Self Organising Map based Method for Understanding High Aesthetic Value Abstract Image Features

A minor thesis submitted in partial fulfilment of the requirements for the degree of Master of Computer Science (Intelligent Systems)

Allan Campbell
School of Computer Science and Information Technology
Science, Engineering, and Technology Portfolio,
Royal Melbourne Institute of Technology
Melbourne, Victoria, Australia

November 4, 2013

Declaration

This thesis contains work that has not been submitted previously, in whole or in part, for any other academic award and is solely my original research, except where acknowledged.

This work was carried out between July and November 2013, under the supervision of Associate Professor Vic Ciesielski.

Allan Campbell
School of Computer Science and Information Technology
Royal Melbourne Institute of Technology
November 4, 2013

Acknowledgements

Thanks to Perry Barile for providing the program used to calculate feature values for all sets of images analysed.

Thanks to Karen Trist for her comments.

Contents

List of Figures

List of Tables

Abstract

The goal of this project was to develop a method that: takes the raw pixel data of images as input, allows the aesthetic separability of a set of images to be evaluated independently of any set of features, provides for the analysis of a number of computed features by visual inspection to determine their relevance to the aesthetic value of the images analysed, and for features determined to be relevant, to quantify the range of feature values that correlate with high and low aesthetic images. The method was developed in three stages, using images of progressively greater complexity. First the method was applied to a synthetic problem case designed to illustrate the method with low complexity images. Second the method was applied to a set of images of moderate complexity and the details of the method expanded upon. Lastly the method was applied to a difficult problem involving complex images of limited size. The method uses the Self Organising Map to project raw pixel data of images onto a feature map. The aesthetic class of these images is overlayed on the feature map, yielding a semantic map. The semantic map allows the aesthetic separability of the images to be evaluated by visual inspection. Feature values computed from the images are then averaged within the clusters in the semantic map and visualized in gray-scale heat maps. The heat maps are used to identify features relevant to aesthetic value by visual inspection. We call this the Pixel Array Self Organising Map (PASOM) method. For the set of images analysed, brightness features were identified by the PASOM method as being discriminatory between high and low aesthetic images, consistent with findings in prior work. High aesthetic value images tend to have higher brightness. In addition higher values for features measuring image texture at various scales was shown to be associated with high aesthetic value images. These findings were corroborated by a professional artist/photographer as being consistent with the principles for attaining aesthetic value in visual media. The PASOM method yields a semantic map and a visualization of feature value variation across the semantic map that together make possible a detailed analysis of features associated with aesthetic value of images, not presently available by any other method to our knowledge.

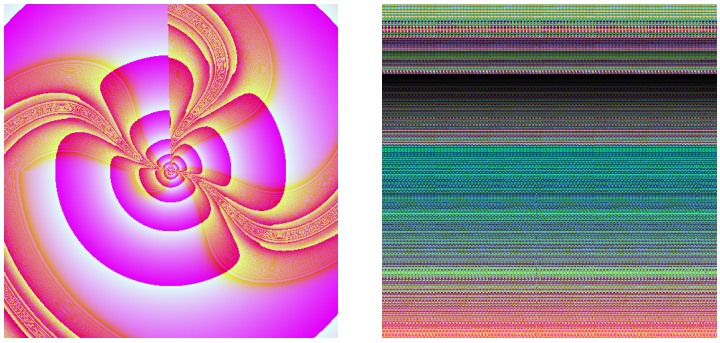
Keywords: Self Organising Map, Computational Aesthetics, Abstract Images, Feature Analysis

Chapter 1

Introduction

The struggle to create beautiful things is an essential element of design in industry. If there is some kind of order in what we perceive that gives rise to what we experience as aesthetic value and if we can understand this order, then perhaps we can automate the creation of high aesthetic value artifacts. Working towards this goal is a valuable endeavour, but having computers make aesthetic distinctions comparable with the subtlety of human judgement is an open problem. Aesthetics in computer generated art is perhaps a domain in which we can explore this challenge. IMAGENE [37] is an evolutionary art generating program, in which a human-in-the-loop supplies a fitness function by making aesthetic selections that guide the generation of further images. The result of an interactive run of the program is a set of images that have been shown in prior work [5] to have significantly varying degrees of aesthetic appeal. Examples of such are shown in Figure 1.1.

The challenge exists to remove the requirement for a human-in-the-loop and replace the manual fitness function with one or more encoded aesthetics, so that art generated by the program more narrowly exhibits a high aesthetic appeal as defined by the norms of the programmed aesthetic. How could an aesthetic be encoded in a machine to enable the expression of that aesthetic to be automated? What are possible laws of aesthetic appeal? Researchers hypothesise that aesthetic value is determined on two levels: high level semantic content (what we understand an image to represent) and low level features (visual properties such as color, texture, image size) [29]. The problem of machine comprehension of the semantic content of images is itself an open problem. In this work we seek to further understanding of how low level visual properties of images contribute to their aesthetic value.



High aesthetic image Low aesthetic image

Figure 1.1: IMAGENE examples

A common approach in machine vision and machine learning problems that take images as input is to compute measures of visual properties termed **features** from the raw pixel data constituting an image. The resulting set of values quantify characteristic features of an image and these feature values are input to vision and learning algorithms as representative of the image itself. The variety of possible features that could be calculated is limited only by the imagination. For example a simple feature might measure the color of the single brightest pixel of an image, another feature might measure the coordinates in the image of the single brightest pixