the entire set. Recognition accuracy is approximately 81%. Each result is the average of 10 iterations. For each iteration we choose



Figure 7: Some letters used for training and testing character recognition

7 of 10 images at random for training from each category, placing all remaining images in the test set. Features are learnt at random positions and scales from the training images. C2 vectors for the training set are built, and following this, the SVM is trained. Finally, C2 vectors are built and used to classify the test images.

It was observed that images that contained a non-standard form of a letter, such as the fourth image in figure 7, were often misclassified. The image data set for training and testing created by manually cropping letters from various folios of *the Book* indicates the need for an automated cropping mechanism if this is to be extended to a more comprehensive study, or applied to other manuscripts.

4.2.3 Feature extraction and hand classification

In this stage, we use 4 features to classify a character as belonging to a particular scribe: character width, character height, maximum brush width and minimum brush width. These features are extracted using distances between the opposing edges in a glyph. It may be noted that these are computed directly from the S1 layer, which is essentially a collection of edges occurring in the input image at different orientations and scales, using simple distance measures. The glyph dimensions are normalised by the average for each trained character. During training we identify the hands for which each character in the dataset as proposed in earlier work by Meehan. A two class training and testing (that is, hands are labelled '1' and '2' by the classifier corresponding the Meehan's hands 'A' and 'C' respectively) is performed using the listed features in addition to the recognised character as a feature. These features are weighted by a Support Vector Machine similar to the one used for the character classification in the previous phase. Further details in appendix A.3.

4.2.3.1 Results

The hand classification scheme was validated using a set of 54 images of characters (separate from the character recognition dataset) from two hands described above. The Support Vector Machine was trained and tested over 6 runs, in each of which, 16 randomly selected images were used for testing, and the remaining for training. Averaging across all the runs, approximately 39 images (72%) were identified correctly.

4.3 CONCLUSION AND FUTURE WORK

In art history, the definition of the style of a particular artist has mostly been attempted using dialogue, and only recently has there been effort in bring objectivity into such efforts. In the particular case of distinguishing the individual styles of the scribes in manuscripts such as *the Book*, the more forensic approaches to calligraphy analysis have been applied by many art historians. Some computational approaches to such analyses do significantly reduce manual effort required for such studies and at the same time, brings it into a measurable context. In this chapter, we apply a specific hierarchical object recognition model to this problem of classification, and build on features extracted from this hierarchy to demonstrate classification of characters by hand styles in *the Book*. The possibility of using locational proximity as a factor and other typographic features, such as heights of ascenders and descenders, and pen angle, are potential avenues of exploration for improvement of performance the methods discussed in this chapter. There is a lack of consensus on the number of scribes in *the Book*. While the use of unsupervised classification techniques could potentially infer an optimal number of hands based on some similarity measures, the stumbling block for such an approach remains the extraction and computation of a reasonable number of typographic features to make this approach viable.

INTERLACE

ABSTRACT Interlace is one of the most common of stylistic elements in traditional designs in Insular manuscripts such as *the Book*. In fact, it is abundantly found not only in manuscripts, but also in sculpture and architecture, and is found in articles of clothing and other accessories to this day. This chapter explores some digital methods that may be used to quantitatively analyse an interlace design. The goal is to provide some measures that allow more detailed and objective analysis of such designs with minimal manual effort on the part of an analysis expert. We also develop semi-heuristic measures that would aid art historians in analysing and comparing interlace quantitatively.

5.1 INTRODUCTION

Through Antiquity and the Middle ages, there was a lot of emphasis on the scribal traditions of calligraphy and illumination in manuscript production. Interlace is the most commonly used design element in a large number of artistic traditions, including Insular, Norse and Islamic. In its most basic form, it is highly regular and repetitive. John Romilly Allen (1847–1907) is often considered the one of the earliest prominent scholars of interlace. Born in London, Allen worked as a civil engineer in Edinburgh, where he developed an interest in early stone structures in Scotland, Wales, Britain and Ireland. He began to study the principles of interlace on stone, metalwork and manuscripts. He formulated a theory that the structure of interlaced ribbons was derived from ‘simple elements’ [1]. Five years later he made a two-way classification: *plaitwork* and *knotwork*, and introduced the concept of the *break*. Allen was also responsible for inspiring the work of George Bain (1881–1968). Bain declared his debt to Allen in the dedication to his best known work, the *Celtic Art: The Methods of Construction* [4]. Altogether over 225 different patterns are presented in this book, with hundreds of modification suggestions, 110 historical and modern artefacts showing designs in use, a great number of letters including six complete alphabets and 25 decorative initials, and a number of animal and human figures used in the original Celtic works. In the 1940s, towards the end of his teaching career he produced a set of tutorials for schools on how to draw the different categories of ornament in ‘Celtic Art’, with knotwork, spirals, key patterns, and zoomorphic designs among them. Bain’s zoomorphic reproductions, many of them excellently drawn on a magnified scale, carry occasional warnings about ‘errors’ or ‘faults’ in the interlace. His reconstruction of a faded 3cm wide sub-cruciform panel in the centre left of folio 33r in *the Book* is very impressive (see [1951, p. 107]). George Bain’s methods are often cited when theories about the historical methods for constructing interlace are discussed. This is due to his book’s own theoretical methods – not always easy to implement – and to its wide span of forms, as well as its usefulness in providing magnified copies of small elements of the illuminated manuscripts. Easier tutorials were to follow — George Bain’s son, Iain Bain [5] and Meehan [96] — and ‘Celtic knotwork’ tutorials continue to be published to the present day. Iain Bain published tutorials on Celtic Art in the late 1980s and believed that the setting-out lines and not the grid-lines should form part of the final strand. He states that Romilly Allen’s diagrams of the plait were just diagrams, not prescriptions for a construction method, and that the Celtic artists finished their lines freehand, and thus approximately – practising an ‘approximate geometry’ [5, pp. 7–18]. He discussed a technique, common in me-

dieval times, and known to him from engineering, involving using grids for transferring and re-scaling designs.

In Françoise Henry's works on the art of the early Christian period, her focus in the study of interlace was on zoomorphs and vegetative forms. She did not evolve a theory about interlace because her work on art had a wider context – provenances, influences, and chronology. But her illustrations, often hand-done, are an important resource for studying zoomorphic interlace, both in its style and construction. In the first volume of her series on Irish art [59, 60, 61] she gives a geometrical deconstruction of Celtic motifs in the pre-Christian and Christian eras (see [59, pp. 216–24]).

In 1993 an article by Cromwell [30] the grid based construction method was formalised into a pseudo-algorithm, which was later implemented in several popular software (including [97]). Derek Hull, an engineer interested in the patterns on materials, in the book *Celtic and Anglo-Saxon Art: Geometric Aspects* deals with methods of construction of interlace and with structural and 'combinatorial' properties of interlace and other motifs.

A cursory look at some wildly irregular designs (such as in figure 9b) leads us to the conclusion that grid based designing would be very simplistic. The more general representation would be to use a graph, and this is elaborated by Kaplan and Cohen [74], where an algorithm for construction of knotwork based on this representation is presented. Some non-grid based constructions that are thus made possible are illustrated in figure 10. Wild knotwork designs are compared to fractal knotwork designs and the traditional technique of N-interlacement. It is shown that all three styles may co-exist in one design. Browne [21] deals with the case of generative rules for "Wild knots" and fractal knots which are outside the scope of graph-based and grid based construction methods, while Dunham [38] applies similar concepts to designs in hyperbolic space.

Despite all this attention to knotwork construction, applied algorithmic and mathematical work has been restricted to the generative side, i.e. to methods for construction of interlace design and visualisation or rendering using computer graphics. It may be noted that so far, analysis and reproduction of interlace designs were done manually and with painstaking efforts. There has been a study on the complexity of manipulation of knots mentally by McLeay and Piggins [95], which concludes that visual analysis (comparison and classification) of knots becomes more complicated with an increase in the number of crossings. This suggests that manual inspection and detecting similarity between knots has limits on feasibility and effort.

As for the question of usage of mechanical aids in construction and execution of Celtic designs, recently Kulkarni and Stork [82] studied on the use of mechanical and optical aids in Robert Campin's Mérode Altarpiece. They conclude that Campin used a simple mechanical aid,

such as a minutely kinked straight edge or a mahl stick commonly used in the early Renaissance, rather than a very complex optical process using projection.

DISCUSSION Mathematicians, engineers, historians, archaeologists and recently computer scientists have delved into the area of understanding and reproducing interlace. Bain [4] among others outlined the methods of reproducing the style of interlace in *the Book* and Lindisfarne Gospels. These methods were essentially reverse engineered from the exemplars. Procedures for reproducing other design elements and some popular combinations, such as combining interlace with the art of illumination are also demonstrated by him. Given its regular nature, knot designs can be used as space-fillers, and as such are extensively found in manuscripts filling not only rectangular spaces but also curved ones, where a regular rectangular grid based design is stretched or otherwise transformed to fit the available space. It is notable that the approach taken to constructing interlace by Bain is based on grids. An instance of the procedure developed therein is illustrated in figure 8. However, the deconstruction and subsequent analysis of interlace has traditionally been the domain of the art historian. The same is true of key patterns, the computer-aided generation of which has received more interest (see [79]) than deconstruction. Computer vision has great potential to overcome some of the subjectivity in analysis of knotwork across manuscripts and sculpture, and cover the gap between the generative side and allow the designs that exist in the Book and other manuscripts to be seen in a new light.

In chapter 3, we discussed that the development of a unified approach in image processing is a necessity to tackle the wide variety of problems that are of interest to the art historian. In the previous chapter on character and hand analysis, we introduced a two-stage hand classification application for letters taken from *the Book*. In the following sections, we discuss the problem of analysis of interlace design in more detail, and propose a way to extract key points in interlace using a hierarchical object localisation model that is essentially the same as used in the previous chapter.

5.1.1 Questions

Art historians have a lot of questions about *the Book* and its construction. In the specific context of interlace, some of them are listed below.

- How much of it is hand drawn? What was the extent of usage of mechanical tools and optical aids?
Most of the past work in this direction has been limited by the accuracy of visual inspection, albeit of the trained art historian's eye. There is the question of subjectivity of the historian performing the analysis, which we hope to reduce or remove

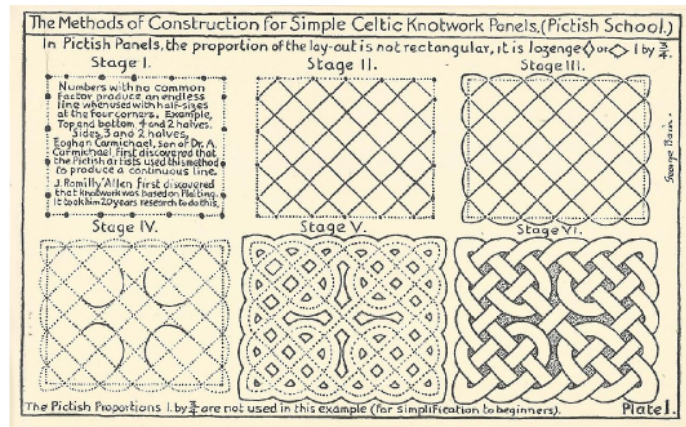


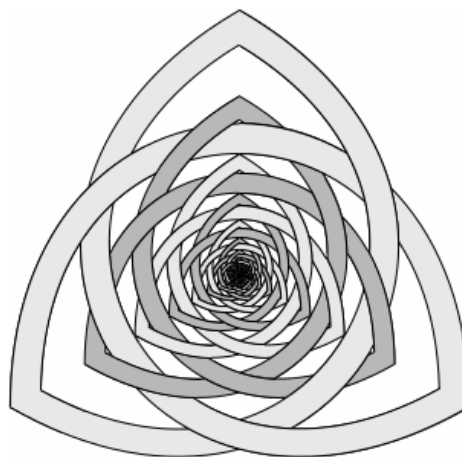
Figure 8: Constructing knotwork in stages by George Bain [4]



(a) Regular interlace design



(b) Irregular interlace design



(c) Recursive ("wild") interlace design [21]

Figure 9: Some interlace designs

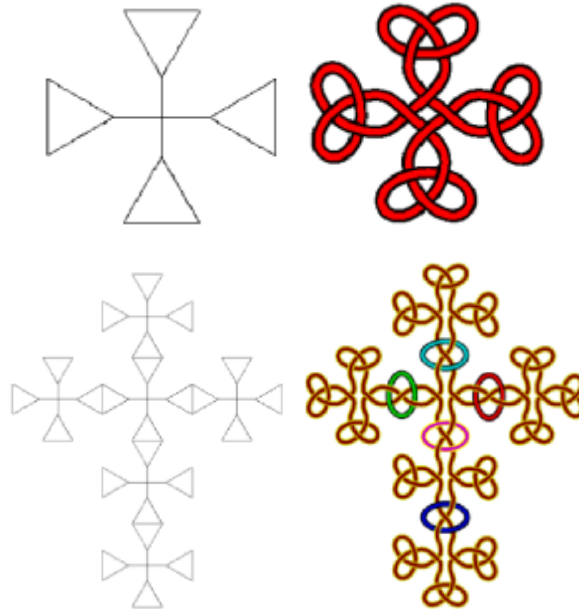


Figure 10: Graph based construction of interlace by Kaplan and Cohen [74]

by proposing pattern recognition approaches. An ideal solution would be to deconstruct designs in *the Book*, and model them using a graph based approach (as in Kaplan and Cohen [74]). Once such models are available for existing pieces of knotwork, some metrics of geometric accuracy of the constructions could be developed.

- What is the scale of ornamentation in the Book and how does it compare with other manuscripts?

The most challenging page in the Book, the Chi-Rho folio (33r) poses the particular question, is the scale of its design more minute than other folios, and similar designs in other manuscripts? Further to this, a measure of the scale of designs in different parts of a manuscript might be an indicator to the expertise of the artist, and would produce to evidence toward the identification of areas in a manuscript executed by the same artist. This also has some correlation with the accuracy of actual execution of design (outlined above), as below a minimal scale, mechanical tools would be essential to the process.

- How similar are the interlace designs in the Book and other manuscripts? Are there instances of reproduction of designs in one manuscript in others?

The phenomenon of copying a template design in later manuscripts or works of art is common. A measure of similarity between designs in two different manuscripts would aid determin-

ing causality in this aspect, and conceivably also refine dating of manuscripts.

We discuss a few proposed methods to answer these questions in the following sections.

5.1.2 *Nomenclature*

The basic terminology we use in this chapter in the context of interlace is listed below¹.

KNOTWORK OR INTERLACE A collection of threads that exist in three dimensions, but are constrained to a plane, in such a way that at various points on the plane, two of these threads cross each other in an interweaved fashion.

THREAD A strand of continuous material that has alternating crossings with other threads to form an interlace (or knotwork).

CROSSING Crossings are points on the plane where two threads cross each other. At such locations, neither of the threads in question actually lie on the plane any more: each lies on either side of it at a very small distance.

GRID A subset of possible interlace designs are possible to represent with a regular grid of points at which crossings occur in an interlace.

GRAPH Another subset of possible designs, larger than that mentioned above, may be represented as a pair of sets: vertices, and edges connecting these vertices. Each edge represents a crossing in the actual interlace. This is more general than a grid, as the vertices may be located anywhere on the plane.

BREAK This is a point on the plan where there a grid point exists, and as such a crossing may be expected, however no threads actually cross.

5.1.3 *Problem definition and scope*

We approach the analysis of interlace designs by modelling a interlace motifs using a grid. Crossings are the most naive and recognisable features an image that can lead us to reconstructing the grid. For convenience, we follow the convention of treating crossings as grid

¹ The nomenclature used by Kaplan and Cohen [74] is taken as a reference.

points, as opposed to edges (which is often used in models that represent interlace as graphs [74]). The measure of deviation of detected crossings from a grid would be a good indicator of the accuracy in execution of the design. Simultaneously, the location of crossing points would also provide a starting point for the comparison of potentially similar designs of interlace. Thus, locating interlace crossings (or a similar recurring pattern or anchor) seems to be the first step in a scheme that may provide some insight into the art historical questions put forth earlier (see 5.1.1). We enumerate some approaches that might be used to locate a repetitive structure such as a crossing in an interlace in section 5.1.4, and subsequently detail the method we use in our study in section 5.2.

5.1.4 Possible solutions

A range of image processing methodologies has been discussed earlier in sections 3.1 and 3.2. A couple of ways of looking at this problem include the following.

- Modelling segments of threads as active contours (see Kass et al. [75]), and reconstructing threads from these segments
Given the high level of degradation (of the inks used and the parchment) this method would no doubt have low accuracy. It also lacks the possibility of closely integrating with an intelligent degradation removal scheme. The iterative nature of degradation removal or reconstruction and its dependence on good feature extraction presents a stumbling block for such purely bottom-up (or top-down) approaches.
- Iterative interlace model fitting
A top-down approach to the problem at hand would be fitting a graph (or grid) model of interlace directly to the image using generative techniques. However, the parameters of any such model would be significant in number, and any iterative scheme to optimise these would rely heavily on heuristics, and thus, be very task specific.

As an alternative approach, the family of hierarchical architectures discussed in the chapter 3 seems to combine the best of top-down and bottom-up methods. The clear distinction between task independent and task specific learning in vision models of this family is one of the major arguments for favouring such an approach. The localisation property of some of the models (particularly [119, 117]) could be directly used in identifying locations of crossings in interlace designs. This family also has the added advantage of being amenable to modification for the specific task of degradation removal, that would significantly improve feature extraction, and also make possible performing degradation related corrections in the image domain (as op-

posed to the feature domain). Such robustness in interlace analysis to degradation found in manuscripts is discussed in chapter 6.

In this chapter, we use the localisation property of the model to elicit the structure of interlace designs and then turn our focus to the main contribution of this chapter: developing metrics for analysis of interlace based on features extracted from the model hierarchy.

5.2 LOCALISATION OF INTERLACE CROSSINGS

George Bain [4] often begins the construction of interlace designs using a grid of crossings (for instance the construction procedure outlined in [4, p. 29]). Crossing points provide reliably recurring and recognisable features that may be easily located in images and at the same time convey structural information essential to the design of a motif. As such the problem of finding such positions in an image may be treated as a particular application of the image processing problem of object localisation. In section 4.2.2, we described a hierarchical object recognition model proposed by Mutch and Lowe [99] as a basis for character recognition and further extended it to the problem of hand classification. Another known application of the same model is the localisation of objects of interest in images. In this case we use the model to locate interlace crossing points. The model specifications described in detail in appendix A.1.

5.3 METRICS FOR INTERLACE ANALYSIS

In this sub-section, we discuss the use of localisation of crossings in answering the questions of an art historical nature listed in section 5.1.1. As part of this study, three semi-heuristic quantitative metrics were developed for accuracy, similarity and scale of interlace designs, and are detailed below.

5.3.1 *Grid fitting*

After crossing points have been localised, a potential grid needs to be fitted to this set of points. We model the grid based on a parallelogram, so cases of stretched interlace can also be approximated, aside from rectangular grids. This model is given by

$$\vec{p}_g = m \cdot \vec{a} + n \cdot \vec{b} + \vec{c} \quad (1)$$

where $\vec{p}_g$ is an grid point, $\vec{c}$ is one of the grid points taken as a origin of the grid, $\vec{a}$ and $\vec{b}$ are two vectors representing the sides of the parallelogram, and $m, n \in \mathbb{Z}$. Figure 11 illustrates this model and grid points.

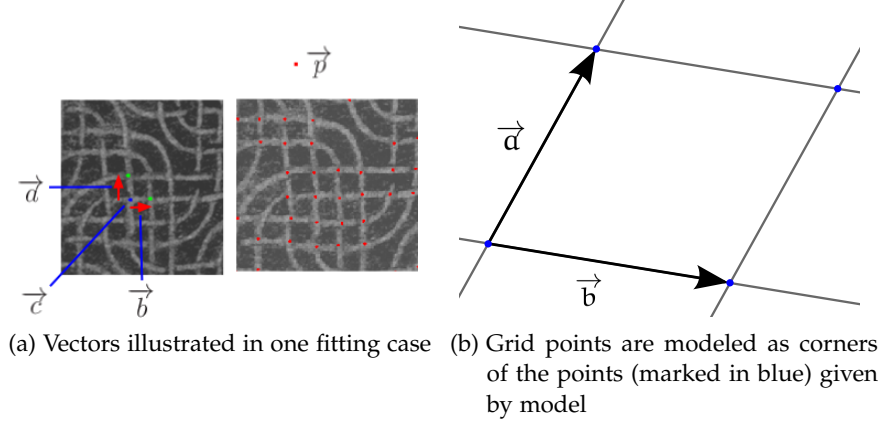


Figure 11: Parallelogram model for grid

We may use gradient descent to perform fitting. We start off fitting with initial estimates of $\vec{a}$, $\vec{b}$ and $\vec{c}$ from a user. These are improved upon in each iteration till no further decrease in error, defined in eqn. 2 below, is possible.

$$E = \sum_{\vec{p}} \min_{\vec{p}_g} \|\vec{p} - \vec{p}_g\|^2 \quad (2)$$

For each piece of interlace, we compute E , the mean squared distance between points on the grid and the extracted crossings. In general, this numerical value gives us a measure of how good the given model fits the data. This may be used to disambiguate between intentional and unintentional irregularities in the structure. Currently we do not penalise missed points explicitly, as there is the case that degradation the input image leads to mis-localisation or missed detection of feature points.

5.3.2 Mechanical or optical aids

In the specific case of knotwork, the (above) measures of fit may be used to quantify the extent of mechanical aids used in the construction. It is expected that any hand drawn piece of interlace (that is, a construction that uses no mechanical or optical aids of any kind) would have a natural deviation from being geometrically perfect. This deviation does have influence on visual aesthetics, which is also a part of the larger aims of this study. Conversely, use of mechanical tools during execution of the design would imply very low deviations from the grid.



(a) Corners of design in folio 27v



(b) Corners of design in folio 114v

Figure 12: Candidates for similarity computation

5.3.3 Scale

The scale of ornamentation in a given piece of interlace is computed by finding the average minimum distance between two extracted points of interest over this piece. This has implications in terms of determining complexity of design. For instance one argument is that other manuscripts (similar in style to *the Book*) usually have larger ornamentation, and such a statistical measure may be used as evidence to support claims on these lines.

5.3.4 Similar knotwork detection

In *the Book*, we have a few candidates for testing similarity metrics. Samples from two folios in particular are shown in figure 12. The computation of similarity between two designs is treated as a matching problem, and a simple heuristic such as a mean squared distance (i.e. averaged over the set of valid point pairs) between points of two given pieces of interlace is a good measure of dissimilarity between two sets of extracted feature points. Validity may be determined using scale computed earlier.

5.4 RESULTS

5.4.1 Extraction of crossings

The object localisation model described earlier in section 4.2.2 was trained and tested on a collection of images of interlace designs. For localization in images larger than 140×140 pixel we added a slid-

ing window functionality, where the full image is scanned vertically and horizontally through a window of size 140×140 pixels which moves in steps of 5 pixels in these directions exhaustively. Duplicate detections were consolidated using a neighbourhood suppression algorithm. The training and testing procedures used are detailed in the section 4.2.2.

DATASET Currently we use manually tagged crossings in 21 pieces of interlace design in *the Book* as the image dataset to be used for training and testing purposes. We note that we use 60% of the samples for training purposes. These images are cropped to 140×140 pixel images (2700 in total) each containing at least one crossing point, preferably near the centre of the image. This manual training is a required part of the localisation process. However, given that localisation quality is not good enough to perform statistical measures with the necessary level of accuracy, we do perform a manual clean up of the extracted points. Here, we correct missed detections and localisations. Both creation of the dataset and manual clean up are labour intensive, and are severely limiting factors in terms of the number of images that may be processed. Also it was noted that after the number of images in the dataset was increased over 2000, localisation accuracy was adversely affected, aside from increasing computation time from the order of 10s of minutes to hours for training alone. During testing of images, where crossings are localised, the detected locations within each cropped image must be converted back into location coordinates in the domain of the original interlace sample, to ensure consistency, and so that these can be used in computation of the proposed statistical metrics.

Results of localisation of crossings are shown in figures 16 and 17 (before manual clean up), and figures 19 and 18 (after manual clean up). On visual inspection, we note that localisation is not perfect. Aside from this, around 11% of the detected crossings are within 3 pixel radius of their expected locations, and a further 63% lie within a 5 pixel radius. We consider anything outside this as a missed detection or missed localisation. Approximately 31% of crossings are lost in the detection process: this is termed as missed detection.

As a reference, the same localisation technique applied on the UIUC car data set² had a recall at equal-error rates (recall = precision) of 99.9% (for the single-scale test set). In this case, scores were computed using the scoring programs provided with the UIUC data. To be considered correct, the detected position must lie inside an ellipse centred at the true position, having horizontal and vertical axes of 25 and 10 pixels respectively. It may be noted that the conditions for our

² A standard dataset for object recognition performance evaluation. Available at <https://cogcomp.cs.illinois.edu/Data/Car/>. It contains 1050 images: 550 images with side views of cars and 500 images containing no cars.

data set are significantly more stringent due to the small scale of images available: a typical car object in the UIUC image set covers areas of around 40×20 pixels, whereas a crossing point does not cover an area more than a few pixels in either direction.

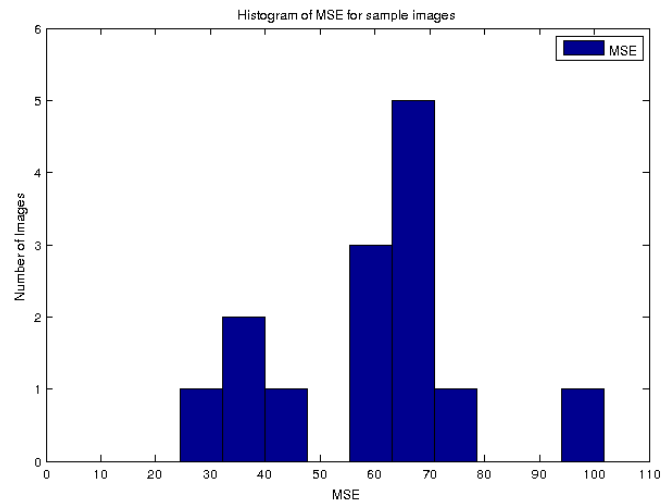
EFFECT OF DEGRADATION Evidently, degradation has a lot to contribute to missed detections and missed localisations. One of the reasons behind selecting this group of models is that some of these models were initially built for reconstruction of degraded text (see [48]). This points to the possibility of integration of these slightly different models to handle image noise, due to degradation resulting from ageing of the manuscripts. We discuss this in more detail in the following chapter, and propose measures to improve the performance of the above model therein.

5.4.2 *Statistical measures*

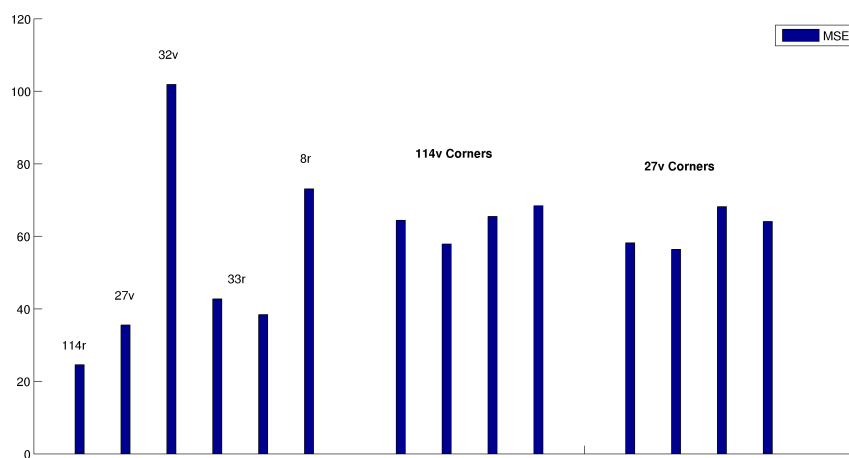
Given the quality of detection of crossing points, we perform some manual clean up of extracted crossings for each interlace. This includes repositioning detected points, and adding missed crossings. The ultimate goal would be complete elimination of this step. However, it is necessary for the time being, to ensure that the developed statistical measures are accurate and potentially allow hypotheses to operate independent of missed and mis-localised crossings.

GRID ESTIMATION We perform grid estimation (fitting) on each of the sets of extracted crossing points for the interlace designs in our dataset. We note the cases (figures 20a, 20g and 20f) that show the variation of the worst through the best grid fitting. In the foremost of these, the irregular grid makes a grid model a bad choice, and as such would be best represented as a graph. In comparison, in the last of these three, we note that possibly excepting the corners, the grid is a very good approximation for the design. We also note the fitting error values are 101.85, 73.05 and 35.58 respectively, which correspond to expectations from visual inspection.

MECHANICAL AND OPTICAL AIDS A plot of fitting error values for some interlace designs are shown in figure 13 and a corresponding histogram is shown in figure 13a. We note that the range 30-50 corresponds to an error of 4-5 pixels in radial distance per crossing, which may be considered a range for high accuracy in execution of the interlace design. At the opposite end of the spectrum, an absolute error 100 would correspond to ~ 7 pixel error per crossing, which can be attributed to human error and also points towards consistent deformation in the grid that suggests warping.



(a) Histogram of mean squared errors (MSE) for dataset



(b) Plot of mean squared errors

Figure 13: Results for mean squared errors over dataset

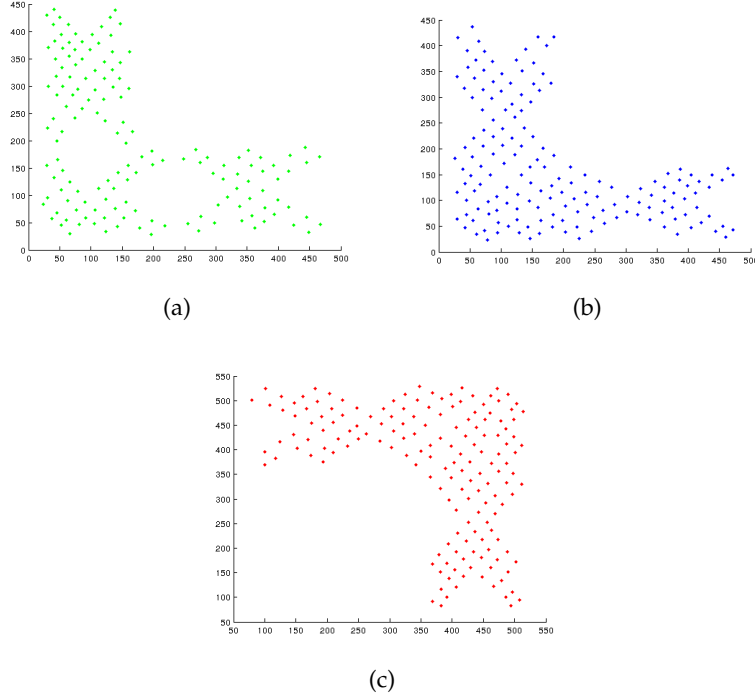


Figure 14: Registration

SIMILAR KNOTWORK DETECTION We perform registration of extracted (crossing) point sets from two potentially similar bits of interlace, as exemplified in figure 14. We treat plots of the extracted points as images, and register pairs of images with one of the potential images as reference. Once this is done, the transforms are applied to the sets of points, so now each point in either set has a potential nearest point in the other. This allows for a pairwise nearest neighbour squared distance evaluation that can be averaged over the number of valid pairs. Valid pairs are determined by ignoring nearest neighbours that are at a distance of over 2 times the average of scale of the two grids. The average similarity scores for the two sets in figure 12 are noted in table 1. From visual inspection, we note that the four designs from 114v are significantly different from each other, and those in 27v are more similar; this is reflected in the computed scores. Given the limited number of occurrences of highly similar knotwork designs in the images available from *the Book* for this study, testing the efficacy of this measure is difficult. However, it should be possible to extend the study to include other documented examples of interlace from similar manuscripts, subject to availability of high quality images.

SCALE The scale of various samples of interlace are plotted in figure 15. As such this data deserves comparison with interlace designs

Folio	MSE Scores	MSE averaged
27v	132.8381 162.6653 177.1295	157.5443
114v	228.8788 202.5366 277.2321	236.2158

Table 1: Similarity scores for interlace designs

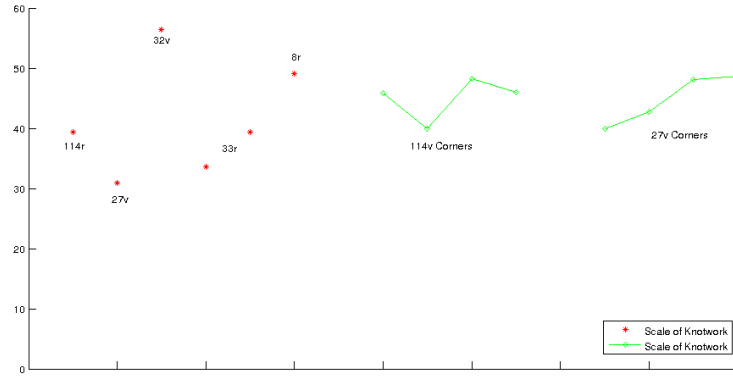


Figure 15: Scale of interlace design

in other manuscripts. Some early conclusions that may be drawn are that there is some sense of consistency in scale across folios, and even more so within folios.

5.5 CONCLUSION AND FUTURE WORK

There has been a significant amount of work in the domain of understanding the construction of interlace, and this study aims to complement these approaches by suggesting some methods to analyse existing knotwork in various manuscripts that use this element of design. Although using these methods, it is not yet possible to completely reconstruct interlace design starting from digitally scanned or photographed images of such manuscripts, the possibility of achieving some objectivity has been demonstrated. We rely on the localisation property of a hierarchical object recognition model proposed by Mutch and Lowe [98] to extract points of interest in images containing interlace, and use developed heuristic metrics to make possible comparison and measurement of accuracy of design execution. This study is limited in scope to grid based designs, and the possibility of extending this to a more general graph based representation has also been discussed as a future possibility. The detection of warping over large areas of parchment are beyond the scope of current work. However, it should be possible to model parchment warping as local affine transforms, and such a fitting would definitely take us a step closer to analysis of such occurrences. Its analysis may also benefit from visu-

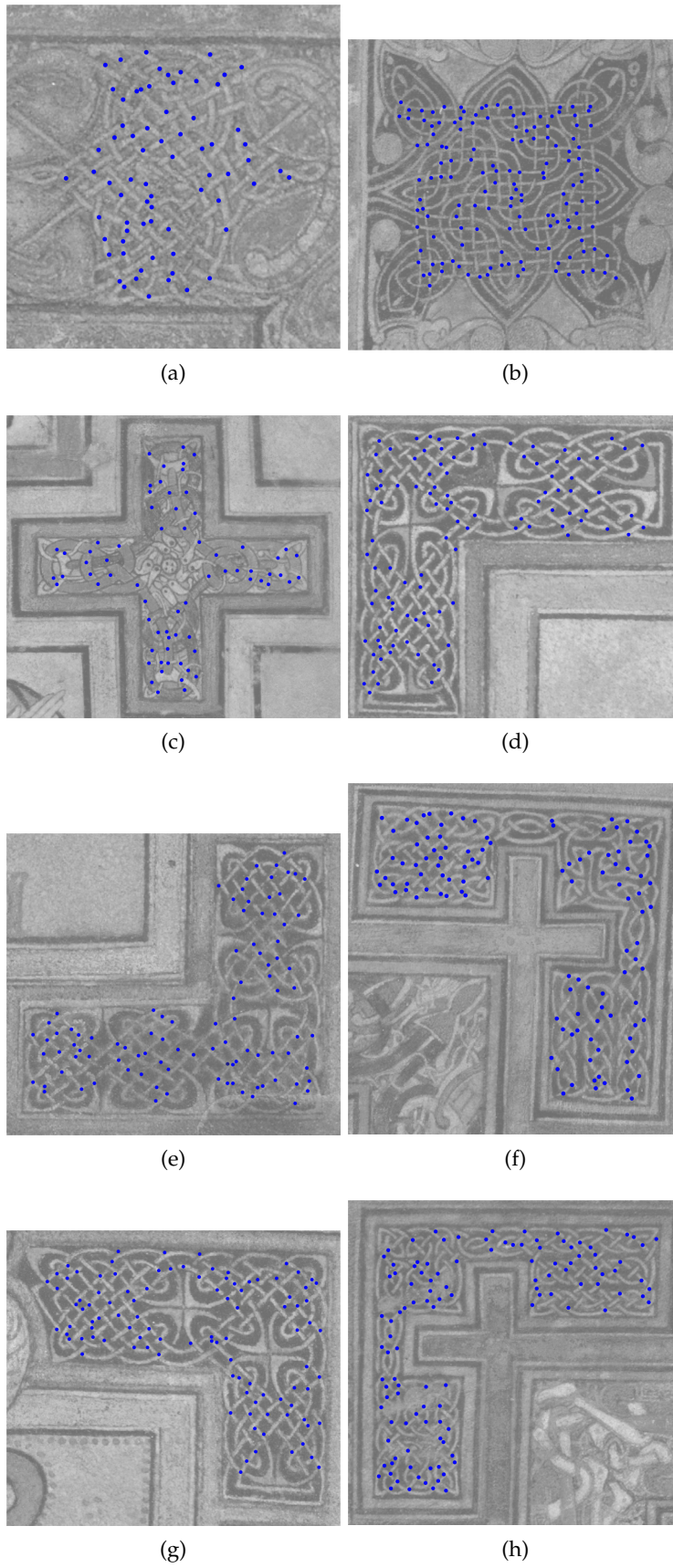


Figure 16: Extraction of crossings. Extracted points are marked in blue.

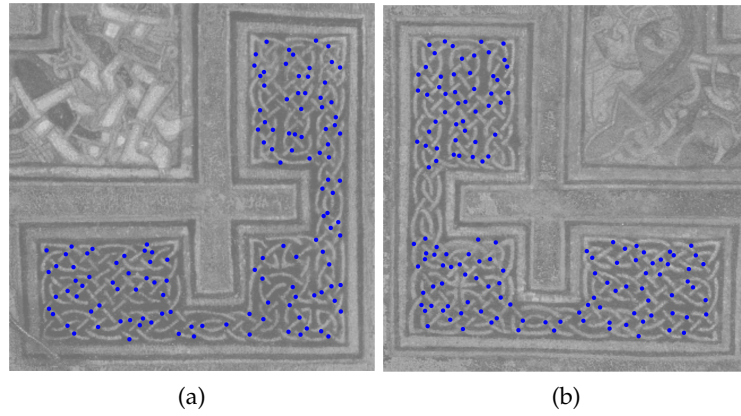


Figure 17: Extraction of crossings (contd.) Extracted points are marked in blue.

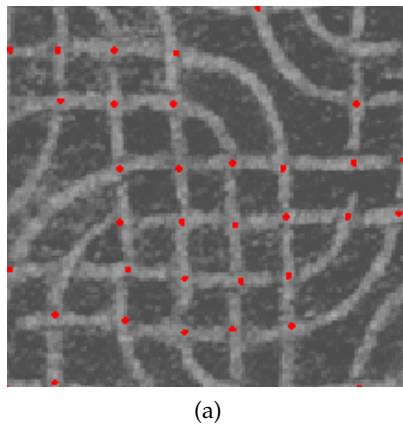


Figure 18: Extraction of crossings. Crossing points are marked in red.

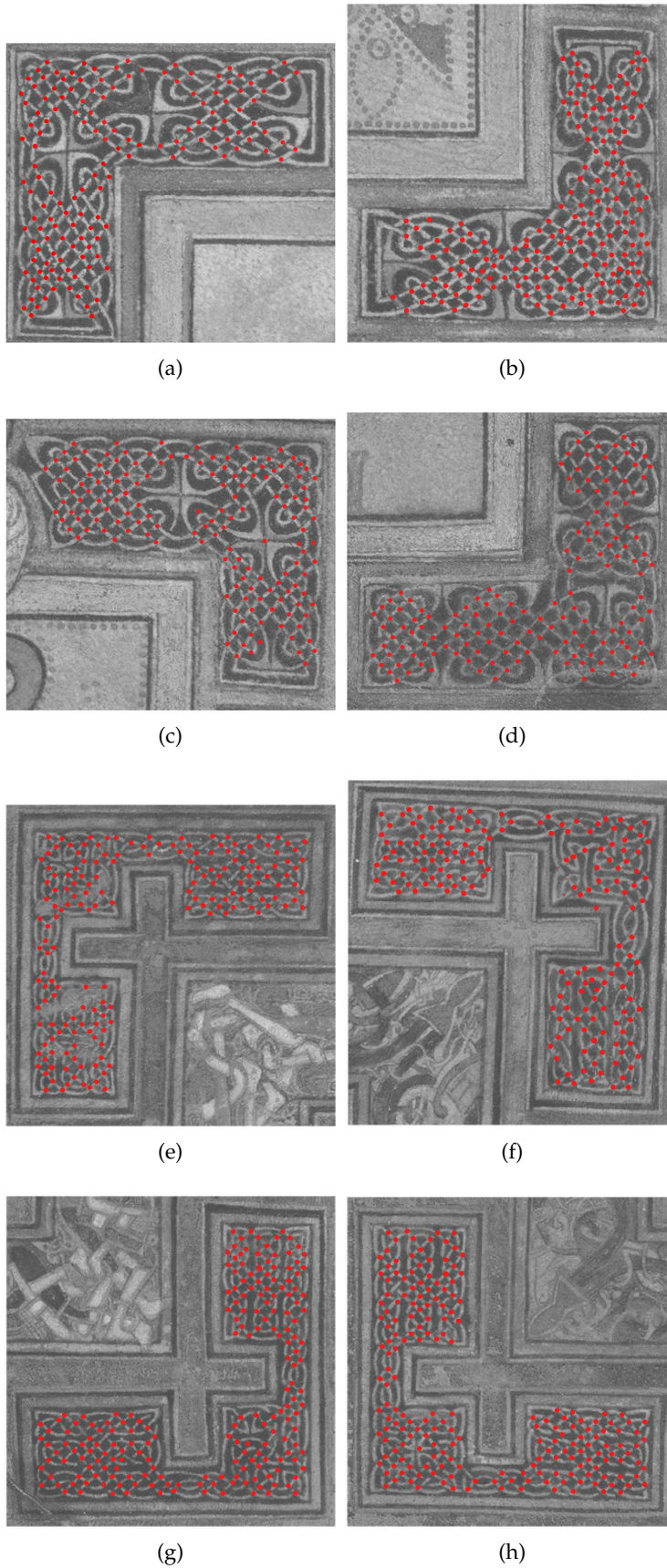


Figure 19: Extraction of crossings (contd.). Crossing points are marked in red.

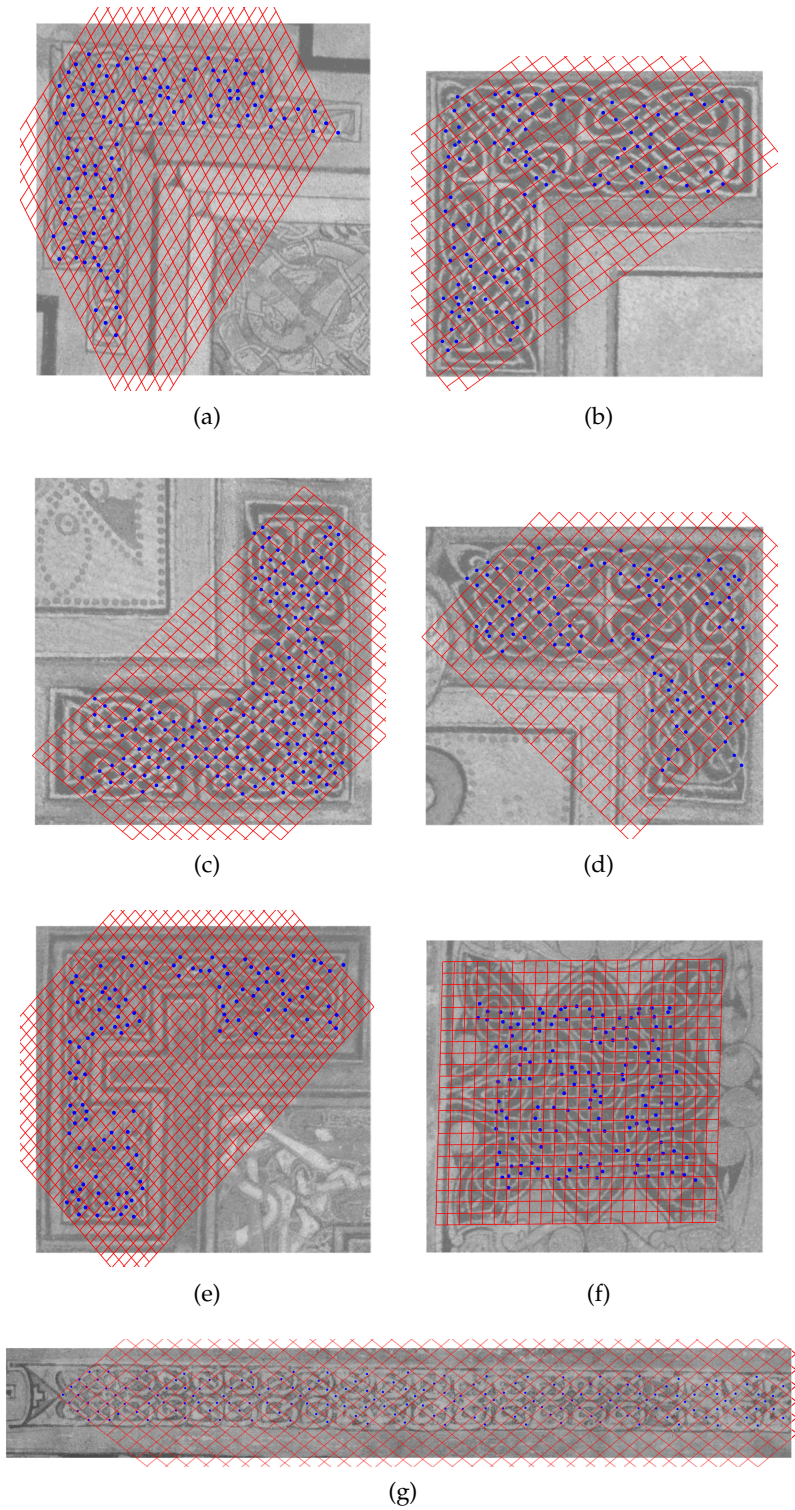


Figure 20: Some grid fitting results. Corrected extracted grid points are marked in blue, and fitted grid is depicted red.

alisation, that shows a deviation map of points from their respective grid points. Another application of very closely related object recognition models that may serve as a degradation removal engine, and thus improve the locational accuracy is discussed in the following chapter.

LEARNED DEGRADATION REMOVAL

ABSTRACT In this chapter, we discuss some of the latest developments in the field of image restoration applied to the domain of manuscript analysis and their limitations. Typically the various types of degradation found in manuscripts have been approached using methods that are aware of the nature of the causative processes. Following in the theme of this thesis, we propose a hierarchical model that extends the model used in the previous chapters to the domain of image reconstruction, more specifically, degradation removal for manuscript images. The proposed model is used to improve the accuracy of the model used in the previous chapter to detect interlace points. It is demonstrated that such an approach not only achieves this goal, but also outperforms a naïve image reconstruction scheme and simultaneously allows us to partially restore interlace design in the image domain.

6.1 INTRODUCTION

Traditionally restoration of degraded art work has been the domain of the expert conservationist. Many cultural works are sensitive to environmental conditions such as temperature, humidity and exposure to light, especially ultraviolet light. The most important sources of degradation are — chemical interaction of the media (paints, pigments, inks, etc. the support base, parchment, canvas, wood, etc.), photodegradation (interaction of the object with incident light), physical contact with other materials (interaction with chemically active materials or organic materials). The function of a conservation expert is to nullify or at least reduce the rate of deterioration of an object. This can be achieved through either non-interventive or interventive methodologies. Interventive conservation refers to any act by a conservation expert that involves a direct interaction between the conservator and the cultural material, and could involve cleaning, stabilizing, repair, or even replacement of parts of the original object. For paintings and manuscripts, steps such as surface cleaning by varnish removal, or consolidation by securing flaking paint are common practices. Such interventive actions are carried out for a variety of reasons including aesthetic choices, stabilization needs for structural integrity, or for cultural requirements for intangible continuity. One of the guiding principles of conservation of cultural heritage has traditionally been the idea of reversibility, that all interventions with the object should be fully reversible and that the object should be able to be returned to the state in which it was prior to the conservator's intervention. Although this concept remains a guiding principle of the profession, it has been widely critiqued within the conservation profession and is now considered by many to be "a fuzzy concept." Another important principle of conservation is that all alterations should be well documented and should be clearly distinguishable from the original object. Conservators routinely use chemical and scientific analysis for the examination and treatment of cultural works. The modern conservation laboratory uses non-invasive equipment such as microscopes, as well as invasive techniques such as the use of spectrometers and various x-ray regime instruments to better understand objects and their components. The data thus collected helps in deciding the conservation treatments to be applied to the art work in question.

Imaging science and image processing technologies offer an orthogonal avenue of exploration of restoration of digitised cultural artworks. As tools for artwork restoration, image-processing techniques could serve two purposes. They can be used as a guide to the actual restoration of the artwork (computer-guided restoration). Alternatively, they can produce a digitally restored version of the work, which itself is valuable, although the restoration is only virtual and cannot be reproduced on the real piece of work (virtual restoration).

In either type of application, since the only interaction of the digital techniques and the art object is at the digitisation stage, both guidelines of conservation are adhered to. Barni et al. [7] propose two digital image processing applications for restoring artworks. The first belongs to the class of computer-guided restoration techniques, and the second represents an example of virtual artwork restoration. The first application attempts to predict the final result on an artwork by learning from the application of a particular cleaning methodology to a small section of the artwork. Restorers can use it by applying a set of different cleaning methodologies to very small patches of the painting. This affords art restorers to evaluate different degradation reversal techniques for best results using digital image processing as a tool to guide the actual restoration of the artwork. The second application is a virtual system for removing ‘craquelure’, i.e. cracks in the paint resulting from drying, from old paintings and frescos — an attempt to remove cracks that severely deteriorate the aesthetic appeal of paintings. In a similar vein, Mastio et al. [93] discuss a novel method which is able to assess the original beauty of digital reproductions of aged photos, as well as digital reproductions of faded goods. The method is based on the comparison of the degraded image with a non-degraded one showing similar contents; thus, colours and colour of the non-degraded image can be transplanted in the degraded one. The key idea is a dualism between the analytical mechanics and the colour theory. Berns [12] describes an approach that simulates of a work of art before the colour change and demonstrates its use on paintings by Vincent van Gogh and Georges Seurat. The method relies on the use of spectral or colorimetric imaging and optical models such as Kubelka-Munk turbid-media theory to emulate the optical behaviour of the artist’s materials.

Wei et al. [134] propose a novel scheme to virtually restore the colours of ancient murals. This approach integrates artificial intelligence techniques with digital image processing methods. The knowledge related to the mural colours is categorized into four types. The proximity in colour-space and location is used as a factor in segmentation of mural images, while the actual restoration of the colours is performed via transformation of colour histograms. Pappas and Pitas [105] presents five digital colour restoration techniques, which can be used to simulate the original appearance of paintings. Most of the work in colour restoration of manuscript images is focused on the Red-Green-Blue (RGB) colour space. Recently, Tonazzini [128] demonstrated that representing RGB images in different colour spaces can be more effective for this goal. Aiming to maximise the information content of the colour space used, he shows that some particular colour spaces are more suitable for the analysis of degraded documents, allowing for the enhancement of the contents, the improvement of the text readability, the extraction of partially hidden features, a better

performance of thresholding techniques for text segmentation, object segmentation and edge detection. In the area of segmentation, Leydier et al. [89] developed an adaptive algorithm for the segmentation of colour images suited for document image analysis. The unsupervised algorithm reuses information about the computed colour clusters, learns from previous colour statistics and adapts the classifier to any new local palette modification.

Another common research problem in the field of digital analysis of manuscripts and other historical documents is bleed through removal. This phenomenon occurs due to the seepage of ink from one side of a document to the reverse side, resulting in a pattern that interferes with the text or illustration on the latter side. In the last decade there has been an increasing momentum in this research area which, most commonly, takes a machine learning approach. Tonazzini et al. [129] has proposed a colour decorrelation strategy to improve the human or automatic readability of bleed-through degraded documents. A simplified linear model for this degradation is introduced to permit the application of very fast decorrelation techniques colour data, and to compare this strategy to the independent component analysis approach. Tonazzini also discusses the possibility of further improvement by using hyper-spectral and multi-spectral data. Rowley-Brooke and Kokaram [113] present a practical Bayesian approach to the problem, building a computationally inexpensive approach also starting with a linear degradation model.

Some other noteworthy problems are those of de-warping, in the case of heat or moisture based warping of parchments. The degradation of iron gall ink on different parchment materials including vellum (as is the case with *the Book*) has received attention for a long time, though only recently, the chemical processes behind it and consequent physical changes are being understood in detail. The implications of such research on digital degradation removal methods remains to be explored.. Some other topics that have received minor attention include image reconstruction via template formation and classification and model based regeneration.

6.1.1 Questions

The most important motivation for the work described in this chapter is to be able to digitally restore *the Book* to its original state free from deterioration it has undergone over the centuries. On the other hand, a specific question in art history that is dependent on being able to measure degradation in some way, is about the distribution of attention given by viewers and handlers to the various folios. For instance, a large part of the deterioration of the pigments in *the Book* can be attributed to physical handling, specifically, touching the parchment while reading, turning folios or as part of prayer. A measure of the

regions of the folios with highest degradation may provide some insight into which pages were accessed most frequently. A similar question on the accessibility of *the Book* to the general public could also benefit from a statistical comparison of its degradation with that of other manuscripts or intelligent model-based estimates.

6.1.2 *Problem definition and scope*

Degradation stemming from various sources is ultimately manifested in the digital image as a loss in quality that is often modelled by the umbrella term ‘noise’. Image restoration is the specialised branch of image processing that aims to remove the effects of noise in images. In the previous chapter, our analysis of interlace was based on an object localisation framework applied to the task of locating interlace crossings. It was observed that the large amount of degradation present in the images affected the quality of these detections adversely. In continuation with the study, the topic examined in this chapter is the use of image reconstruction to improve the accuracy of the model, with the goal of increasing detected crossings in interlace designs.

6.1.3 *Possible solutions*

A majority of image reconstruction methods assume a basic to detailed knowledge of the processes that cause the visual artefacts that are to be removed, for instance noise or process modelling as seen in [105, 129, 113]. It is important to note that the human visual system learns the structure and appearance of objects in nature by example, and is able to detect objects even in the presence of large amount of distortion, occlusions and even artificial noise. As previously discussed in chapter 3, Fukushima [46, 47, 48] developed a hierarchical neural network that can perform character recognition and simultaneously restore object shapes in images in the presence of other occluding objects. This multi-layered hierarchical neural network processes achieves this effect by interaction of bottom-up and top-down signals. Occluded parts of a pattern are restored mainly by feedback signals from the highest stage of the network, while the unoccluded parts are reproduced mainly by signals from lower stages. It is important to note the parallel between the nature of occluded images and loss of parts of text and interlace due to degradation. While the problem of degradation removal may be attempted via standard approaches such as [72] (bottom-up) or more degradation-process aware methods discussed above, an approach such as that taken by Fukushima allows for the possibility of recognition to function in tandem with degradation removal. It also opens up the possibility of developing more complicated art historical applications that rely on degradation removal as a precursor, such as assessing aesthetics in absence of degradation.

Consequently, and in keeping with the theme of this thesis, we approach the problem of reconstruction using iterative refinement and lateral interactions, which are features of a few hierarchical models as explained below. The detrimental effect of image noise in analysis of interlace clearly has to gain by taking the latter route. In the following sections, we develop a model that applies Fukushima's combination of top-down and bottom-up interactions to the (purely) bottom-up model we used in the previous two chapters. The resulting model is used to improve the performance of the interlace analysis application in chapter 6.

6.2 DEGRADATION REMOVAL FOR INTERLACE

The occlusion reconstruction model proposed by Fukushima [48] is a multi-layered hierarchical neural network that extends an object recognition model to allow reconstruction in the image domain for occluded portions of an image. It learns features from a set of template images, and tries to reconstruct missing bits in images that are plagued by occlusions. The main feature that sets this work apart from its siblings in the family of hierarchical image processing models, is that it uses both bottom-up and top-down processing. This is achieved as by mirroring the bottom-up network, and allowing it to learn top-down influences as well. The model proposed by Fukushima is shown in figure 35. It may be observed that occlusions are a specific kind of undesirable artefact in images. In the domain of manuscript analysis, degradation by loss of pigment, etc. is the undesirable artefact that we hope to reverse. With this in mind, we attempt to combine the NeoCognitron approach to reconstruction with the localisation and recognition efficiency of the model described in the section 4.2.2. The proposed architecture is illustrated in figure 21 and discussed below.

The proposed network is composed of alternating layers of "S" (simple) and "C" (complex) cells: layers for image, Gabor filter (S1), local invariance layer (C1), intermediate features (S2) and finally a global invariance layer (C2), as described in sections 4.2.1. In addition to the aforementioned layers, we compute two new layers S3 and C3 that function as S2 and C2 do. The mirroring in Fukushima's network allowed learned features in the top layers to influence lower layers and make corrections iteratively, ultimately resulting in reconstructed image without occlusions. In the model we propose, the equivalent inhibitory feedback connections are added to the object recognition and localisation model to account for such top-down effects. As illustrated in figure 21, connections from each C layer is propagated using inhibitory and excitatory connections to itself and any S layers immediately up the hierarchy from it. Also, similar influences are provided from each S layer to the adjacent C layers and itself. The output layer

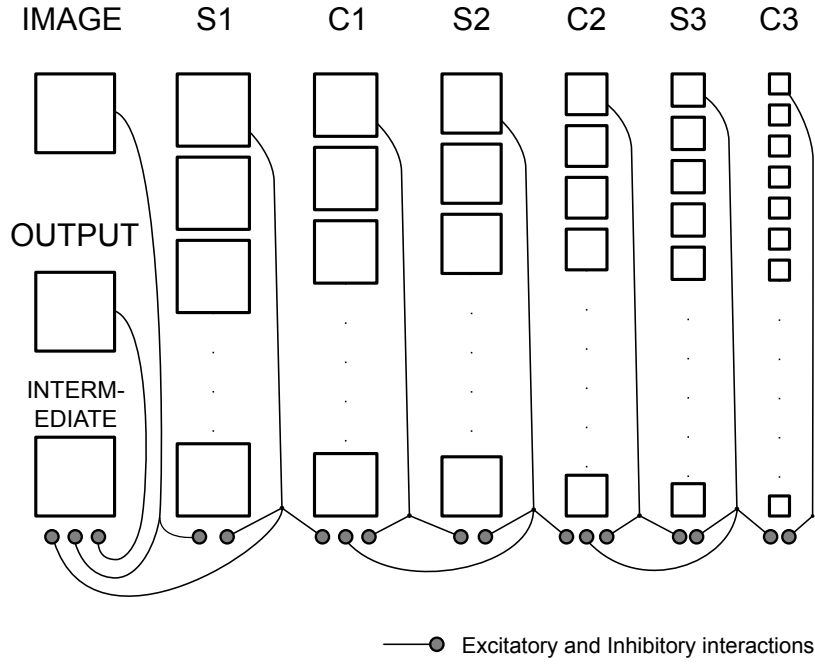


Figure 21: Proposed degradation removal network

is also connected to an intermediate layer via inhibitory and excitatory connections. The intermediate layer is similarly connected to the S1 layer and the input image layer as well. The connections detailed above serve to provide the lateral and top-down influences needed for the final reconstructed image to be available at the output layer. Some other functional aspects of the model are described in appendix A.2.

We evaluate this architecture for the particular application of interlace analysis that was discussed in the previous chapter. An interlace crossing is a the repeated pattern that might be used for template formation, and may be reconstructed in the case of (literally) misplaced crossing points¹ or broken designs where part of the interlace is missing. The goal is to learn the specific features needed to faithfully replicate a crossing at an appropriate place, make corrections to the location if necessary and potentially place crossings at portions of the parchment where a part of an interlace design is missing or noisy.

6.2.1 Implementation

The proposed model operates in two stages. In the first stage, we simply attempt to train the model described above for degradation removal in the same way training for crossing localisation was performed in the previous chapter. Stage 1 learns the features that define a crossing point in terms of morphology and structural features. During this phase, the model also learns to predict the location of cross-

¹ For a discussion on crossing points, refer to section 5.1.2.

ing points in interlace motifs. Training of the forward connections of the network is performed from the lower stages to the higher stages: after the training of a lower stage has been completely finished, the training of the succeeding stage begins. During the training of the connections of the forward paths, responses of the cells of the backward paths are all suppressed. In the second stage, we proceed to the testing phase: degraded portions of interlace were also presented for localisation. the localisation of interlace crossings (discussed in the previous chapter), but using this degradation trained model.

We use the Cortical Network Simulator (CNS) package for MATLAB. The advantage of using a framework such as CNS is that it allows use of Graphics Processing Unit (GPU) cards present in most modern computing machines and leverage their parallel computation capabilities, which significantly reduces processing time for training and evaluation of our hierarchical models. The CNS is designed to build cortically organised models of vision for varied applications. For instance, it has been used for implementations of Convolution Networks, some older HMAX models for object recognition.

6.2.2 *Training and testing*

As in the previous chapter, the data set for training uses manually tagged crossings in 21 pieces of interlace design in *the Book* from various folios. Training is performed on 60% of the images, and the remaining 40% are used in testing the performance of the approach. Special attention is devoted to making sure there are enough noisy (degraded) images in the dataset. This ensures that the repeated pattern that the neural network tries to learn about is a crossing, and not an interlace design in its entirety. The requirement of the network to learn how to predict the location of the knotwork implies that there should be a layer in the network to learn features that use the relative locations of interlace crossing points and learns patterns within such data.

We test the localisation performance of the proposed model by comparing the number of missed localisations in this approach with the model used in the previous chapter. We also demonstrate the viability of the degradation removal by comparing the output of the top-down half of the network with the input image.

6.2.3 *Results and discussion*

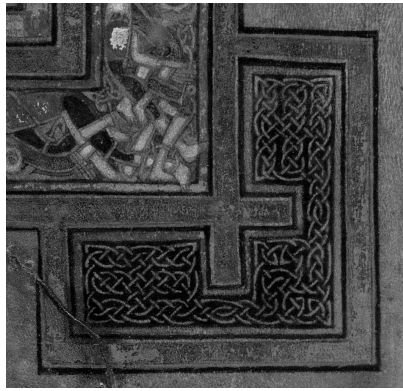
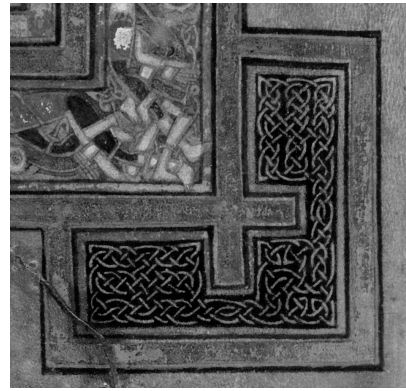
Degradation removal using the proposed reconstruction procedure was attempted for images with a significant amount of degradation, specifically interlace pattern loss. The results are shown in figure 22. In a comparison with a Convolutional Neural Network (CNN), which has been used for de-noising and degradation removal for images

previously [71, 72], visual inspection suggests that this approach is a significant improvement over a CNN based de-noising technique [72]. The results are shown in higher detail in figures 22d, 22e and 22f. The proposed degradation removal scheme shows a high degree of selectivity in reconstruction: specifically it exhibits contrast enhancement for the parts of the image containing interlace. Secondly, it is also observed that while the CNN based approach does remove speckle-type noise, the top left of the interlace design in figure 22f is visibly as degraded as in the input image, figure 22d. In comparison, the interlace strands at the same location in figure 22e are conspicuous. This behaviour was also verified in stage 2 of the implementation, missed interlace crossings decreased from 30% in the naive version (chapter 5) to $\sim 28\%$ using CNN and down to $\sim 23\%$ with the proposed degradation removal model. The CNN removed parchment textural noise, however was unable to improve location of crossing points (compared with the naive crossing point detection) by a significant amount. This demonstrates that learning structural features from the hierarchy contributes significantly to decrease in localisation errors.

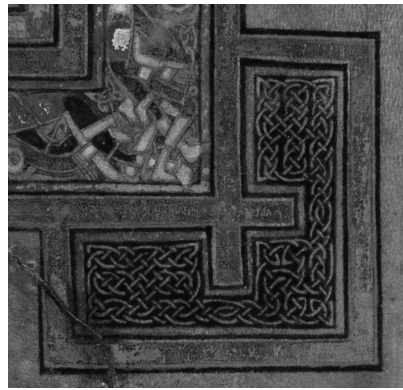
6.3 CONCLUSION AND FUTURE WORK

The physical and chemical processes that cause cracks in and loss of pigments on manuscripts are a major hindrance to robust image processing applications that aim to analyse the design and aesthetics of such works of art. Image segmentation and feature extraction in such applications are affected adversely by degradation. Although colour space transforms have been applied to circumvent some of these issues, a more morphological approach as detailed in this chapter provides an alternative trajectory for manuscript analysis. It is important to note that the model used in this chapter is an expansion of the model used earlier. Attacking the problems of degradation and classification simultaneously is an important step in moving closer to the way in which humans perceive images. The application of this model to other art historical problems promises a way to allow answering a wider variety of questions in the presence of degradation, whatever be the cause.

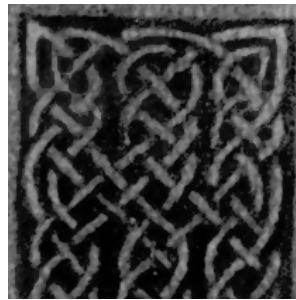
In the following chapter, we discuss some minor applications of the localisation and classification model.

(a) Interlace design from *the Book 27v*

(b) Degradation removal result by proposed model



(c) Result of denoising using Convolution Neural Network

(d) Detail from the interlace design from *the Book 27v*

(e) Detail from the degradation removal result by proposed model



(f) Detail from the result of denoising using Convolution Neural Network

Figure 22: Degradation removal results

MISCELLANEOUS APPLICATIONS

ABSTRACT In this chapter we discuss some novel applications as minor extensions of the image hierarchy discussed thus far. The analysis of spirals in *the Book* has attracted some attention in relation to the issue of use of mechanical tools for their construction. The use of specialised rotational symmetry detection layers as a way to quantify the precision of spiral construction is discussed as a quantitative measure in this context. As another application, the object localisation model is used to locate faces in folios at differing scales, and serves as a precursor for the next chapter. This chapter also introduces a basic clustering approach to grouping characters into lines and paragraphs as a demonstration of the customisability of our hierarchical architecture.

7.1 ACCURACY OF SPIRAL DESIGNS

While interlace is one of the most frequently used design element in the Insular tradition, spirals and elements with rotational symmetry are also extensively used in manuscripts, sculpture and other art forms. In chapter 5 the art historical inquiry into the extent of use of mechanical and optical devices was introduced in the context of the construction of interlace designs. In asking the same question of spirals, we find that in comparison with interlace, the spiral is found in a wider variety of forms and is often integrated into more complicated designs. However, the characteristic feature of rotational symmetry is often adhered to.

In this section, we discuss a technique to assess the aspects of symmetry in spirals and propose a rotational symmetry model that subscribes to a hierarchical architecture. Its application to measuring the symmetry of some common spiral forms in *the Book* is demonstrated as a replicable method to find perfect rotational symmetry and measure the deviation from mathematically or constructionally accurate spirals.

7.1.1 Rotational symmetry in images

Symmetry is an important element of perceptual grouping. The reason is that image structures are unlikely to be arranged in such a manner as to produce spatial symmetry by chance. It is much more likely that if a group of image structures exhibits symmetry, then they are physically related, and so are perceived in relation to one another (i.e., they are grouped together perceptually). This has motivated research on the detection of symmetries in images. There are three types of symmetries. If a figure is invariant under a reflection about a line, then it is said to possess bilateral symmetry. A figure is said to possess rotational symmetry of order n , if it is invariant under rotations of $\frac{2\pi}{n}$ radians about its centre of mass, and is said to possess radial symmetry if it has both bilateral and rotational symmetry.

In nature, perfect symmetry is very rare. For an artist, construction of designs with perfect symmetry would require an immense amount of practice. Furthermore hand eye coordination is known to deteriorate with decrease in scale of design. From a visual inspection of some of the spirals constructed in *the Book*, it is evident that this observation holds true. As detailed in section 5.1.1, the usage of mechanical or optical tools in the construction of designs has remained an open question to date. In the specific case of spirals, quantification of the closeness to perfect symmetry would provide valuable evidence to support such studies. The problem of detection of symmetry in images has been attempted by image distance metrics [141], gradient vector fields [108] and neural networks [49, 83]. While the earlier ap-

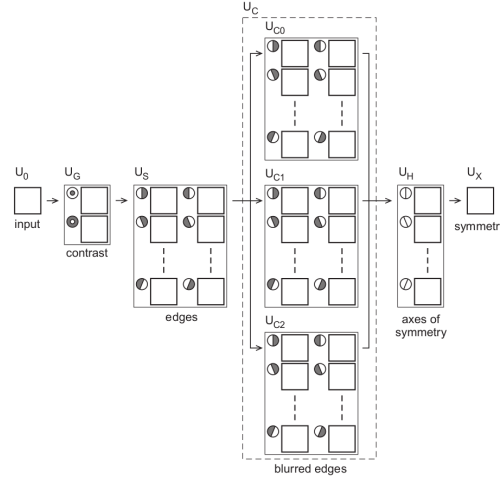


Figure 23: Symmetry axis detection model by Fukushima and Kikuchi [49]

proaches are more morphological and iterative in nature, Kunihiro and Fukushima [83] propose the application of a blur operative hierarchical network to this problem. This approach is further refined in [49] and draws its hierarchical structure from Fukushima's earlier NeoCognitron [83, 49]. This resemblance to the hierarchical models and its robustness to noise and deformations in the input image make it the approach of choice to measure the symmetry of spiral designs. The model is described in detail in the following section.

7.1.2 Symmetry axes extraction network

Fukushima's model is a multi-layered network with a hierarchical architecture, which is illustrated in figure 23. It resembles the architecture of the lower stages of the NeoCognitron [45]. Each layer consists of a number of cell-planes. Incidentally, a cell-plane is a group of cells that are arranged retino-topically and share the same set of input connections. As a result, they all have receptive fields of an identical characteristic, but the locations of the receptive fields differ from cell to cell. The model consists of a number of layers connected in an hierarchical manner: an input layer (U_0), a contrast extracting layer (U_G), an edge-extracting layer of a simple-type (U_S), an edge-extracting layer of a complex type (U_C), a symmetry-axis-extracting layer (U_H), and a symmetry-extracting layer (U_X).

The stimulus pattern is presented to input layer U_0 , which consists of photoreceptor cells. The output of layer U_0 is fed to contrast-extracting layer U_G . Cells of contrast-extracting layer U_G resemble retinal ganglion cells or lateral geniculate nucleus cells. Layer U_G consists of two cell-planes: one with concentric on-center receptive fields, and the other with off-center receptive fields. The former cells extract positive contrast in brightness, whereas the latter extract negative con-

trast from the image presented to the input layer. To be more specific, the spatial distribution of the input connections to an on-center cell take the shape of a Mexican hat, and off-center cells have connections of opposite polarities. These connections are so designed that the total sum of the connections converging to a single cell is equal to zero. This means that the DC component of spatial frequency of the input pattern is eliminated in the contrast-extracting layer U_G . As a result, the output from layer U_G is zero in the area where the brightness of the input pattern is flat.

The output of layer U_G is sent to an edge-extracting layer (of a simple-type) U_S . Layer U_S consists of S-cells, which resemble simple cells in the primary visual cortex. Each S-cell receives connections from both on- and off- center cells of U_G and extracts edges of a particular orientation. Layer U_S thus decomposes the input image into edge components of various orientations. The input connections to S-cells are determined with the same process as the supervised learning for the NeoCognitron.

BLURRING BY C-CELL LAYERS The output of layer U_S is fed to layer U_C (edge-extracting layer of complex-type), where a blurred version of the response of layer U_S is generated. Layer U_C consists of C-cells, which resemble complex cells in the primary visual cortex. Connections from S-cells to C-cells resemble the classical hypothesis by Hubel and Wiesel [63]. Each C-cell integrates the responses of a group of S-cells of the corresponding cell-planes within its receptive field. A blurred version of the response of layer U_S thus appears in layer U_C .

EXTRACTION OF AXES OF SYMMETRY In the proposed model, symmetry axis extraction layer U_H follows layer U_C . Each cell of layer U_H analyses the response of U_C and calculates H , which is the weighted sum of h (the degree of symmetry) for different edge-orientations and for different resolutions, and checks if its receptive-field centre is on an axis of symmetry. Each cell in a cell-plane checks if its receptive-field centre is on an axis of symmetry of a particular orientation. Finally, the symmetry extraction layer U_X , which consists of only one cell-plane, integrates the responses of all cell-planes of layer U_H . Layer U_X responds at locations of the axes of symmetry of the input image independently of their orientations.

MODIFICATIONS The network described above was designed to process images with bilateral symmetry in a robust fashion. However, the application to the measurement of rotational symmetry needs an approach that can find distinctive features in images and match rotated occurrences of the same using the measure of similarity developed by Fukushima and Kikuchi [49]. The model we use is illustrated

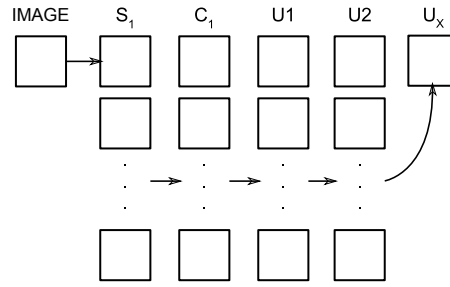


Figure 24: Rotational symmetry axes model

in figure 24. Its structure and function of the image layer U_o (equivalent to U_G in the [49]) through to C_1 (equivalent to U_C) is the similar to that of the symmetry axes extraction model. We implement these layers as we did in chapter 5, using the same model properties as described therein. The user is expected to provide a distinctive feature as an anchor in the input image as shown in figure 25. Based on the selection, all the feature vectors at locations in the hierarchy that spatially map to this location are matched with feature vectors of other locations starting from the C_1 layer and proceeding down the hierarchy for the closest match, and expanding the search into corresponding the receptive field. The U_H and U_X layers from [49] are now replaced by simple distance measures that are implemented in layers $U1$ and $U2$ to locate the centre of symmetry, and the angular distances between the matches (further details in appendix A.4). These layers collect structural features from the lower layers and compare them in a rotationally invariant way. At a later stage an alternative to manual input was introduced for automatic selection of image anchor using Scale Invariant Feature Transform [90] features.

7.1.3 Results

The feature hierarchy during the computation of the axes of symmetry for a typical example spiral design from *the Book* is shown in figure 25. A collection of 40 spirals spread across various folios of *the Book* is used to train this network. Unfortunately, it is not possible to validate the performance without ground truth data corresponding to the measurements performed, namely the spatial location of centre of symmetry, and three angular distances at which the anchor was found to be repeated. On an average, approximately 11% of the axes were not detected due to distortion or degradation. Of the remaining 89% the standard deviation σ of angular separation of the axes was 48° , which points to a large deviation from precision in execution of these spiral designs.

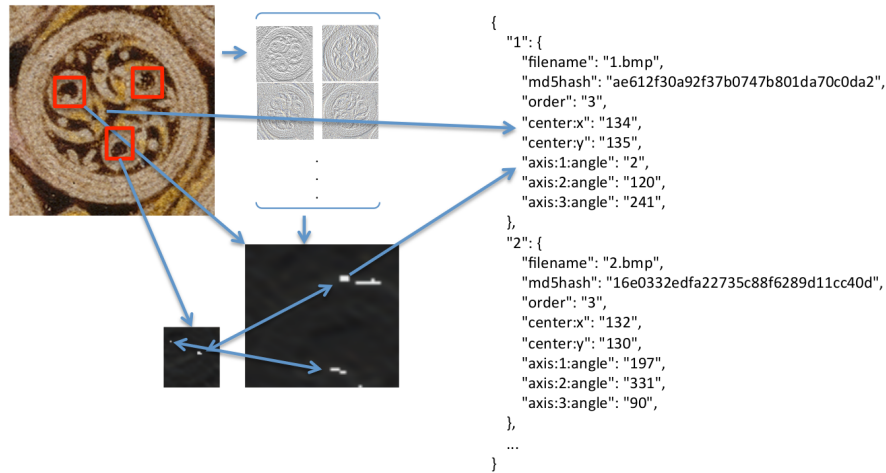


Figure 25: Feature extraction in the elements with circular symmetry. The image domain features selected by the user are marked in red. The lines in blue represent relations between recurring features or between the output and feature.

7.2 LOCALISATION OF FACES

Human faces and figures form an important part of the design repertoire of the Insular artists. In *the Book* and other manuscripts, the appearance of faces seems to have some correlations with the text and even functions as allusions to characters or events elsewhere in the narrative of the text. The attentional preference afforded to faces by the human visual system has been well documented [25]. This raises the question about the intentionality behind the face illustrations in particular folios. Any art historical analysis into such aspects may readily benefit from image processing techniques: the localisation of human faces has been one of the most well researched problems in computer vision. State of the art methods [131][112] are very often formulated as neural networks or machine learning solutions. The use of neural networks is promising as they are functionally and organisationally similar to the hierarchical image processing approaches discussed in this thesis.

As noted earlier, the hierarchical object localisation model we used in chapter 5 is generalisable enough to be used to locate faces in manuscripts. The generalisability of the model to the problem of localisation of faces and their demonstrated robustness in this particular task [118] makes it unnecessary to adopt a different image processing route. The procedures of training and testing detailed in chapter 4 are followed. The specific differences are noted below. The image datasets used for training and validation of the model had no common images. The training set consisted 52 fully visible faces and 10 partially visible faces in images of *the Book*, the Lindisfarne Gospels and other such manuscripts. The dataset used for validation of face

localisation (which is reused in the next chapter) contains 33 fully and 9 partially visible faces at various sizes ranging from approximately 30×30 pixels to 120×120 pixels. The model was trained with fully visible faces only and was able to locate 24 full faces and 1 partial face. When the model was retrained with partial faces as well, the number of fully visible faces detected decreased to 21 and 4 out of the 9 partially visible faces were now detected. Also, it was observed that in this case, the number of false positives (i.e. locations that were detected as faces even though they did not contain any) increased from 3 to 15. Since our definition of a partial face is the occurrence of a face with no more than one eye visible, the loss of features in these cases causes a corresponding drop in accuracy. The localisation results for 32v of *the Book* is shown in figure 26. It is noteworthy that while faces are localised at varying scales of occurrence, one of the faces is missed due to the low contrast difference between the face and eyes combined with the presence of a colour between the eyebrows and the eyes that is not consistent with the rest of the face.

7.3 LAYOUT ELICITATION

In *Page design of some illuminations in the Book of Kells* [103, pp. 243–256], Robert D. Stevick proposed some rules that the artists may have followed to organise the textual and illustration content in various folios of *the Book*. He seems to suggest that these almost recursive rules imbue some aesthetic qualities. While Larson and Picard [85] draw some guidelines that improve aesthetics of digital content comprising text and images, the analysis of existing layouts in manuscripts would allow art historians to validate such assumptions about the effect of layout on aesthetics.

This image processing field of layout analysis deals with the elicitation of perceptual groups of visual information. Each group may contain other groups within itself. The parallelism between this recursive organisation of a document and the models discussed and developed in this thesis is obvious. Due to their handwritten nature, text in manuscripts traditionally has considerable variation in the shapes and sizes of characters, fluctuations in the length and position of text lines, changing scripts and writing styles, and variance in the layout itself. Among the recent advances in the subject is [51] who approaches the problem using Scale Invariant Feature Transform [90] features that are classified and voted into layout entities. Breuel in [18, 17] compiles various geometric algorithms for layout analysis of images of printed documents, which already subscribe to a well defined layout. Other older methods [122, 102] also employ various geometric rules to connect contiguous areas of text into groups.



Figure 26: Faces detected in MS 58 folio 32v

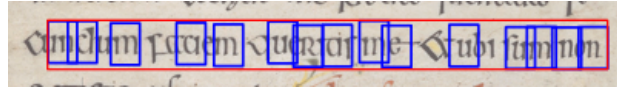
7.3.1 *Detecting lines and paragraphs*

The application of layout analysis of text in manuscripts uses localisation input from the object recognition and localisation model from section A.1. Since the number of lines and lines in a paragraph are unknown initially, unsupervised methods are used. Specifically two functional layers WX and WY (see figure 32) perform hierarchical clustering of localisation output in the horizontal and vertical directions respectively. We use threshold for distance cut off for joining is assumed to be 2.2 times the dimension of a letter (to allow for the spacing between words) in WX and 0.1 in WY. Operating under the assumption of absence of skewing in the images of manuscripts we approach the problem by adding two model layers to the localisation model, that perform successive clustering on the results of localisation for the characters (instead of interlace points). The data set used for training and testing this application is the same as the one used in character recognition. The first layer groups detected characters horizontally into lines, and the next groups lines vertically into paragraphs. The output of the two layers is demonstrated in figure 27 for some examples of text. It may be noted that our character localisation application was not trained with the complete alphabet not with uncommon variants of particular characters that the scribes sometimes used. This accounts for the letters that were not detected in the shown figure.

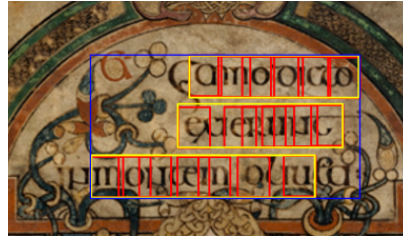
It should be noted that while the model is trained for character localisation as earlier, the clustering process is not strictly similar to the operations that are performed by the cells in the hierarchical model layers. However, the possibility of integration with non hierarchical methods is promising for applications that require unsupervised methods, such as the problem of estimating the number of scribes based on calligraphic styles.

7.4 CONCLUSION AND FUTURE WORK

The three applications explained in this chapter apply the concepts developed in earlier chapters to other minor manuscript analysis problems demonstrating the specific properties of adaptability and localisation. The model for symmetry predicts large deviations from the perfect rotational symmetry, but raises the question of intentionality of irregular designs. A study with larger scope would be aid discovery of any specific trends. The modification of the hierarchy to support detection of symmetry of order larger than 4 (which is the upper limit for our study) would be the logical next step. The results from face localisation are used in the next chapter as an indicator of visual attention. As for layout analysis, the two clustering layers do not provide adaptive feedback about locations, where localisation may



(a) Combining individual characters (marked in blue) into a line (marked in red) (MS58 folio 114r)



(b) Combining characters in to a line and lines into a paragraph (MS58 folio 24v)

Figure 27: Layout analysis of textual content

have failed, to the lower part of the hierarchy, and this may be considered as a direction of future research that integrates the concept of top-down and bottom-up learning in the same iterative learning architecture. Also, the restriction of clustering along different axes is limiting: letters arranged as rectangular blocks spatially separated from each other will always be merged into horizontal lines first if possible, which may not represent the true document hierarchy.

AESTHETICS

ABSTRACT The importance of beauty, which we now refer to as *aesthetics* — of man, nature, or artefacts — has been recognized since antiquity. The aesthetics of visual arts has been a core focus of philosophy, art history and art theory since Plato. In the last decade, computer scientists have attempted to develop automated methods to infer the aesthetic quality of structured information such as websites as well as unstructured information in the form of paintings. Studies on photographic images also draw a correlation between several heuristic measures of aesthetics and aesthetic judgements of viewers. This chapter of the thesis aims to apply and extend these developments to the aesthetic evaluation of manuscripts. The hierarchical architecture detailed in the previous chapters is developed further to estimate a popular indicator of aesthetics: gaze behaviour. This application also demonstrates the abstract nature of questions that may be explored using the hierarchical image processing approach. An eye tracking experiment conducted to collect gaze data and validate such an object-aware gaze prediction strategy is also described.

8.1 INTRODUCTION

Early attempts to identify attributes of objects that critically affect aesthetics were inspired by an objectivist perspective on aesthetics: the idea that beauty is a function of certain characteristics of an object. Among those identified features were balance and proportion [3, 15], novelty and prototypicality [58], as well as contrast and clarity [54]. Birkhoff [15] suggested that aesthetics of an object depends on two factors: namely order, or simplicity, and complexity, the latter being a deterrent to the aesthetics of an object. Contrastingly, Christian von Ehrenfels and Hans Eysenck argued that simplicity and complexity act multiplicatively to determine the aesthetic appraisal [39, 40]. The Gestalt approach (e.g., [3]) rejected the idea that it suffices to describe elemental parts of an object to describe the whole figure. Instead, it was suggested that *Gestalten* are perceived first and their constituent parts second.

In the last few decades, the field of neuroscience has made possible the validation of various such hypotheses in aesthetics. Several validated observations on the preferences towards visual aesthetics exhibited by the beholder are described by Palmer et al. [104]. Chatterjee [26] discusses the parallel organisational principles of the brain and the intent and practices of artists, the description of informative anecdotes, and the emergence of experimental neuroaesthetics. This field has drawn correlations between brain function and the performance of aesthetic judgements via medical imaging techniques such as functional magnetic resonance imaging [68] and magnetoencephalography studies [23]. Ishizu and Zeki [64] observe that there is a correlation between the activity of particular areas of the brain and the experience of beauty even though the source of such an experience may vary widely. Such empirical studies are valuable to the ground up formulation of theories of fun, intrinsic motivation and creativity [115] and ultimately of aesthetic experience.

In the area of classification aesthetic appraisal and judgement of paintings, Pihko et al. [106] compared subjective aesthetic judgements, emotional evaluations, gaze patterns and electrodermal reactivity between art history experts and laypersons when they were asked to explore and evaluate a series of paintings that ranged in style from representational to abstract. While the degree of abstraction had a demonstrable effect on aesthetic judgements of the laypersons, the opinions of the experts showed no such correlation with abstraction level. The gaze patterns of both groups changed as the level of abstraction increased: the number of fixations and lengths of saccades increased. Another study based on tracking eye movements evaluates the influence of bottom-up and top-down processes on visual behaviour while subjects, naïve to art criticism, were presented with representational paintings [92]. Wallraven et al. [132] also conducted a

similar study on the aesthetic appraisal of a large number of paintings from different art periods. In the experiment, observers were asked to rate these paintings on visual complexity and aesthetics, while their eye movements were tracked. The recorded gaze behaviour was compared with two predictive metrics - a visual saliency model developed by [66], and Colour Mutual Information. Visual saliency is the distinct subjective perceptual quality that makes some parts of images immediately grab our attention, and is an indicator of the aesthetic properties of images (see [13, 139]). While models of visual saliency have also been used by Wong and Low [139] to estimate aesthetic classification, bottom-up computational models such as [67, 66], that perform a low level structural analysis, have room for improvement by incorporating higher level information about contained objects, such as presence and location of faces [25].

8.1.1 *Questions*

One of the biggest challenges for art history, neuroscience and computer vision is that of understanding and simulating the way the brain perceives images even up to the level of understanding aesthetics. The effect of the presence of specific shapes and objects at particular locations, the use of particular colours and combinations thereof, repetition, symmetry, negative space, familiarity of the constituents, correlation with other works and forms of art all affect the perception of visual art. While general trends have been observed, it remains to be seen if existing indicators of aesthetics are as applicable to manuscript art. The specific and inherent importance of textual content in conjugation with illustrations does present some interesting questions: what is the nature of competition between faces, interlace, spiral, other design elements and text for visual attention in manuscripts?

8.1.2 *Problem definition, scope and possible solutions*

We consider the problem of modelling visual attention. While the study by Wallraven et al. [132] is oriented towards connecting aesthetic appraisal of paintings to their content via studies on gaze behaviour and to answer the question posed in the previous section, restrict our focus to manuscript images with combinations of text, interlace, human figures and other design elements. Aesthetics and consequently visual attention is affected by familiar objects and patterns [34], and existing predictors of attention are agnostic about such content. In this chapter we propose an improvement to Itti-Koch saliency metric [66] by incorporating localisation features for several objects specific to manuscript design: faces, characters and interlace. Validation of such a scheme for improved prediction would require data

about how the eye moves when viewing such manuscripts, however no such eye tracking data is available for early medieval manuscripts. The necessary eye movement data for validation of our proposed model may be collected as in studies by Wallraven et al. [132] and others. In the following section, we describe a human study for this purpose. Subsequently, we examine how existing predictors of gaze may be improved using features from an image processing hierarchy. These features are finally combined in a weighted scheme with Itti-Koch saliency and validated using the eye tracking data gathered from volunteers who are asked to make aesthetic judgements on images of manuscripts.

8.2 EYE TRACKING EXPERIMENTAL DESIGN

A total of 30 images of manuscript folios were shown to 25 volunteers. All the participants, with normal or corrected-to-normal vision, viewed the images of various manuscripts while their eye movements were recorded by a head-mounted tracker. Eye tracking data was collected with an SR Research EyeLink II eye tracker using a display PC for stimulus display with the EyeLink Toolbox [28] for MATLAB® [94] that leverages PsychToolbox [16, 78] and a synchronized host PC for recording of the data. All images were shown centred on a 22 inch monitor at a resolution of 1680×1050 pixels - as the paintings were of different formats, images were resized so that the image with the largest dimensions fit comfortably on the screen, while maintaining the relative (real) size ratios of the various manuscripts.

Before each experiment, a calibration of the eye-tracker had to be done in order to align the camera of the EyeLink system to each participant. For this, we used a standard 5 point calibration sheet, which consisted of five points with fixed pixel positions that were distributed on the screen and which participants were asked to fixate in a random order.

Each of the 30 (randomized) manuscript image trials consisted of a fixation dot that was shown for 2 seconds followed by an image that was shown for 15 seconds. Participants were asked not to direct their gaze outside of the image area during this period. After each image was displayed for the specific duration, three questions were shown on the screen and participants had to answer by entering their preference using the keyboard. The three aesthetic judgements expected of the observers were posed in the form of the following questions:

1. How visually complex was the image?
2. How aesthetically appealing was the image?
3. How dynamic (invoking sense of movement) was the image?

Participants rated the three aesthetic factors on a Likert scale of 1 (very simple, not aesthetic, static) to 7 (very complex, aesthetic, dy-

namic, respectively). No time limit was imposed for the recording of these responses. After the experiment, participants were asked to fill out a questionnaire which gathered data about how familiar participants were with art history and manuscripts such as the ones displayed. For the purposes of this study, the recorded ratings are ignored, though some general trends are noted later. The introduction of bias in a participant's behaviour during an experiment such as this by the task definition has been well documented [33].

The goal of this experiment is to collect eye movement data that would be representative of humans on average. Consequently, it is important to formulate the questions in such a way that the experiment itself does not influence the gaze by drawing specific focus of the subject to particular design elements, regions, textual content, illuminated letters, faces, etc. that may be present in the presented images. This is achieved by asking questions that are sufficiently abstract and require the subject to process image in entirety. The questions listed above fit this criterion and were based on the tasks defined in earlier studies [92, 132]. Ultimately the task of answering the aforementioned questions is also intended to draw the observers' focus to aesthetic appraisal.

Figures 28 and 29 shows the raw gaze information and fixation maps obtained for two of the manuscript images presented to observers in the eye tracking trials. Some general observations arrived at include higher attentional preference towards did appear to be attracted to high-contrast regions in the foreground, specifically faces and large or complicated designs such as illuminated letters.

8.3 GAZE PREDICTION

In the previous chapters, we applied an object localisation framework to the location of faces, letters, interlace crossings and some other Insular design elements. While Itti et al. [66] explore the bottom-up modelling of saliency, these models are mostly agnostic about specific objects and their structural features. Using the localisation output from the previous chapters could complement these approaches by making them more object-aware. Here, we continue building on top of the localisation model hierarchy and present a gaze prediction scheme that is tuned to the presence of design elements commonly occurring in manuscripts. We submit that features from the layers of the model discussed in section 5.2 (that are finally used in localisation task) can be directly used in the extended hierarchy to predict average fixation maps.

Keeping in mind that such a direct expansion of the model would require significantly higher computational resources¹ than required for the previous applications, we evaluated two very similar approaches.

¹ See also appendix A.6.1.

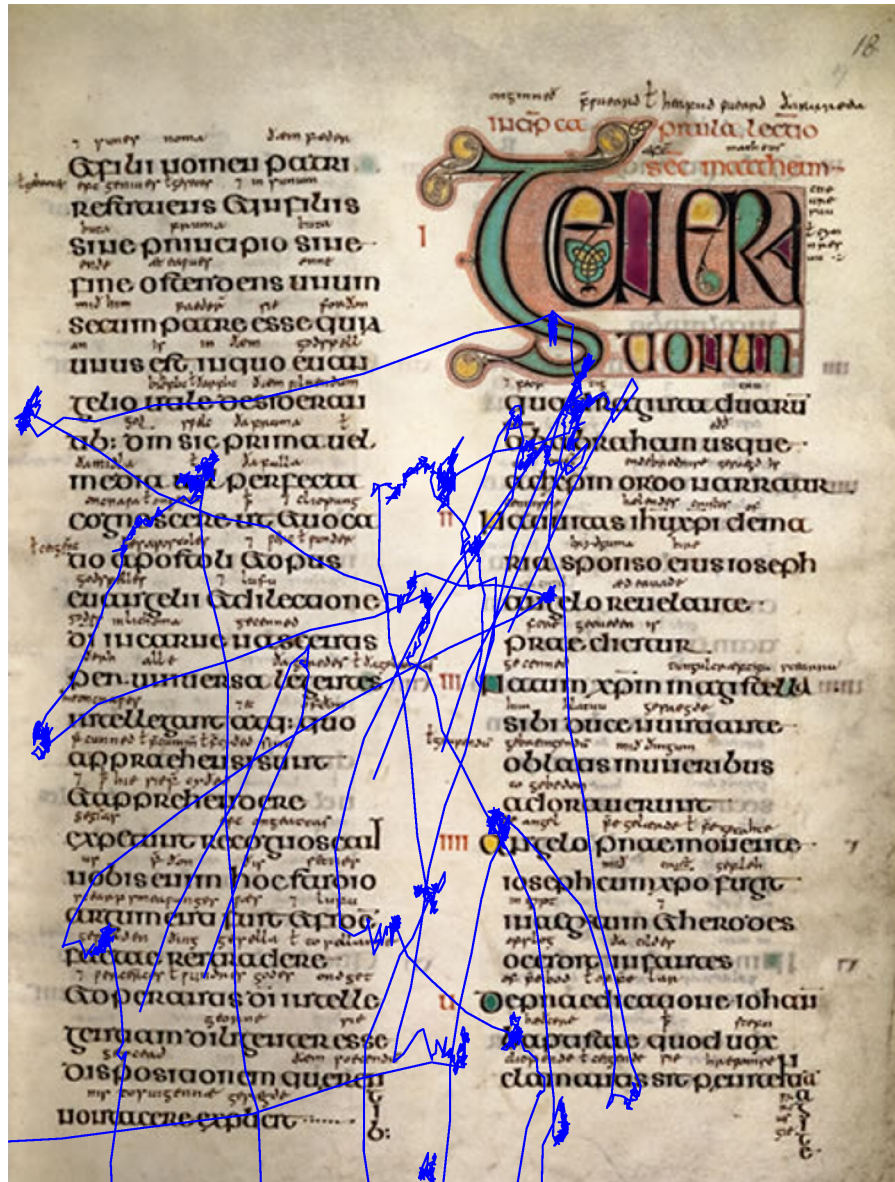


Figure 28: Raw gaze data for folio 18 of the Lindisfarne Gospels for one observer

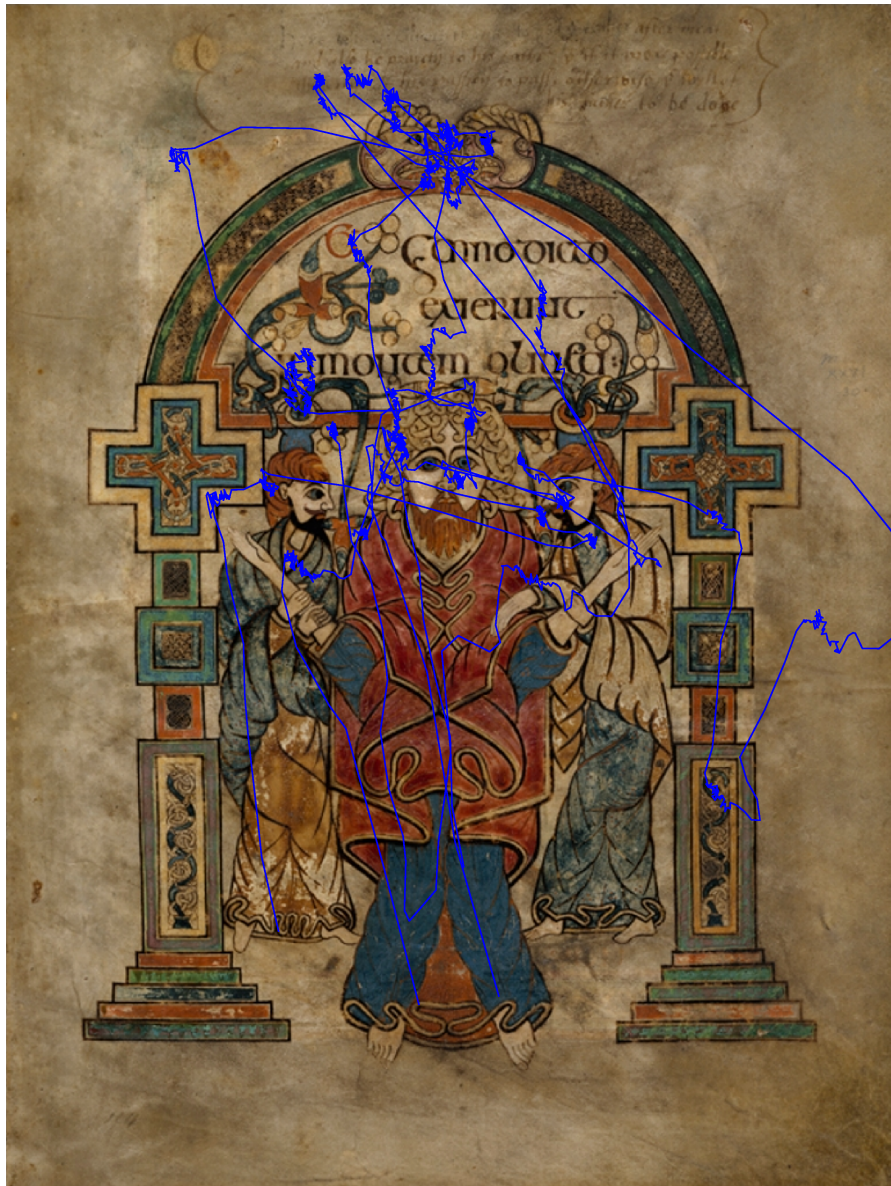


Figure 29: Raw gaze data for folio 114r of *the Book* (Trinity College Library MS 58) for one observer

In the first approach, we treat each of the localisers for faces, interlace and characters as black boxes, and only use their localisation and scale output. For compatibility with Itti-Koch's metric, the generated attention maps for faces and letters are approximated by spatially-ellipsoid Gaussian wells, and interlace by spatially-circular Gaussian wells, compensated for scale. A combination of both using boosted features (location, class of object, and scale) is attempted and shows a comparatively better prediction than either approach individually. This approach are described in more detail in appendix A.5. In the second approach, we complement the object localisation model with layers to combine the output of previous layers with the Itti-Koch saliency metric. This approach is discussed below.

8.3.1 *Extended hierarchy for gaze prediction*

In the case of gaze prediction, the data features we hope to extract from the hierarchical object model include the location and scales of the various objects of interest: faces, letters, interlace crossings. The core model for object recognition and localisation is extended by one layer that combines the saliency metric from Itti et al. [66] with the features that are learnt for the purpose of object localisation. Earlier, we noted that Mutch and Lowe [99] introduced some retention of scale and positional features by restricting the receptive fields of cells in the C2 layer, so they may be used for further classification at the SVM layer. This retention is important for scale detection, and this would not have been possible in the precursor model [118], where all positional and scale information was removed at the C2 layer (hence its name: global invariance layer). Once the SVM has been trained for the localisation of the three classes of objects of interest, we choose the SVM score at each position from each of the three binary SVM as the localisation input to the final gaze prediction layer. The SVM class is used as a feature representing the type of object in the gaze prediction layer. Scale information is gathered from the C2 layer by picking the scale of highest weighted features that were not dropped during localisation training. The Itti-Koch saliency was computed separately, and for each position in the image, the three metrics were combined by the final layer: a regression SVM treating each pixel in the image as training vector. This was done over the set of all images. It is noteworthy that in this approach, there is no need for the two dimensional Gaussian wells used in the the non-hierarchical approach for gaze prediction: this function is now performed by the SVM scores without the need for any further processing. Also there is no need for adjusting for scale, which is now directly used as an independent feature.

TRAINING The core model (see A.1) is trained for localisation. It may be noted that in this case the all-pairs m -class linear SVM with $m = 3$ is implemented by $m(m-1)/2 = 3$ binary SVMs are created that are used for the pair wise classifications: faces vs. letters, letters vs. interlace crossings and crossings vs. faces. In the next stage, the regression SVM is trained over each pixel in the image over all images in the collection.

8.3.2 Results

Of the 30 images that were displayed during the eye tracking experiment discussed in the previous section, 24 were used for training the weighted combination, and the remaining 6 were used for validating the approach, of which 2 images — folio 18 of the Lindisfarne Gospels and folio 114r of *the Book* — are discussed as examples. Figures 30a and 31a show the average attention maps for these two images. Figures 30b and 31b show the corresponding Itti-Koch saliency maps. Figures 30c and 31c show detected object locations; as mentioned before, for compatibility with Itti-Koch’s metric, the generated attention maps for faces and letters are approximated by ellipsoid Gaussian wells, and interlace by circular Gaussian wells, compensated for scale. It is clear that there are centres of attention that are predicted by either metric and not the other. Figures 30d, 31d (approach 1, detailed in appendix A.5) and 30e and 31e (approach 2 detailed in section 8.3.1) that represent the learned combination of the features detailed above with Itti-Koch saliency demonstrate the validity of the hypothesis that object awareness is important for attention modelling, and that a hierarchical model (approach 2) can perform these computations as an extension of the core localisation model of the framework. We also observe that the integrated model from our second approach approximates average gaze data better than the first approach, aside from being more extensible to similar applications.

8.4 CONCLUSION AND FUTURE WORK

As we have progressed through this thesis, we have discussed a variety of art historical applications of a hierarchical architecture for image processing. It is noteworthy, that the area of the image being analysed in each chapter has been increasing over this course: we began with detection of interlace points and characters in the text of *the Book*, proceeded to slightly larger regions considered at time in the In this chapter, we deal with a complete manuscript image at a time. The task specific layers for gaze prediction can be simply implemented on top the task-independent part of the model hierarchy used in the previous chapters. The use of graphics processing units with larger memory or multiple graphics processing units on

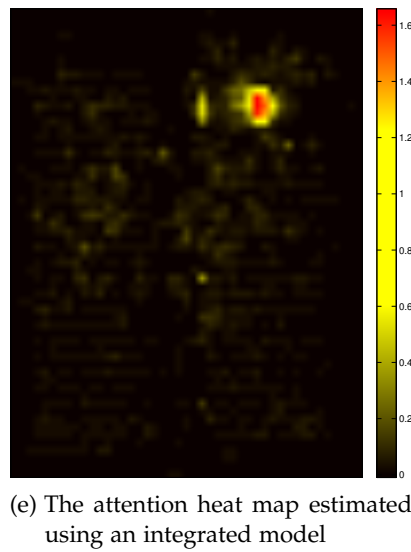
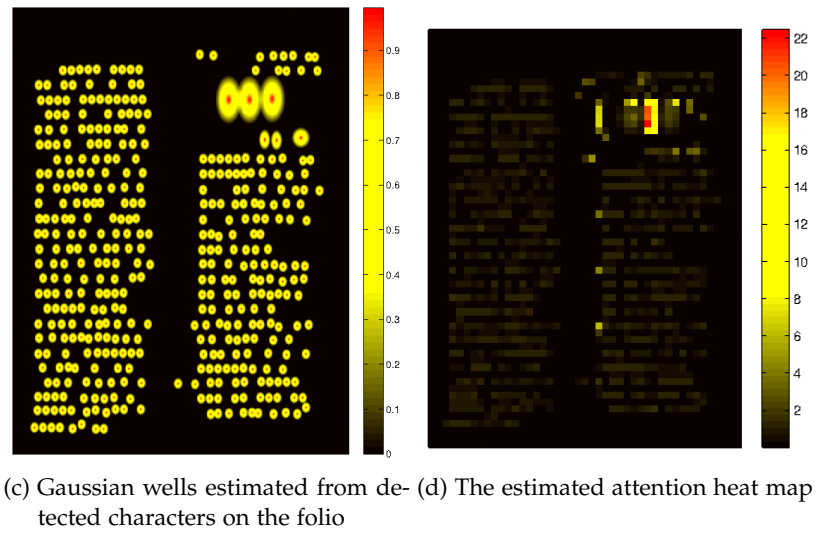
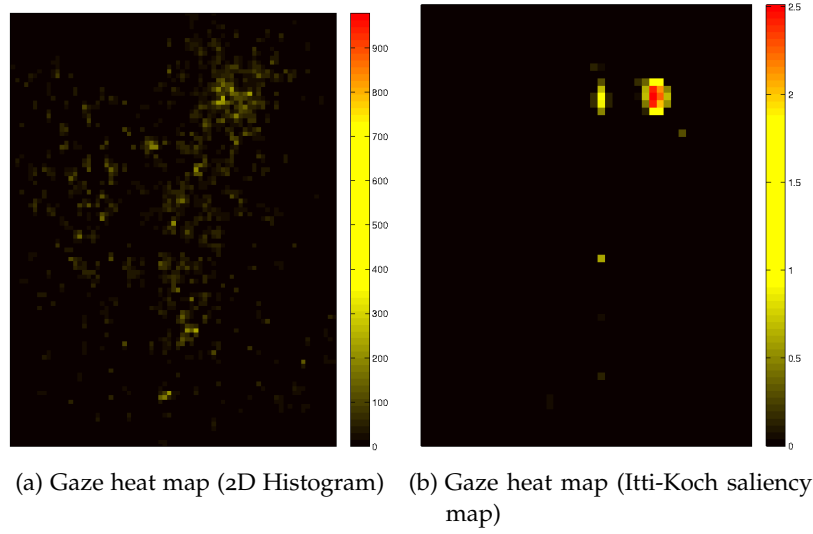


Figure 30: Gaze prediction for Lindisfarne Gospels (folio 18)

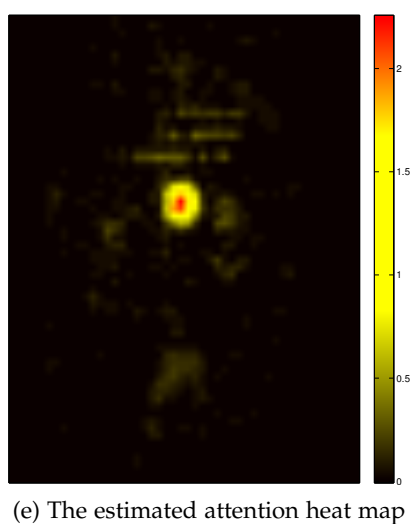
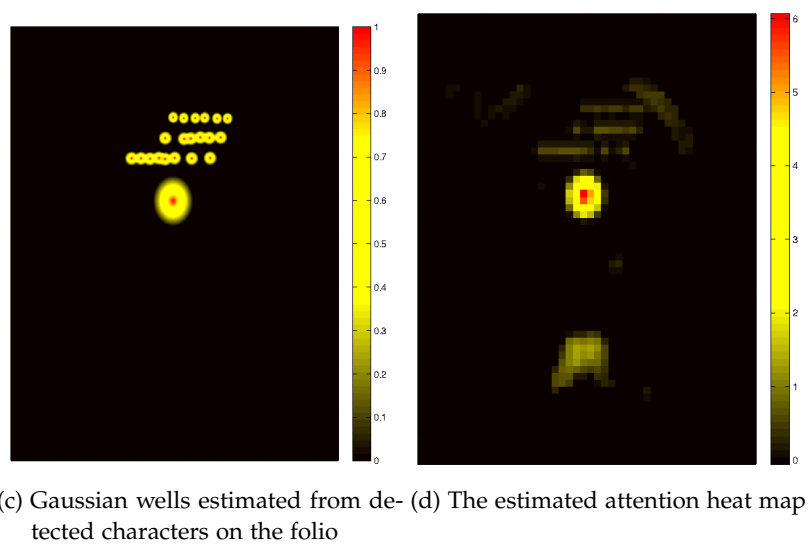
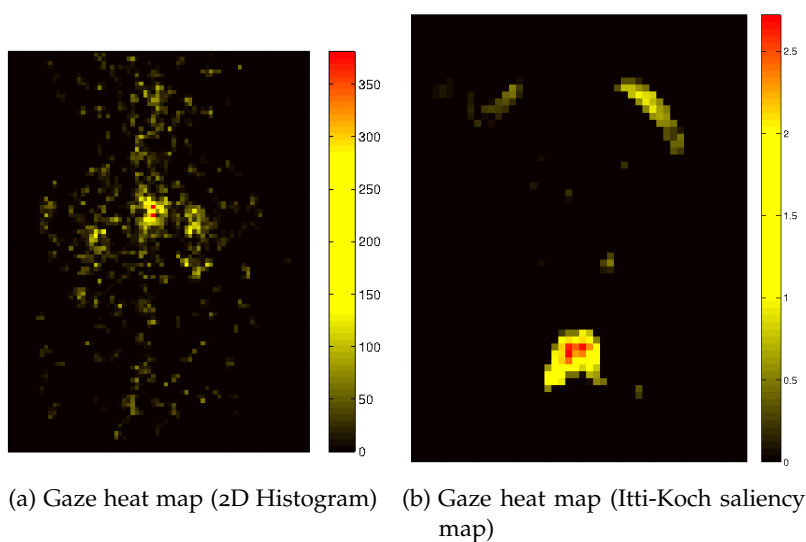


Figure 31: Gaze prediction for 114r of *the Book* (Trinity College Library MS 58)

the same computer might enable us to validate this application, and further extend it make even more content-aware gaze predictions by including features present lower in the hierarchy than the C2 layer. Another difference — our study also ignores colour information which would immensely increase computational resources needed for a realistic application of inferring attentional focus given by humans to such images.

The significance of this chapter is that it demonstrates the extensibility of the hierarchical image processing methods and applications thereof to art analytical problems of varying complexity - ranging from the simple, viz. computation of spatial metrics that can answer questions about the accuracy of spiral designs to the complex question of predicting gaze and attention that have implications in aesthetics.

CONCLUSION AND FUTURE WORK

This chapter summarises the conclusions of this research and provides directions for future work on image processing techniques and applications to manuscript analysis, and more generally to art history.

In this work, the viability of a hierarchical image processing model for application to character and hand analysis, interlace deconstruction, design element recognition, learned degradation removal, and gaze prediction has been demonstrated in the field of manuscript analysis. A unified approach was favoured for tackling this diversity of problems, which still only represents a subset of the problems art historians look at. The scope of the contributions in this study is limited by a design-morphology centric approach. It is hoped that in the future, in combination with exploration of the social-political context for a work of art, the methods described herein will lead to more comprehensive understanding of manuscript art, which is at the same time founded in the quantifiable and measurable.

9.1 CONCLUSION

This research was conducted with the aim to move toward the grand vision of automated digital understanding of art. The analysis of the more specific art form of manuscripts was explored as the most viable step in research in this direction. While the nature of research for manuscripts is still vast, a selection of the problems that art historians consider was presented as potential areas where the digital domain could provide measurable data. While looking for solutions to such problems in the field of image processing, it was noted that most methods that could be applied to solving such problems were either one sided (top-down or bottom-up) or fundamentally different from each other or both. The latter limitation is rather severe, considering that almost all image applications are modular, and any robust solution would require the modules to influence each other. The varied nature of art historical questions, ranging from pixel scaled (such as linear measurements in the image domain) to abstract such as the definition of aesthetics, or even its precursors, and the aforementioned incompatibility encouraged focusing on image processing methods that allow both a wide variety of inferences and consistency of representation. In this work, we presented a popular computer vision architecture that fits these requirements and proceeded to develop art history applications, and demonstrated its generalisability

by integrating useful properties of very similar models into it. Some applications considered for this study are discussed below.

A very popular question in regard to the Book of Kells (which was taken as an exemplar for all applications) is the number of scribes who wrote the textual content. This problem of hand classification was approached by combining an object recognition model Mutch and Lowe [99] with structural features that have been used by historians to determine the stylistic differences between scribes. If presented with enough samples of text by assumed hands, such a model is able to predict letters as belonging to one of these hands with reasonable accuracy. In analysis of interlace structure, the same model is extended using developed heuristic measures of error, scale and similarity. Such metrics are aimed at provide a numerical foundation to problems such as detecting similarity of designs across two manuscripts, among others. As part of this exercise, it was noted that the natural processes of degradation that have affected *the Book* over the centuries does play a major role in accuracy of the model used in extracting the structure of interlace motifs. The problem of image reconstruction, specifically that of degradation removal was examined with the aim to improve these results. An extension made to the model was able to reconstruct damaged regions of interlace designs and simultaneously boost the performance of interlace crossing detection. The development of measures for accuracy in spiral designs and grouping characters into lines and lines into paragraphs were studied next; both these problems required independent additions to the core model and demonstrated that questions more abstract than linear measurements are also possible using the proposed architecture. Finally the question of prediction of gaze for manuscripts was explored using information about various design elements combined with known saliency metrics, that performed significantly better as a result of the variety of new features made available for such predictions by such an integration.

The structural relationships between the image processing models used in each of the aforementioned applications are discussed below.

COMPARISON OF MODELS From the previous chapters, if we were to draw a comparison between the models used for each application, we would note that the chapters 4 and 5, and the application of face localisation in section 7.2 use the same core object recognition and localisation model. While hand classification in chapter 4 requires some additional information from the S1 layer that is finally combined with the object classification Support Vector Machine, the interlace metrics directly operate on the localisation output of the same core model. In chapter 6 we extend the core model by adding two layers that perform the function of degradation removal. The metrics for rotational symmetry developed in section 7.1 uses only a small section of the

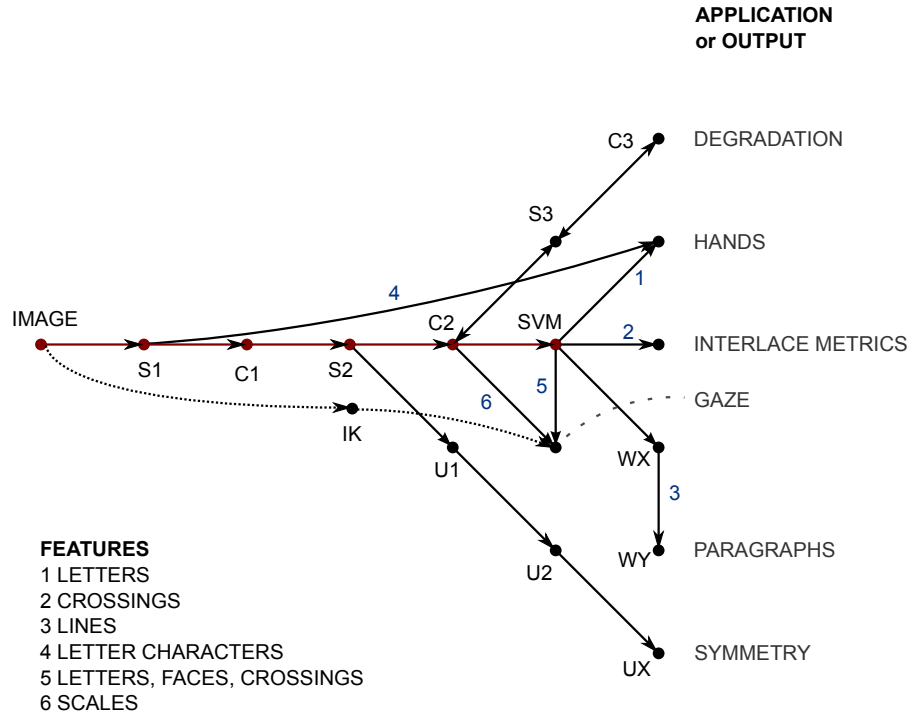


Figure 32: Modular comparison of (flattened) models for applications examined in this study. The layers and connections marked in red collectively represent the core object recognition and localisation model. Extensions and other operations developed for the applications covered in the previous chapters are marked in black.

core model and implements symmetry detection as layers built in the place of S2, C2 and the SVM layers of the core model. The extraction of layout uses the local maxima obtained from the SVM layer of the core model in two consecutive stages of clustering, along the horizontal and vertical directions. And finally, the gaze prediction application combines localisation features for faces, interlace and interlace crossings with the output of the Itti-Koch [67] saliency model. These organisational relationships between the various applications are illustrated in figure 32.

It is noteworthy that there are benefits to such a shared representation of features. The first of the two most important benefits of the models discussed in this study is the integration of a degradation removal (or in more general terms, an image reconstruction) model with the object localisation framework, which by equivalence also allows object categorisation. The second is the generalisability exhibited in gaze prediction. Therein, we combine information about various objects that we are interested in along with an existing saliency metric to provide answer to an abstract question such as gaze prediction. While earlier methods were at the least agnostic of objects present in images, and at the most ignored objects completely, a combined

hierarchical approach such as taken in this study allow deeper introspection than possible before.

So far, we have summarised the key aspects of this study and in the following section we briefly collate its limitations and discuss lines of inquiry that would benefit from further research.

9.2 FUTURE WORK

The 16 full page images of *the Book* used for this research¹ were scanned using very high resolution camera from transparencies that were originally captured by Faksimile Verlag, Lucerne, Switzerland, in 1990. Care was taken while choosing the images to ensure that the content were consistent with the aims of the research. However, the use of such a small set of images does raise the question of generalisability of the methods developed. In addition, each of the applications is dependent on a training process, which as often reported in the respective chapters, requires precise and manual preparation of images. While the predecessors of the models have been rigorously tested over a broad range of images, scales, in the presence of noise and so on, the core focus in this thesis has been the unified approach to solving diverse art historical problems. The viability of such an approach has been validated via specific applications discussed, which are by no means comprehensive, either in terms of features that were extracted for classification problems, or in terms of validation by testing over other manuscripts.

For instance, consider the case of hand classification. Studies based in classifying or describing typeform [36, 19] and calligraphic characteristics [62, 103, pp. 183–195] refer to a myriad of properties of the handwritten or typewritten letter. All of these may be added to the features extracted from the S1 or even other layers higher in the hierarchy to complement the features used in our study. It is important to note that the argument for greedy integration of more features that may be used for classification (and by extension for all other classification problems) is weak: usually classification accuracy has been shown to increase with increased number of features till it reaches a maximum, after which any further increase in features in the classifier causes suboptimal accuracy. However, in our case, the model already performs feature selection (dropping of unnecessary features) and given that only four structural features are used in the final classifier, adding more features will definitely improve the robustness of the hand classification scheme. Thus we conclude that extraction of more features is less prior than generalisation of the model, specifically to degradation. However, any practical art historical application would necessitate extraction a higher and more comprehensive set of structural features from letters, and this is a very promising avenue

¹ See appendix A.7 for more details.

for digital art historians. In contrast with the supervised approach taken here, unsupervised hand classification warrants exploration as a method to estimate a statistically optimal number of hands in *the Book* and other manuscripts, assuming a more comprehensive hand feature extraction is attempted as a precursor.

The application of interlace analysis has huge potential applicability for design comparison, further extension (from grid based) to graph based designs and improvement of the heuristic metrics developed here in. In our study we limited the interlace types in model training to single stranded and grid based designs. A multi-strand interlace design would invite complications in localisation of crossings, however, with a possible tuning of the Gabor filter parameters (in the S1 layer) and receptive field area it may be possible to locate the crossings with reliable accuracy.

The degradation removal model begs application to all the other models, and while the importance of learning in the reconstruction process is important, a potentially rewarding research problem would be to extend the range of task independent learning to degradation removal layers. It is surmised that images presented in training stage would have to be comprehensive: including natural images, containing all sorts of objects, allowing for changes in lighting, scale and position of presented objects and so on (see [118] for similar studies).

A multitude of applications in the same vein as in chapter 7 may be considered for future research, including performing similar measurements on other design elements: circles, key patterns, star rosettes, etc. The grand goal would be to design layers or metrics for each of the design elements in the Insular tradition (the most important of which are illustrated in figure A.8). The importance of a unified computational architecture, as developed in this thesis, will be felt in such diverse applications, where groups of layers that are reused across applications may be bundled into modules.

While the previous chapter only touches on the area of aesthetics of images, the possibility of using features extracted from various parts (differing in two dimensional position in the image or a parallel layer, or scale, etc.) of the model to rate aesthetic properties of an image are demonstrably within reach. While other approaches have attempted such questions earlier [34], using a more detailed and extensible approach that is the subject of this research would provide the necessary flexibility to test hypotheses about the effect of a particular factor (say, the presence of a cat) aesthetic properties of an image via user studies such as the one conducted in chapter 8.

Finally, one of the most important limitations of this study is that all the hierarchical image processing models were designed to work with grayscale images. While the naïve solution of stacking three independent such networks is definitely worth exploration on a small scale, the increase in size of the network would make it infeasible to

implement such networks in a realtime constrained application. The more 'gestalt' approach of adding a dimension to the hierarchy corresponding to colour should also be considered for problems such as colour palette generation, colour based degradation removal and so on. The introduction of such a domain would no doubt open up such models to more complex applications, such as a colour-aware predictor of gaze behaviour. However, such research would probably have to invest in high memory computing or cluster computing based approaches for such comprehensive applications.

In conclusion, the fields of neuroscience, image processing and art history are approaching a confluence where computing technologies will be able bridge the gap between the theories across the three disciplines and practical applications. It is hoped that this study will help clear the brambles along the way, and may be lay a brick or two too.

APPENDICES

A.1 OBJECT RECOGNITION AND LOCALISATION MODEL

This biologically inspired computer vision model was proposed by Mutch and Lowe [99] as an improvement on Mutch and Lowe [98], which builds up on a series of improving visual models: [117] based on the earlier HMAX model by Riesenhuber and Poggio [111]. These models are part of a group which also includes convolution networks (see [86]) and NeoCognitron by Fukushima [45]. All these approaches start with an image layer of (grayscale) pixels, and successively compute features in layers higher in the model, alternating between S- and C- layers (named by analogy with the simple and complex cells discovered by Hubel and Wiesel [63]). Simple (S) layers use convolution with local filters to compute higher-order features by combining different types of units in the previous layer, whereas complex (C) layers increase invariance by pooling units of the same type in the previous layer over limited ranges, at the same time, reducing the number of units by sub-sampling.

The model is illustrated in figure 33. The layers in order are the Gabor filter (S1), local invariance (C1), an intermediate feature (S2) and global invariance (C2). Finally we use an all-pairs linear Support Vector Machine (SVM) classifier after C2 vectors are normalised (mean to 0 and variance to 1).

Recent models have moved towards greater quantitative fidelity to the ventral stream. HMAX was designed to account for the tuning and invariance properties of neurons in Inferior temporal cortex. Rather than attempting to learn its bottom-level (S1) features, HMAX uses hard-wired filters designed to emulate simple cells found in the primary visual cortex (striate cortex, V1). Subsequent C layers are computed using a Hard MAX — a C unit's output is the maximum value of its afferent S units. This increases feature invariance while maintaining specificity. HMAX is also explicitly multi-scale: its bottom-level filters are computed at all scales, and subsequent C units pool over both position and scale. Serre et al. introduced learning of intermediate-level shared features, made additional quantitative adjustments, and added a final (non-biologically motivated) SVM classifier to make the model useful for classification. The base model is similar to [117] and the differences from [117] are listed in section A.1.

The model consists of five layers: an initial image layer and four subsequent layers, each layer built from the previous by alternating

template matching and max-pooling operations. Each of the layers is described below.

IMAGE LAYER We convert the image to grayscale and scale the shorter edge to 140 pixels while maintaining the aspect ratio. Next we create an image pyramid of 10 scales, each a factor of $2^{1/4}$ smaller than the last (using bi-cubic interpolation).

GABOR FILTER (S1) LAYER The S1 layer is computed from the image layer by centring two dimensional Gabor filters with a full range of orientations at each possible position and scale. The base model follows [117] and uses four orientations. Where the image layer is a 2D pyramid of pixels, the S1 layer is a four dimensional structure, having the same three dimensional pyramid shape, but with multiple oriented units at each position and scale (see figure 33). Each unit represents the activation of a particular Gabor filter centred at that position/scale. This layer corresponds to V1 simple cells. The Gabor filters are 11×11 in size, and can be described by:

$$G(x, y) = \exp\left(-\frac{X^2 + \gamma^2 Y^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda}X\right) \quad (3)$$

where $X = x \cos \theta - y \sin \theta$ and $Y = x \sin \theta + y \cos \theta$. x and y vary between -5 and 5 , and θ varies between 0 and π . The parameters γ (aspect ratio), σ (effective width), and λ (wavelength) are all taken from [117] and are set to 0.3 , 4.5 , and 5.6 respectively. Finally, the components of each filter are normalised so that their mean is 0 and the sum of their squares is 1 . We use the same size filters for all scales (applying them to scaled versions of the image). The response of a patch of pixels X to a particular S1 filter G is given by:

$$R(X, G) = \left| \frac{\sum X_i G_i}{\sqrt{\sum X_i^2}} \right| \quad (4)$$

LOCAL INVARIANCE (C1) LAYER This layer pools nearby S1 units (of the same orientation) to create position and scale invariance over larger local regions, and as a result can also sub-sample S1 to reduce the number of units. For each orientation, the S1 pyramid is convolved with a 3D max filter, 10×10 units across in position and 2 units deep in scale. A C1 unit's value is simply the value of the maximum S1 unit (of that orientation) that falls within the max-filter. To achieve sub-sampling, the max-filter is moved around the S1 pyramid in steps of 5 in position (but only 1 in scale), giving a sampling overlap factor of 2 in both position and scale. Due to the pyramidal structure of S1, we are able to use the same size filter for all scales.

The resulting C1 layer is smaller in spatial extent and has the same number of feature types (orientations) as S1; see figure 33. This layer implements a biologically plausible model for V1 complex cells.

INTERMEDIATE FEATURE (S2) LAYER At every position and scale in the C1 layer, we perform template matches between the patch of C1 units centred at that position/scale and each of d prototype patches. These prototype patches represent the intermediate-level features of the model. The prototypes themselves are randomly sampled from the C1 layers of the training images in an initial feature-learning stage. Prototype patches are like fuzzy templates, consisting of a grid of simpler features that are all slightly position and scale invariant. During the feature learning stage, sampling is performed by centring a patch of size 4×4 , 8×8 , 12×12 , or 16×16 ($\times 1$ scale) at a random position and scale in the C1 layer of a random training image. The values of all C1 units within the patch are read out and stored as a prototype. For a 4×4 patch, this means 16 different positions, but for each position, there are units representing each of 4 orientations (see the ‘dense’ prototype in figure 34). Thus a 4×4 patch actually contains $4 \times 4 \times 4 = 64$ C1 unit values. Preliminary tests seemed to confirm that multiple feature sizes worked somewhat better than any single size. Smaller (4×4) features can be seen as encoding shape, while larger features are probably more useful for texture. Since we learn the prototype patches randomly from unsegmented images, many will not actually represent the object of interest, and others may not be useful for the classification task. The weighting of features is left for the later Support Vector Machine (SVM) step. It should be noted that while each S2 prototype is learned by sampling from a specific image of a single category, the resulting dictionary of features is shared, i.e. all features are used by all categories. During normal operation (after feature learning) each of these prototypes can be seen as just another convolution filter which is run over C1. We generate an S2 pyramid with roughly the same number of positions/scales as C1, but having d types of units at each position/scale, each representing the response of the corresponding C1 patch to a specific prototype patch; see figure 33. The S2 layer is intended to correspond to cortical area V4 or posterior IT. The response of a patch of C1 units X to a particular S2 feature/prototype P , of size $n \times n$, is given by a Gaussian radial basis function:

$$R(X, P) = \exp \left(-\frac{\|X - P\|^2}{2\sigma^2\alpha} \right) \quad (5)$$

Both X and P have dimensionality $n \times n \times 4$, where $n \in \{4, 8, 12, 16\}$. As in [117], the standard deviation σ is set to 1 in all experiments.

The parameter α is a normalising factor for different patch sizes. For larger patches $n \in \{8, 12, 16\}$ we are computing distances in a

higher dimensional space; for the distance to be small, there are more dimensions that have to match. We reduce the weight of these extra dimensions by using $\alpha = (n/4)^2$, which is the ratio of the dimension of P to the dimension of the smallest patch size.

GLOBAL INVARIANCE (C2) LAYER Finally we create a d -dimensional vector, each element of which is the maximum response (anywhere in the image) to one of the model's d prototype patches. At this point, all position and scale information has been removed.

SUPPORT VECTOR MACHINE In the SVM classifier, the C2 vectors are classified using an all-pairs linear SVM. Data is sphered before classification: the mean and variance of each dimension are normalised to zero and one respectively. Test images are assigned to categories using the majority-voting method.

DIFFERENCES FROM SERRE ET AL. Mutch and Lowe [99] detail several methods to significantly reduce the number of features that are passed between layers. These methods vastly improve speed of the computations and have some positive effects in multi-scale detections and localisation. However, we discuss them in detail below, as some modifications for the image sets we use are affected by these.

It may be noted that in [118], the following are assumed, and the improvements also work around these restrictions.

- Image height is always scaled to 140 pixels.
- A pyramid approach is not used (different sized filters are applied to the full-scale image).
- The S1 parameters σ and λ change from scale to scale.
- S1 filters differ in size additively.
- Cells in the C1 layer sub-sampling ranges do not overlap in scale.
- Cells in the S2 layer have no α parameter.

SPARSIFY S2 INPUTS In the base model, an S2 unit computes its response using all the possible inputs in its corresponding C1 patch. Specifically, at each position in the patch, it is looking at the response to every orientation of Gabor filter and comparing it to its prototype. Based on the principle that features should be as sparse as possible, we reduce the number of inputs to an S2 feature to one per C1 position. In the feature learning phase, we remember the identity and magnitude of the dominant orientation (maximally responding C1 unit) at each of the $n \times n$ positions in the patch. This is illustrated

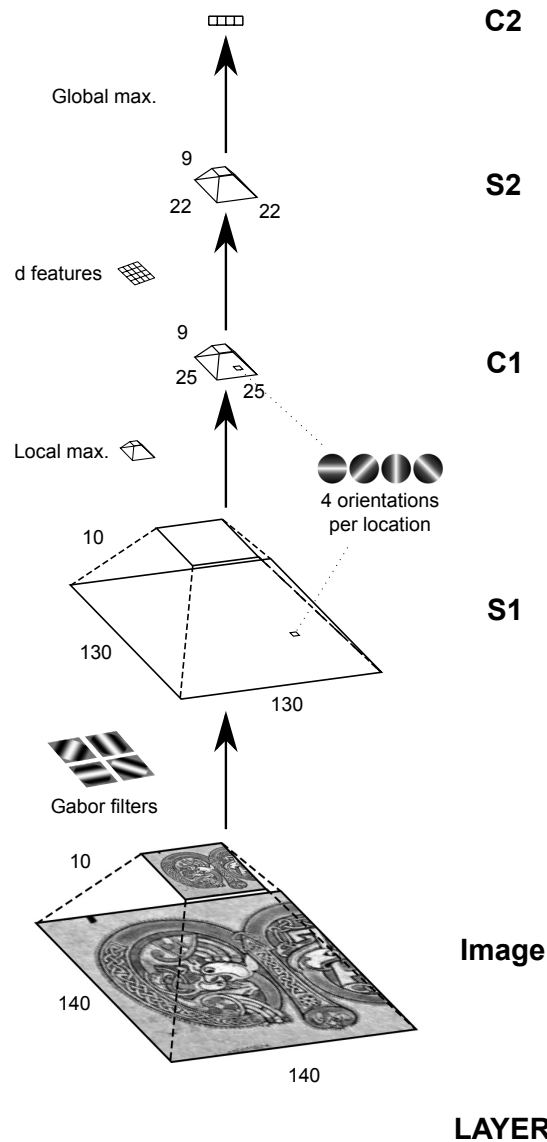


Figure 33: Overview of the base model (from Mutch and Lowe [98]). Each layer has units covering three spatial dimensions (X/Y/scale), and at each 3D location, an additional dimension of feature type. The image layer has only one type (pixels), layers S1 and C1 have 4 types, and the upper layers have d (many) types per location. Each layer is computed from the previous via convolution with template matching or max pooling filters. Image size can vary and is shown for illustration.

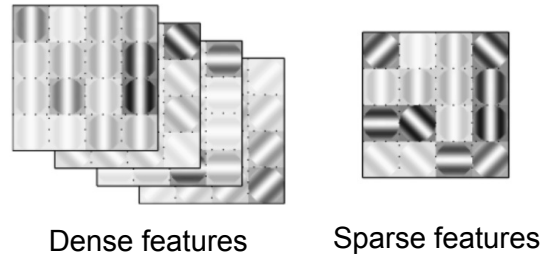


Figure 34: Dense vs. sparse S2 features (from Mutch and Lowe [98]). Dense S2 features in the base model are sensitive to all orientations of C1 units at each position. Sparse features are sensitive only to a particular orientation at each position. A 4×4 S2 feature for a 4-orientation model is shown here. Stronger C1 unit responses are shown as darker.

in figure 34; a 4×4 prototype patch now contains only 16 C1 unit values, not 64. When computing responses to S2 features, equation 5 is still used, but with a lower dimensionality: for each position in the patch, the S2 feature only cares about the value of the C1 unit representing its preferred orientation for that position. In conjunction with this we increase the number of Gabor filter orientations in S1 and C1 from 4 to 12. Since we're now looking at particular orientations, rather than combinations of responses to all orientations, it becomes more important to represent orientation accurately. Cells in the visual cortex also have much finer gradations of orientation than $\pi/4$.

INHIBIT S1/C1 OUTPUTS The second modification is similar — we again ignore non-dominant orientations, but here we focus not on pruning S2 feature inputs but on suppressing S1 and C1 unit outputs. In the cortex, lateral inhibition refers to units suppressing their less-active neighbours. We adopt a simple version of this between S1/C1 units encoding different orientations at the same position and scale. Essentially these units are competing to describe the dominant orientation at their location. We define a global parameter h , the inhibition level, which can be set between 0 and 1 and represents the fraction of the response range that gets suppressed. At each location, we compute the minimum and maximum responses, $R_{\min}$ and $R_{\max}$, over all orientations. Any unit having

$$R < R_{\min} + h(R_{\max} - R_{\min})$$

has its response set to zero. As a result, if a given S2 unit is looking for a response to a vertical filter (for example) in a certain position, but there is a significantly stronger horizontal edge in that rough position, the S2 unit will be penalised.

LIMIT POSITION/SCALE INVARIANCE OF S2 FEATURES As is true of many “bag of features” models, the base model disregards

all geometry above the level of S2 units. It simply uses the maximum response to each S2 feature at any position or scale. This gives complete position and scale invariance, but S2 features are still too simple to eliminate binding problems: we are still vulnerable to false positives due to chance co-occurrence of features from different objects and/or background clutter. We wanted to investigate the option of retaining some geometric information above the S2 level. In fact, neurons in V4 and IT do not exhibit full invariance and are known to have receptive fields limited to only a portion of the visual field and range of scales. To model this, we simply restrict the region of the visual field in which a given S2 feature can be found, relative to its location in the image from which it was originally sampled, to $\pm t_p\%$ of image size and $\pm t_s\%$ scales, where t_p and t_s are global parameters. This approach assumes the system is “attending” close to the centre of the object. This is appropriate for datasets such as the Caltech 101¹, in which most objects of interest are at similar positions and scales within the image.

SELECT FEATURES THAT ARE HIGHLY WEIGHTED BY THE SVM
Our S2 features are prototype patches randomly selected from unsegmented training images. Many will be from the background, and others will have varying degrees of usefulness for the classification task. We wanted to find out how many features were actually needed, and whether cutting out less-useful features would improve performance, as we might expect from machine learning results on the value of sparsity. We use a simple feature selection technique based on SVM normals. In fitting separating hyperplanes, the SVM is essentially doing feature weighting. Our all-pairs m -class linear SVM consists of $m(m-1)/2$ binary SVMs. Each fits a separating hyperplane between two sets of points in d dimensions, in which points represent images and each dimension is the response to a different S2 feature. The d components of the (unit length) normal vector to this hyperplane can be interpreted as feature weights; the higher the k^{th} component (in absolute value), the more important feature k is in separating the two classes. To perform feature selection, we simply drop features with low weight. Since the same features are shared by all the binary SVMs, we do this based on a feature’s average weight over all binary SVMs. Starting with a pool of 12,000 features, we conduct a multi-round “tournament”. In each round, the SVM is trained, then at most half the features are dropped. The number of rounds depends on the desired final number of features d .

¹ A standard dataset for object recognition performance evaluation. Available at http://www.vision.caltech.edu/Image_Datasets/Caltech101/. It contains pictures of objects belonging to 101 categories with about 40 to 800 images per category, and an average size of 300×200 pixels.

MODIFICATIONS In applying the above model to character recognition, it was noted that the inhibition of S1/C1 non-dominant orientation features did improve speed but had an adverse impact on recognition accuracy. This is expected as the scale of letters in our application is higher than that of objects present in images used in the validation of the aforementioned studies. This issue will be revisited in chapter 6 where lateral interactions are studied in detail. All the other modifications suggested before were incorporated into the model.

POSITION INVARIANCE AND LOCALISATION There are applications such as localisation, where the invariance produced by the network to position in the original image is counterproductive. Complete invariance would suggest that cells cannot reveal the position of the object. While very precise geometric constraints may not be useful for broad object categories, there is a substantial loss of useful information in completely ignoring feature location as in bag-of-features models. We have shown a considerable increase in performance by using intermediate features that are localized to small regions of an image relative to an object coordinate frame. When an object may appear at any position or scale in a cluttered image, it is necessary to search over all potential reference frames to combine appropriately localized features. In biological vision this attentional search appears to be driven by a complex range of saliency measures. For our computer implementation, we can simply search over a densely sampled set of possible reference frames and evaluate each one. This has the advantage of not only improving classification performance but also providing quite accurate localization of each object. The strong performance shown on the UIUC car localization task indicates the potential for further work in this area.

A.2 DEGRADATION RECONSTRUCTION MODEL

The proposed degradation reconstruction network introduced in section 6.2 relies on the iterative top-down propagation of learned object features to reconstruct degraded and noisy parts of images. In comparison with the object recognition framework in the previous section, the model pyramid is taller, i.e. the C1 and C2 layers perform the max-filter operation on a smaller neighbourhood of S1 and S2 layers respectively. While the earlier model has a filter radius of 10×10 units in position 1 and 2 units deep in scale, the degradation reconstruction model performs a slower abstraction by reducing the filter radius to 8×8 units in position at C1 and to 4×4 units in position at C2. The network layer sizes are changed to accommodate this smaller filter size. The layer sizes are computed as shown in table 2. Also, the C2 layer in uses the same logic as the C1 layer (max-filtering), where

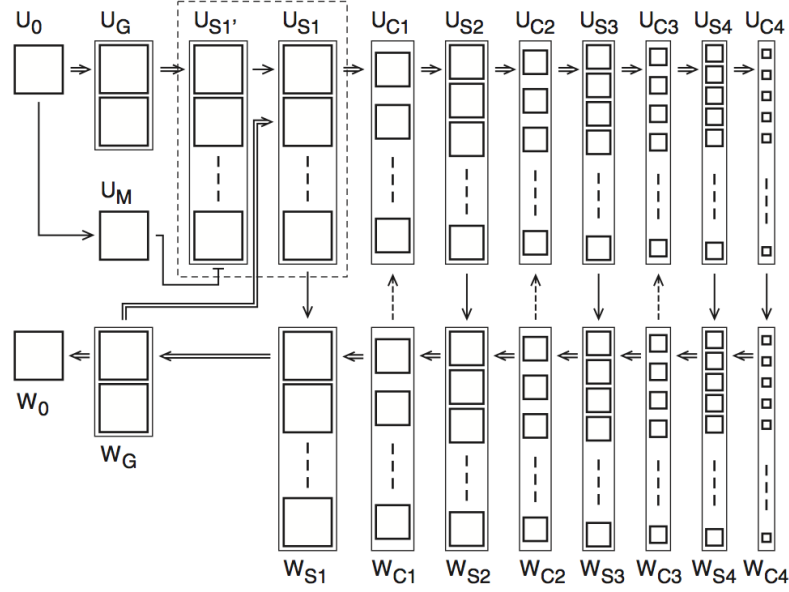


Figure 35: NeoCognitron: an occlusion reconstruction network model [48]

Model	Object recognition	Degradation removal
Image	140×140	170×170
S1	130×130	160×160
C2	25×25	39×39
S2	22×22	36×36
C2	d	17×17
S3	-	14×14
C3	-	d

Table 2: Comparison of layer sizes (in position)

as C3 layer is equivalent functionally to the object recognition model C2 layer.

The forward and backward projections between S2–C2 and S3–C3 implement a full connectivity with $d \times 14 \times 14 \times 8$ total weights in each direction. Backward projections of layers S1/C1 and image/output layers are non-overlapping. For each feature, different backward projections access the hyper-neighbourhood in the next higher layer. Lateral projections in the first three layers originate in the hyper neighbourhood of a feature cell. Its activities are copied from feature cells using wrap-around. In the C3 layer, the lateral projections access all d feature cells.

SPARSIFICATION IN S2 In the previous section, we discussed the specifics of a hierarchical object recognition and localisation model that we use in various applications throughout this thesis. Several

methods to reduce the number of features propagated between layers were also proposed and implemented therein. These optimisations reduced the computational inefficiency of a model which relies on all possible sub-feature combinations. For practical applications, such as object recognition, optical character recognition and the interpretation of handwritten text, where often a high speed of operation is a requirement, the network size plays an important role since real-time conditions must be met for the network recall. However, the problem of degradation requires backward projections of excitatory and inhibitory types. The sparsification optimisation at the S2 layer is counter-productive to this type of network, where a loss of features would prove detrimental to the reconstruction in the image domain. As such, this hypothesis was validated and S2 layer sparsification was omitted as a part of the proposed degradation removal model.

A.3 HAND CLASSIFICATION METRICS

As discussed in section 4.2.3, the following morphological information is chosen for the hand classification problem are directly extracted from the hierarchy.

1. Character width
2. Character height
3. Maximum brush stroke width
4. Minimum brush stroke width

The extraction of these metrics does imply the need for pixel level accuracy in the location of points in the image that could be used to compute the above. From the description of the models above, it is clear that C layers introduce positional invariance, and consequently, any layer above S1 would not aid us in achieving the goal of accuracy. Specifically we look for edges the bounding box for each letter. Operating under the assumption that the letters presented to the system are vertically upright or inverted, and not rotated, it can be concluded that only S1 (Gabor filter) cells with orientations of $n\pi/2$ are required for our purposes. The top-, bottom-, left- and right- most edge pixels are selected from the S1 layer, and character width and height are derived from these (for obvious reasons, edges lying on the image boundary are ignored).

For the measurement of brush stroke width, we use the knowledge that at the edges of a brush stroke, the S1 edges would be parallel. The minimum and maximum of all stroke widths as obtained from above are used as measures for the hand classification along with the character width and height calculated earlier.

A.4 SPIRAL METRICS

A.4.1 *Rotational symmetry extraction model*

The model we use is a combination of the HMAX based object recognition model of Mutch and Lowe [99] and the neural network architecture proposed by Fukushima and Kikuchi [49] for detection of symmetry axes in images. For the purposes of extraction of symmetry from an image, the goal is to integrate the metrics developed for symmetry detection into the more recent HMAX based model.

As with the model described in section A.1, we have layers for the input image, Gabor filter (S1) and local invariance (C1) as the first two layers of the hierarchy. Following these layers, we build two task specific layers for symmetry computations: U1 and U2 as described below.

U1 LAYER At every position and scale in the U1 layer, we perform template matches between the patch of U1 units centred at that position/scale and each of d prototype patches P , as we would have done in the C1 layer. The computational difference is noted in the response produced by the Gaussian radial basis transfer function:

$$R(X, P) = \exp \left(-\frac{\|X - P'\|^2}{2\sigma^2\alpha} \right) \quad (6)$$

In comparison with equation 5, here P' may be any rotated version of P . We consider specific rotational angles of $2\pi/2 = \pi$, $2\pi/3$ and $2\pi/4 = \pi/2$ corresponding to rotational symmetries of order 2, 3 and 4 respectively for our study.

U2 LAYER Again, this layer performs computation similar to the C2 layer. Once the user selected or SIFT [90] based anchor features are provided, the U1 features required for the localisation of the anchor are compared with other similar feature groups in U2. The closest matches are presented as other occurrences of the anchor, and UX finally calculates the centre and angular distances between these points.

A.5 GAZE PREDICTION BY BOOSTED LEARNING

Boosting is a machine learning ensemble meta-algorithm for supervised learning, which combines weak learners to strong ones. For gaze prediction, we assume that the weak classifiers are :

1. The Itti-Koch saliency metric [66]
2. Scale-aware localisation probability for

- a) faces
 - b) letters
 - c) interlace crossings
3. Scale
 4. Type of object: face, letter, or crossing

The scale of the detected objects are computed as previous section, except that in this case, we select only the top feature instead of a larger set. Gradient boosting (regression) [44] is used for the combination of these ‘classifiers’ into a regression model. As in the previous section 8.3.1, each location in the images is treated as a separate training sample.

A.6 IMPLEMENTATION OF MODELS

A.6.1 *Graphics Processing Units: memory and processor*

The use of GPUs in this work reflects the recent trend in moving away from the traditional processing units found in computers the Central Processing Unit (CPU) and toward GPUs for computations that are amenable to parallelisation. Most computer graphics and a large number of computer vision applications are of this nature, where the same set of computing operations are performed simultaneously over a large amount of data. In the case of our hierarchical models, cells in the network perform a fixed set of well defined operations while combining input from a large number of cells in preceding (‘afferent’) layers. Parallel computing on GPUs allow for speed improvements of up to 400 times that of CPUs if programmed with optimisation in mind. The Cortical Network Simulator framework (see next subsection for more details), that is used for running the models in this thesis, performs a large number of standard memory access and instructional optimisations, and speed improvements of ≈ 6 to 45 times have been observed over the course of this study for the models in question. In the case of simpler models we traded lower memory requirements for improved speed. For instance, the a stripped down version of the core object recognition model ran ≈ 40 times faster than on the CPU on a NVIDIA GTX 460 with 1 gigabyte of memory. When the recognition/localisation model as described earlier was partly run on this GPU (layers from image to C2) and CPU (Support Vector Machine), the speed up over CPU was on average 45 times. But with complex models such as that of gaze prediction, the memory required increased to ≈ 3.3 gigabytes and had to be run on a more recent GTX 970 (4 gigabytes of memory).

A.6.2 Frameworks

All the implementations of the models developed (or extended) for each of art historical applications discussed in part iii of this thesis were built using the Cortical Network Simulator [100]. This network simulator allows MATLAB [94] to seamlessly use the parallel processing capabilities of Graphics Processing Units (GPUs). Usually high-level numerical computing tools and programming languages (including MATLAB) can access GPUs memory and run instructions on its processors via the Compute Unified Device Architecture (CUDA) [101]. CNS provides the necessary abstraction layer between CUDA and MATLAB, and allows a programmer to define the network organisation, its layers, constituent cells, their parameters and connectivities. As such it immensely simplifies the task of designing, training and testing object recognition models that have a neural network like structure. The hierarchical models discussed throughout this thesis do move away from the traditional neural network mathematical formulations, but the organisation and connectivities are very similar, and the CNS is flexible enough to encode both types.

A.7 THE BOOK OF KELLS IMAGES

The image of *the Book* used for this research were provided for academic purposes by the Digital Resource and Imaging Services, Trinity College Dublin. These were made available as print quality high resolution scans of the Fasimile Verlag transparencies of the Book of Kells, originally captured in 1990. Given that the transparencies themselves were 20 years old, the scans are affected by film grain, which does pose a problem for image processing applications that focus on accurate measurements at high resolutions. One of the major motivations behind devoting specific attention to the development of degradation removal as an application of a hierarchical architecture was to deal with the image noise which ultimately was a combination of the degradation of *the Book* itself and film grain. The images were on an average $\sim 5800 \times \sim 7500$ pixels in resolution, with the exception of the Chi-Rho page, which was $\sim 12000 \times 15000$ pixels. All images accompanied by the colour calibration profiles required for any colour accurate analysis of its content.

A.8 DESIGN ELEMENTS IN INSULAR MANUSCRIPTS

Jacobsthal [69, 70] built a very comprehensive grammar of design elements - the structural alphabet of the Insular artists. The most basic classification is into three categories: geometric, floral and plastic. It may be noted that the Insular artists were not limited to using individual design elements in isolation, but freely combined them in

a most organic fashion. Some of these design forms are illustrated in figure 36. While this collection is by no means comprehensive, it does point to the infeasibility of applying model based image processing approaches for identification and stylistic characterisation of such a diverse set of designs: such an approach would require building computer models for each of the elements, and yet allow for the exponentially bigger set of their combinations.

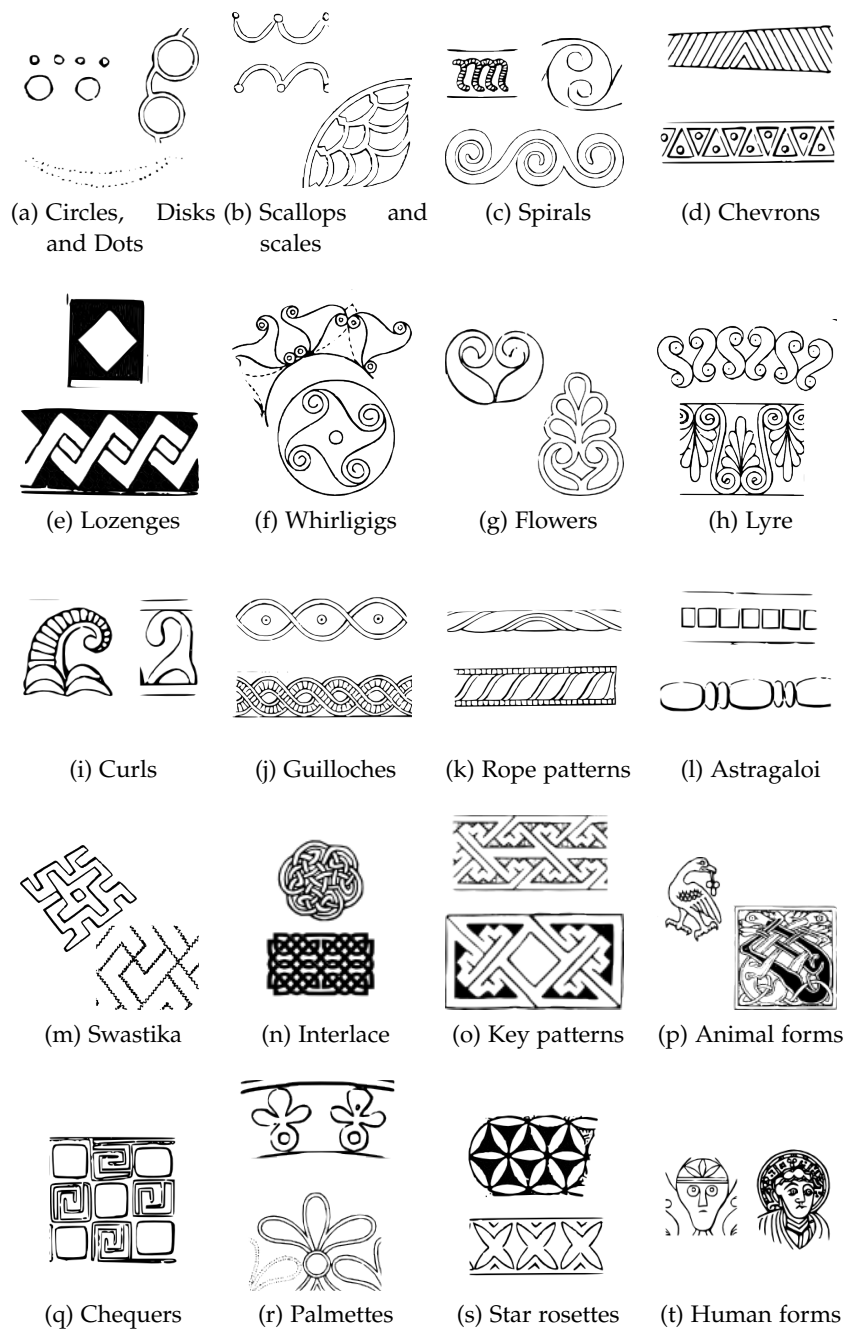


Figure 36: Design elements used of Insular art and sculpture (collected from [70] and other sources)

GLOSSARY

Back propagation method A common method for the training of artificial neural networks that, used with an optimisation method, updates the synaptic weights (strengths) between neural cells in a network. The goal of any learning algorithm is to find a function that best maps a set of inputs to its correct output. The back propagation method measures the mismatch between the expected output and the computed output and iteratively corrects the parameters of a (multi-layered) network to minimise such an error.

Classification In machine learning and statistics, classification is the problem of identifying to which of a set of categories a new observation belongs.

Convolutional Neural Network A type of feed-forward artificial neural network where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field. Convolutional networks were inspired by biological processes and are variations of multilayer perceptrons which are designed to use minimal amounts of preprocessing. They are widely used models for object recognition in images and videos.

DC component In frequency domain signal analysis, the DC component refers to the zero frequency component. The name DC component comes from electronics, where it stands for Direct Current (as opposed to Alternating Current). The DC component in any signal, is its time average (time average is equivalent to zero frequency).

Gabor filter In image processing, a Gabor filter, is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. Simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions. Thus, image analysis with Gabor filters is thought to be similar to perception in the human visual system.

Gaussian function In mathematics, a Gaussian function has a bell-like shape. Gaussian functions are widely used in statistics where they describe the normal distributions, in signal processing where they serve to define Gaussian filters, in image processing where two-dimensional Gaussians are used for blur operations.

GPU Graphics Processing Unit: a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images intended for output to a display. Modern GPUs are very efficient at manipulating computer graphics and image processing, and their highly parallel structure makes them more effective than general-purpose CPUs for algorithms where processing of large blocks of data is done in parallel.

Gradient descent Gradient descent is an optimization algorithm, that attempts to find local minima of a function by taking steps proportional to the negative of the gradient (or of the approximate gradient) of the function at the current point. Conversely, if instead one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent. This method is often used in conjunction with learning rules like the Hebbian rule, to improve parameters of an artificial neural network.

Hebbian rule The Hebbian theory of learning in neural networks proposes an explanation for the adaptation of neurons in the brain during the learning process. It posits that any two neural cells that are repeatedly active at the same time tend to become 'associated' so that activity in one facilitates activity in the other. This was ultimately formulated as a learning rule that can be applied to artificial neural networks.

k-means algorithm In unsupervised machine learning, k-means clustering aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Likert scale A Likert scale is a psychometric scale commonly involved in research that employs questionnaires. It is the most widely used approach to scaling responses in survey research. The scale is named after its inventor, psychologist Rensis Likert. When responding to a Likert questionnaire item, respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements. It is assumed that distances on each item in the scale are equal.

MS Manuscript

Neural Network In machine learning and cognitive science, artificial neural networks are a family of statistical learning algorithms inspired by biological neural networks and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown.

Retinotopy The mapping of visual input from the retina to neurons, particularly those neurons within the visual stream.

Segmentation In image processing, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Support Vector Machine In machine learning, support vector machines and networks are supervised learning models that are used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using kernel methods that map their inputs into high(er) dimensional feature space.

BIBLIOGRAPHY

- [1] John Romilly Allen. Notes on interlaced crosses. *Journal of British Archaeological Association*, 34:352–59, 1878.
- [2] Curtis A. Armstrong. Vectorization of Raster Images using B-Spline Surfaces. Master's thesis, Brigham Young University, 2006.
- [3] Rudolph Arnheim. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, 1954. URL <https://books.google.ie/books?id=68hz5phbLMwC>.
- [4] George Bain. *Celtic art: the methods of construction*. Dover Publications, 1951. ISBN 9780486229232. URL http://books.google.com/books?id=_1u9metwG60C.
- [5] Iain Bain. *Celtic Knotwork*. Celtic Interest Series. Constable, 1990. ISBN 9780094698109. URL <https://books.google.ie/books?id=0StqJgAACAAJ>.
- [6] P. Baines and A. Haslam. *Type and Typography*. Biebl Ranweiler portfolio series. Laurence King, 2005. ISBN 9781856694377. URL <https://books.google.ie/books?id=rs6gyb2hPF4C>.
- [7] Mauro Barni, Franco Bartolini, and Vito Cappellini. Image Processing for Virtual Restoration of Artworks. *IEEE Multimedia*, 7:34–37, 2000. ISSN 1070-986X. doi: 10.1109/93.848424. URL <http://doi.ieeecomputersociety.org/10.1109/93.848424>.
- [8] Sven Behnke. *Hierarchical Neural Networks for Image Interpretation*. Lecture Notes in Computer Science. Springer, 2003. ISBN 9783540407225. URL <http://books.google.ie/books?id=nj8z5NuG2SUC>.
- [9] Sven Behnke, Marcus Pfister, and Raúl Rojas. A Study on the Combination of Classifiers for Handwritten Digit Recognition. In *Proceedings of Third International Workshop on Neural Networks in Applications (NN'98)*, pages 39–46, 1998.
- [10] Clive Bell. *Art*. Chatto & Windus, London, 1914.
- [11] A. Bentkowska-Kafel, T. Cashen, and H. Gardiner. *Digital Art History: A subject in transition*. Computers and the history of art. Intellect, 2005. ISBN 9781841501161. URL <http://books.google.ie/books?id=KMLPAAAAAMAAJ>.

- [12] R.S. Berns. Rejuvenating the appearance of cultural heritage using color and imaging science techniques. *Proc. AIC Colour*, 5:369–374, 2005.
- [13] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. A Framework for Photo-quality Assessment and Enhancement Based on Visual Aesthetics. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 271–280, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-933-6. doi: 10.1145/1873951.1873990. URL <http://doi.acm.org/10.1145/1873951.1873990>.
- [14] Susan Bioletti, Rory Leahy, John Fields, Bernard Meehan, and Werner Blau. The examination of the Book of Kells using micro-Raman spectroscopy. *Journal of Raman Spectroscopy*, 40(8):1043–1049, 2009. ISSN 1097-4555. doi: 10.1002/jrs.2231. URL <http://dx.doi.org/10.1002/jrs.2231>.
- [15] G.D. Birkhoff. *Aesthetic Measure*. Harvard University Press, 2003. ISBN 9780766130944. URL <http://books.google.ie/books?id=F8A-UF8QgckC>.
- [16] David H. Brainard. The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436, 1997. doi: 10.1163/156856897X00357. URL <http://booksandjournals.brillonline.com/content/journals/10.1163/156856897x00357>.
- [17] Thomas M. Breuel. Two geometric algorithms for layout analysis. In *Document Analysis Systems V*, pages 188–199. Springer, 2002.
- [18] Thomas M. Breuel. High performance document layout analysis. In *Proceedings of the Symposium on Document Image Understanding Technology*, pages 209–218, 2003.
- [19] Robert Bringhurst. *The elements of typographic style*. Hartley & Marks Vancouver, BC, 1992.
- [20] T.J. Brown and C.D. Verey. Northumbria and the Book of Kells. *Anglo-Saxon England*, 1:219–246, 1 1972. ISSN 1474-0532. doi: 10.1017/S026367510000017X. URL http://journals.cambridge.org/article_S026367510000017X.
- [21] Cameron Browne. Wild knots. *Computers & Graphics*, 30(6): 1027 – 1032, 2006. ISSN 0097-8493. doi: 10.1016/j.cag.2006.08.021. URL <http://www.sciencedirect.com/science/article/pii/S0097849306001555>.
- [22] Michael Camille. Seeing and Reading: Some visual implications of medieval literacy and illiteracy. *Art History*, 8(1):26 – 49, 1985.

- ISSN 01416790. URL <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=19343024&site=ehost-live>.
- [23] Camilo J. Cela-Conde, Gisèle Marty, Fernando Maestú, Tomás Ortiz, Enric Munar, Alberto Fernández, Miquel Roca, Jaume Rosselló, and Felipe Quesney. Activation of the prefrontal cortex in the human visual aesthetic perception. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6321–6325, 2004. doi: 10.1073/pnas.0401427101. URL <http://www.pnas.org/content/101/16/6321.abstract>.
- [24] C. Cennini. *The Craftsman's Handbook*. Dover Art Instruction. Dover Publications, 2012. ISBN 9780486136622. URL <https://books.google.ie/books?id=4Z2jAQAQBAJ>.
- [25] Moran Cerf, Jonathan Harel, Wolfgang Einhaeuser, and Christof Koch. Predicting human gaze using low-level saliency combined with face detection. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 241–248. Curran Associates, Inc., 2008. URL <http://authors.library.caltech.edu/40642/>.
- [26] Anjan Chatterjee. Neuroaesthetics: a coming of age story. *Journal of Cognitive Neuroscience*, 23(1):53–62, 2011.
- [27] R.G. Collingwood. *The Principles of Art*. A Galaxy book. Oxford University Press, 1943. ISBN 9780195002096. URL <http://books.google.ie/books?id=SXpMC9416y4C>.
- [28] Frans W. Cornelissen, Enno M. Peters, and John Palmer. The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4):613–617, 2002. ISSN 0743-3808. doi: 10.3758/BF03195489. URL <http://dx.doi.org/10.3758/BF03195489>.
- [29] Antonio Criminisi, Martin Kemp, and Andrew Zisserman. Bringing pictorial space to life: computer techniques for the analysis of paintings. *Digital art history: A subject in transition*, 1, 2005.
- [30] Peter R. Cromwell. Celtic knotwork: Mathematical art. *The Mathematical Intelligencer*, 15(1):36–47, 1993. ISSN 0343-6993. doi: 10.1007/BF03025256. URL <http://dx.doi.org/10.1007/BF03025256>.
- [31] K. Cullen. *Design Elements, Typography Fundamentals: A Graphic Style Manual for Understanding How Typography Affects Design*. Design Elements. Rockport Publishers, 2012. ISBN 9781592537679. URL <http://books.google.ie/books?id=7Y4kvVbelLYC>.

- [32] Y. Le Cun, B. Boser, J.S. Denker, R.E. Howard, W. Hubbard, L.D. Jackel, and D. Henderson. Handwritten digit recognition with a back-propagation network. In David S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990. ISBN 1-55860-100-7. URL <http://dl.acm.org/citation.cfm?id=109230.109279>.
- [33] D.W. Cunningham and C. Wallraven. *Experimental Design: From User Studies to Psychophysics*. An A K Peters book. CRC Press, 2011. ISBN 9781439865514. URL <https://books.google.ie/books?id=Ui30BQAAQBAJ>.
- [34] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Studying Aesthetics in Photographic Images Using a Computational Approach. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 288–301. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-33836-9. doi: 10.1007/11744078_23. URL http://dx.doi.org/10.1007/11744078_23.
- [35] Whitney Davis. Neurovisuality. *Nonsite.org*, Jun 2011. URL <http://nonsite.org/issues/issue-2/neurovisuality>.
- [36] Catherine Dixon. Describing typeforms: a designer’s response. *InfoDesign–Revista Brasileira de*, 2008.
- [37] Frank Drewes. Tree-based picture generation. *Theoretical Computer Science*, 246(1):1–51, 2000.
- [38] Douglas Dunham. Hyperbolic Celtic Knot Patterns. In Reza Sarhangi, editor, *Bridges: Mathematical Connections in Art, Music, and Science*, pages 13–22, Southwestern College, Winfield, Kansas, 2000. Bridges Conference. ISBN 0-9665201-2-2. Available online at <http://archive.bridgesmathart.org/2000/bridges2000-13.html>.
- [39] Christian von Ehrenfels. Über gestaltqualitäten. *Vierteljahresschrift für wissenschaftliche Philosophie*, 14(3):249–292, 1890.
- [40] Hans Jürgen Eysenck. The empirical determination of an aesthetic formula. *Psychological Review*, 48(1):83, 1941.
- [41] E.C. Fernie. *Art history and its methods : a critical anthology*. Phaidon, 1996. URL <https://books.google.ie/books?id=Qb1euAAACAAJ>.
- [42] Katrin Franke and Mario Köppen. A computer-based system to support forensic studies on handwritten documents. *International Journal on Document Analysis and Recognition*, 3(4):218–

- 231, 2001. ISSN 1433-2833. doi: 10.1007/PL00013565. URL <http://dx.doi.org/10.1007/PL00013565>.
- [43] Katrin Franke, Lambert Schomaker, Christian Veenhuis, Christian Taubenheim, Isabelle Guyon, Louis Vuurpijl, Merijn van Erp, and Geertje Zwarts. WANDA: A generic Framework applied in Forensic Handwriting Analysis and Writer Identification. In *Proc. 3rd International Conference on Hybrid Intelligent Systems*, volume 105, pages 927–938, 2003.
- [44] Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):pp. 1189–1232, 2001. ISSN 00905364. URL <http://www.jstor.org/stable/2699986>.
- [45] Kunihiro Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980. ISSN 0340-1200. doi: 10.1007/BF00344251. URL <http://dx.doi.org/10.1007/BF00344251>.
- [46] Kunihiro Fukushima. Recognition of partly occluded patterns: a neural network model. *Biological Cybernetics*, 84:251–259, 2001. ISSN 0340-1200. doi: 10.1007/s004220000210. URL <http://dx.doi.org/10.1007/s004220000210>.
- [47] Kunihiro Fukushima. Neural Network Model Restoring Partly Occluded Patterns. In Vasile Palade, Robert J. Howlett, and Lakhmi Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 2774 of *Lecture Notes in Computer Science*, pages 131–138. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-40804-8. doi: 10.1007/978-3-540-45226-3_19. URL http://dx.doi.org/10.1007/978-3-540-45226-3_19.
- [48] Kunihiro Fukushima. Restoring partly occluded patterns: a neural network model. *Neural Networks*, 18(1):33 – 43, 2005. ISSN 0893-6080. doi: 10.1016/j.neunet.2004.05.001. URL <http://www.sciencedirect.com/science/article/pii/S0893608004000875>.
- [49] Kunihiro Fukushima and Masayuki Kikuchi. Symmetry axis extraction by a neural network. *Neurocomputing*, 69(16–18): 1827–1836, 2006. ISSN 0925-2312. doi: 10.1016/j.neucom.2005.11.010. URL <http://www.sciencedirect.com/science/article/pii/S092523120600141X>. Brain Inspired Cognitive Systems, Selected papers from the 1st International Conference on Brain Inspired Cognitive Systems (BICS 2004), 1st International Conference on Brain Inspired Cognitive Systems (BICS 2004).

- [50] Donncha Mac Gabhann. The division of hands in the Book of Kells: considering the evidence of the letter A in half-uncial in the text and uncial at line-ends. In *49th International Congress on Medieval Studies*, 2014.
- [51] A Garz, R. Sablatnig, and M. Diem. Layout Analysis for Historical Manuscripts Using Sift Features. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 508–512, Sept 2011. doi: 10.1109/ICDAR.2011.108.
- [52] Carlo Ginzburg and Anna Davin. Morelli, Freud and Sherlock Holmes: clues and scientific method. In *History Workshop*, number 9, pages 5–36. JSTOR, 1980. doi: 10.1093/hwj/9.1.5.
- [53] E.H. Gombrich. *Art and illusion: a study in the psychology of pictorial representation*. A.W. Mellon lectures in the fine arts. Phaidon, 1977. ISBN 9780714817569. URL <http://books.google.ie/books?id=pMQfAAAAIAAJ>.
- [54] E.H. Gombrich. *The Story of Art*. Phaidon Press, 16th edition edition, 1995. ISBN 9780714833552. URL <http://books.google.ie/books?id=rJFUAAAAMAAJ>.
- [55] J. Hawkins and S. Blakeslee. *On Intelligence*. Henry Holt and Company, 2005. ISBN 9780805078534. URL <http://books.google.com/books?id=Qg2dmntfxmQC>.
- [56] J.G. Hawthorne and C.S. Smith. *On Divers Arts: The Foremost Medieval Treatise on Painting, Glassmaking, and Metalwork*. Dover Art Instruction Series. Dover Publications, 1963. ISBN 9780486237848. URL <http://books.google.ie/books?id=MMiLTJqvYnYC>.
- [57] D.O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Taylor & Francis, 2005. ISBN 9781135631918. URL <https://books.google.ie/books?id=uyV5AgAAQBAJ>.
- [58] Paul Hekkert and Piet C.W. van Wieringen. The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica*, 94(2):117 – 131, 1996. ISSN 0001-6918. doi: 10.1016/0001-6918(95)00055-0. URL <http://www.sciencedirect.com/science/article/pii/0001691895000550>.
- [59] Françoise Henry. *Irish Art: In the Early Christian Period (to 800 A.D.)*. Irish Art. Cornell University Press, 1965. URL <https://books.google.ie/books?id=eKafAAAAMAAJ>.
- [60] Françoise Henry. *Irish Art During the Viking Invasions: 800-1020 A.D.* Irish Art. Cornell University Press, 1967. URL <https://books.google.ie/books?id=V6efAAAAMAAJ>.

- [61] Françoise Henry. *Irish art in the Romanesque period, 1020-1170 A.D.* Irish Art. Cornell University Press, 1970. URL <https://books.google.ie/books?id=gsHpAAAAMAAJ>.
- [62] Françoise Henry and John Kennedy. *The Book of Kells: reproductions from the manuscript in Trinity College, Dublin.* Thames and Hudson, 1974.
- [63] David H. Hubel and Torsten N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.
- [64] Tomohiro Ishizu and Semir Zeki. Toward A Brain-Based Theory of Beauty. *PLoS ONE*, 6(7):e21852, 07 2011. doi: 10.1371/journal.pone.0021852. URL <http://dx.doi.org/10.1371%2Fjournal.pone.0021852>.
- [65] Koichi Itoh and Yoshio Ohno. A curve fitting algorithm for character fonts. *Electronic publishing*, 6(3):195–205, 1993.
- [66] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254–1259, Nov 1998. ISSN 0162-8828. doi: 10.1109/34.730558.
- [67] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203, Mar 2001. ISSN 1471-003X. doi: 10.1038/35058500. URL <http://dx.doi.org/10.1038/35058500>.
- [68] Richard H.A.H. Jacobs, Remco Renken, and Frans W. Cornelissen. Neural Correlates of Visual Aesthetics - Beauty as the Coalescence of Stimulus and Internal State. *PLoS ONE*, 7(2):e31248, 02 2012. doi: 10.1371/journal.pone.0031248. URL <http://dx.doi.org/10.1371%2Fjournal.pone.0031248>.
- [69] Paul Jacobsthal. *Early Celtic art: Vol I.* Oxford, Clarendon Press, 1944.
- [70] Paul Jacobsthal. *Early Celtic art: Vol II.* Oxford, Clarendon Press, 1944.
- [71] V. Jain, J.F. Murray, F. Roth, S. Turaga, V. Zhigulin, K.L. Briggman, M.N. Helmstaedter, W. Denk, and H.S. Seung. Supervised Learning of Image Restoration with Convolutional Networks. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007. doi: 10.1109/ICCV.2007.4408909.

- [72] Viren Jain and Sebastian Seung. Natural Image Denoising with Convolutional Networks. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 769–776. Curran Associates, Inc., 2009. URL <http://papers.nips.cc/paper/3506-natural-image-denoising-with-convolutional-networks.pdf>.
- [73] Immanuel Kant. *Critique of the Power of Judgment*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge University Press, 2000. ISBN 9780521348928. URL <http://books.google.ie/books?id=JEXHicDbBDcC>. Translated by Guyer, P. and Matthews, E.
- [74] Matthew Kaplan and Elaine Cohen. Computer Generated Celtic Design. In *Proceedings of the 14th Eurographics Workshop on Rendering*, EGRW '03, pages 9–19, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association. ISBN 3-905673-03-7. URL <http://dl.acm.org/citation.cfm?id=882404.882406>.
- [75] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988. ISSN 0920-5691. doi: 10.1007/BF00133570. URL <http://dx.doi.org/10.1007/BF00133570>.
- [76] Won-Soon Kim and Rae-Hong Park. Color image palette construction based on the HSI color system for minimizing the reconstruction error. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, pages 1041–1044 vol.3, 1996. doi: 10.1109/ICIP.1996.561017.
- [77] E. Kindel. *Typeform dialogues: a comparative survey of typeform history and description*. Hyphen Press, 2008. URL <http://books.google.ie/books?id=QolWHAACAAJ>.
- [78] M. Kleiner, D.H. Brainard, and D. Pelli. What's new in Psychtoolbox-3? *Perception*, 36 ECVF Abstract Supplement, 2007.
- [79] Renate Klempien-Hinrichs and Caroline Von Totth. Generation of Celtic Key Patterns with Tree-based Collage Grammars. *Electronic Communications of the EASST*, 26, 2010.
- [80] Donald Ervin Knuth. Mathematical typography. *Bulletin of the American Mathematical Society*, 1(2), 1979.
- [81] Donald Ervin Knuth. *Digital typography*, volume 78. Csl Publications, 1999.
- [82] Ashutosh Kulkarni and David G. Stork. Optical or mechanical aids to drawing in the early Renaissance? A geometric analysis

- of the trellis work in Robert Campin's Mérode Altarpiece. In *Proc. SPIE 7251, Image Processing: Machine Vision Applications II*, volume 7251, pages 72510R–72510R–9, 2009. doi: 10.1117/12.806247. URL <http://dx.doi.org/10.1117/12.806247>.
- [83] Kuniyiko and Fukushima. Use of non-uniform spatial blur for image comparison: symmetry axis extraction. *Neural Networks*, 18(1):23 – 32, 2005. ISSN 0893-6080. doi: 10.1016/j.neunet.2004.08.001. URL <http://www.sciencedirect.com/science/article/pii/S089360800400142X>.
- [84] Vinciane Lacroix. Automatic Palette Identification of Colored Graphics. In Jean-Marc Ogier, Wenyin Liu, and Josep Lladós, editors, *Graphics Recognition. Achievements, Challenges, and Evolution*, volume 6020 of *Lecture Notes in Computer Science*, pages 61–68. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-13727-3. doi: 10.1007/978-3-642-13728-0_6. URL http://dx.doi.org/10.1007/978-3-642-13728-0_6.
- [85] Kevin Larson and R. Picard. The aesthetics of reading. In *Appears in Human-Computer Interaction Consortium Conference, Snow Mountain Ranch, Fraser, Colorado*, 2005.
- [86] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. ISSN 0018-9219. doi: 10.1109/5.726791.
- [87] Alana S. Lee, Peter J. Mahon, and Dudley C. Creagh. Raman analysis of iron gall inks on parchment. *Vibrational Spectroscopy*, 41(2):170 – 175, 2006. ISSN 0924-2031. doi: 10.1016/j.vibspec.2005.11.006. URL <http://www.sciencedirect.com/science/article/pii/S0924203105001517>. 6th Australian Conference on Vibrational Spectroscopy.
- [88] C.G. Leedham and V.K. Sagar. Using forensic handwriting analysis techniques to enhance automatic handwritten script recognition and processing. In *Handwriting Analysis and Recognition: A European Perspective, IEE European Workshop on*, pages 2/1–2/3, 1994.
- [89] Y. Leydier, F. Le Bourgeois, and H. Emptoz. Serialized unsupervised classifier for adaptative color image segmentation: application to digitized ancient manuscripts. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 494–497 Vol.1, Aug 2004. doi: 10.1109/ICPR.2004.1334174.
- [90] D.G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh*

- IEEE International Conference on*, volume 2, pages 1150–1157 vol.2, 1999. doi: 10.1109/ICCV.1999.790410.
- [91] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press. URL <http://projecteuclid.org/euclid.bsmsp/1200512992>.
- [92] Davide Massaro, Federica Savazzi, Cinzia Di Dio, David Freedberg, Vittorio Gallese, Gabriella Gilli, and Antonella Marchetti. When Art Moves the Eyes: A Behavioral and Eye-Tracking Study. *PLoS ONE*, 7(5):e37285, 05 2012. doi: 10.1371/journal.pone.0037285. URL <http://dx.doi.org/10.1371%2Fjournal.pone.0037285>.
- [93] A. Del Mastio, A. Piva, M. Barni, V. Cappellini, and L. Stefanini. Color transplant for reverse ageing of faded artworks. In David G. Stork and Jim Coddington, editors, *Proc. SPIE 6810, Computer Image Analysis in the Study of Art*, volume 6810, page 681006. SPIE, 2008. doi: 10.1117/12.765680. URL <http://dx.doi.org/10.1117/12.765680>.
- [94] MATLAB. *MATLAB Version 8.1 (R2013a)*. The MathWorks Inc., Natick, Massachusetts, 2013.
- [95] Heather McLeay and David Piggins. The mental manipulation of 2-d representations of knots as deformable structures. *Educational Studies in Mathematics*, 30(4):399–414, 1996. ISSN 0013-1954. doi: 10.1007/BF00570831. URL <http://dx.doi.org/10.1007/BF00570831>.
- [96] Aidan Meehan. *The Celtic Design Book*. Thames & Hudson, 1991.
- [97] Christian Mercat. Les entrelacs des enluminures celtes. *Pour la Science*, 1997.
- [98] J. Mutch and D.G. Lowe. Multiclass Object Recognition with Sparse, Localized Features. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 11–18, 2006. doi: 10.1109/CVPR.2006.200.
- [99] Jim Mutch and David G. Lowe. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80(1):45–57, 2008. ISSN 0920-5691. doi: 10.1007/s11263-007-0118-0. URL <http://dx.doi.org/10.1007/s11263-007-0118-0>.
- [100] Jim Mutch, Ulf Knoblich, and Tomaso Poggio. CNS: a GPU-based framework for simulating cortically-organized networks.

- Technical Report MIT-CSAIL-TR-2010-013 / CBCL-286, Massachusetts Institute of Technology, Cambridge, MA, February 2010.
- [101] NVIDIA. *NVIDIA CUDA Programming Guide 2.0*. NVIDIA Corporation, 2008.
 - [102] Lawrence O’Gorman. The document spectrum for page layout analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1162–1173, 1993.
 - [103] Felicity O’Mahony, editor. *The Book of Kells: proceedings of a conference at Trinity College, Dublin, 6-9 September 1992*, 1994. Published for Trinity College Library, Dublin by Scholar Press. ISBN 9780859679671. URL <http://books.google.ie/books?id=oyPrAAAAAAAJ>.
 - [104] Stephen E. Palmer, Karen B. Schloss, and Jonathan Sammartino. Visual Aesthetics and Human Preference. *Annual Review of Psychology*, 64(1):77–107, 2013. doi: 10.1146/annurev-psych-120710-100504. URL <http://dx.doi.org/10.1146/annurev-psych-120710-100504>. PMID: 23020642.
 - [105] M. Pappas and I. Pitas. Digital Color Restoration of Old Paintings. *Image Processing, IEEE Transactions on*, 9(2):291–294, 2000. ISSN 1057-7149. doi: 10.1109/83.821745.
 - [106] Elina Pihko, Anne Virtanen, Veli-Matti Saarinen, Sebastian Panasch, Lotta Hirvenkari, Timo Tossavainen, Arto Haapala, and Riitta Hari. Experiencing art: the influence of expertise and painting abstraction level. *Frontiers in human neuroscience*, 5, 2011.
 - [107] M.R. Poynton and L. Peltz. *History of Art: A Students’ Handbook*. Routledge, 1997. ISBN 9780415151818. URL <http://books.google.ie/books?id=TY6F7h1fpfIC>.
 - [108] V.S.N. Prasad and L.S. Davis. Detecting rotational symmetries. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 954–961 Vol. 2, 2005. doi: 10.1109/ICCV.2005.71.
 - [109] Donald Preziosi. *Rethinking art history: Meditations on a coy science*. Yale University Press, 1989.
 - [110] Toni M. Rath and R. Manmatha. Features for word spotting in historical manuscripts. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pages 218–222 vol.1, Aug 2003. doi: 10.1109/ICDAR.2003.1227662.

- [111] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2 (11):1019–1025, 1999.
- [112] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(1):23–38, 1998. ISSN 0162-8828. doi: 10.1109/34.655647.
- [113] Róisín Rowley-Brooke and Anil Kokaram. Bleed-through removal in degraded documents. In *Proc. SPIE 8297, Document Recognition and Retrieval XIX*, volume 8297, pages 82970T–82970T–10, 2012. doi: 10.1117/12.908911. URL <http://dx.doi.org/10.1117/12.908911>.
- [114] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, October 1986. URL <http://dx.doi.org/10.1038/323533a0>.
- [115] J. Schmidhuber. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2013). *Autonomous Mental Development, IEEE Transactions on*, 2(3):230–247, 2010. ISSN 1943-0604. doi: 10.1109/TAMD.2010.2056368.
- [116] Peter Selinger. Potrace: a polygon-based tracing algorithm (2003), 2008. <http://potrace.sourceforge.net>.
- [117] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 994–1000 vol. 2, 2005. doi: 10.1109/CVPR.2005.254.
- [118] Thomas Serre, Minjoon Kouh, Charles Cadieu, Ulf Knoblich, Gabriel Kreiman, and Tomaso Poggio. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical report, Massachusetts Institute of Technology, 2005.
- [119] Thomas Serre, Aude Oliva, and Tomaso Poggio. A feed-forward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, 104(15):6424–6429, 2007. doi: 10.1073/pnas.0700622104. URL <http://www.pnas.org/content/104/15/6424.abstract>.
- [120] Morteza Shahram, David G. Stork, and David Donoho. Recovering layers of brush strokes through statistical analysis of color and shape: an application to van gogh’s self portrait with grey felt hat. In *Proc. SPIE 6810, Computer Image Analysis in the Study*

- of Art*, volume 6810, pages 68100D–68100D–8, 2008. doi: 10.1117/12.765773. URL <http://dx.doi.org/10.1117/12.765773>.
- [121] A.D.R. Sheppard. *Aesthetics: An Introduction to the Philosophy of Art*. An OPUS book. Oxford University Press, 1987. ISBN 9780192891648. URL <https://books.google.ie/books?id=ZYCcAQAAQBAJ>.
- [122] Anikó Simon, J.-C. Pret, and A. Peter Johnson. A fast algorithm for bottom-up document layout analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(3):273–277, 1997.
- [123] David Small. Simulating watercolor by modeling diffusion, pigment, and paper fibers. In Walter R. Bender and Wil Plouffe, editors, *Proc. SPIE 1460, Image Handling and Reproduction Systems Integration*, volume 1460, pages 140–146. SPIE, 1991. doi: 10.1117/12.44417. URL <http://dx.doi.org/10.1117/12.44417>.
- [124] D.G. Stork. Computer Vision, Image Analysis, and Master Art: Part 1. *Multimedia, IEEE*, 13(3):16–20, Jul.–Sep. 2006. ISSN 1070-986X. doi: 10.1109/MMUL.2006.50.
- [125] D.G. Stork and M.F. Duarte. Computer Vision, Image Analysis, and Master Art: Part 3. *Multimedia, IEEE*, 14(1):14–18, Jan.–Mar. 2007. ISSN 1070-986X. doi: 10.1109/MMUL.2007.6.
- [126] D.G. Stork and M.K. Johnson. Computer Vision, Image Analysis, and Master Art: Part 2. *Multimedia, IEEE*, 13(4):12–17, Oct.–Dec. 2006. ISSN 1070-986X. doi: 10.1109/MMUL.2006.78.
- [127] James Sutton and Alan Bartram. *An atlas of typeforms*. Lund Humphries, 1968.
- [128] A. Tonazzini. Color space transformations for analysis and enhancement of ancient degraded manuscripts. *Pattern Recognition and Image Analysis*, 20(3):404–417, 2010. ISSN 1054-6618. doi: 10.1134/S105466181003017X. URL <http://dx.doi.org/10.1134/S105466181003017X>.
- [129] Anna Tonazzini, Emanuele Salerno, Matteo Mochi, and Luigi Bedini. Bleed-Through Removal from Degraded Documents Using a Color Decorrelation Method. In Simone Marinai and Andreas R. Dengel, editors, *Document Analysis Systems VI*, volume 3163 of *Lecture Notes in Computer Science*, pages 229–240. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-23060-1. doi: 10.1007/978-3-540-28640-0_22. URL http://dx.doi.org/10.1007/978-3-540-28640-0_22.
- [130] G. Vasari and G. Bull. *Lives of the Artists*. Number v. 1 in Penguin classics. Penguin Books Limited, 2003. ISBN

9780141919973. URL <https://books.google.ie/books?id=40LwqoX0gPoC>.
- [131] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. ISSN 0920-5691. doi: 10.1023/B:VISI.0000013087.49260.fb. URL <http://dx.doi.org/10.1023/B%3AVISI.0000013087.49260.fb>.
- [132] Christian Wallraven, Douglas W. Cunningham, Jaume Rigau, Miquel Feixas, and Mateu Sbert. Aesthetic Appraisal of Art-from Eye Movements to Computers. In *Computational aesthetics*, pages 137–144, 2009.
- [133] N. Warburton. *Philosophy: The Basics*. The Basics. Taylor & Francis, 2004. ISBN 9781135151911. URL <https://books.google.ie/books?id=1XiaAgAAQBAJ>.
- [134] Baogang Wei, Yonghuai Liu, and Yunhe Pan. Using Hybrid Knowledge Engineering and Image Processing in Color Virtual Restoration of Ancient Murals. *IEEE Transactions on Knowledge and Data Engineering*, 15:1338–1343, 2003. ISSN 1041-4347. doi: 10.1109/TKDE.2003.1232282. URL <http://doi.ieeecomputersociety.org/10.1109/TKDE.2003.1232282>.
- [135] J.J. Winckelmann and H. Fuseli. *Reflections on the Painting and Sculpture of the Greeks: With Instructions for the Connoisseur, and an Essay On Grace in Works of Art*. A. Millar and T. Cadell, 1767. URL <https://books.google.ie/books?id=0DXSAAAAMAAJ>.
- [136] J.J. Winckelmann and A. Potts. *History of the Art of Antiquity*. Texts & documents. Getty Research Institute, 2006. ISBN 9780892366682. URL <https://books.google.ie/books?id=mbWo3EaeGPEC>.
- [137] Lior Wolf, Nachum Dershowitz, Liza Potikha, Tanya German, Roni Shweka, and Yaacov Choueka. *Kodikologie und Paläographie im digitalen Zeitalter 2 - Codicology and Palaeography in the Digital Age 2: Schriften des Instituts für Dokumentologie und Editorik 3*, chapter Automatic Palaeographic Exploration of Genizah Manuscripts, pages 157–179. Norderstedt: Books on Demand, Germany, 2011.
- [138] H. Wölfflin. *Principles of Art History: The Problem of the Development of Style in Later Art*. Dover Publications, 1915. URL <https://books.google.ie/books?id=JrWloAEACAAJ>.
- [139] Lai-Kuan Wong and Kok-Lim Low. Saliency-enhanced image aesthetics class prediction. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 997–1000, 2009. doi: 10.1109/ICIP.2009.5413825.

- [140] Sam T.S. Wong, Howard Leung, and Horace H.S. Ip. Model-based analysis of Chinese calligraphy images. *Computer Vision and Image Understanding*, 109(1):69 – 85, 2008. ISSN 1077-3142. doi: 10.1016/j.cviu.2007.03.001. URL <http://www.sciencedirect.com/science/article/pii/S1077314207000422>.
- [141] Hagit Zabrodsky, Shmuel Peleg, and David Avnir. Symmetry as a continuous feature. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(12):1154–1166, 1995.
- [142] Song-Hai Zhang, Tao Chen, Yi-Fei Zhang, Shi-Min Hu, and R.R. Martin. Vectorizing Cartoon Animations. *Visualization and Computer Graphics, IEEE Transactions on*, 15(4):618–629, 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.9.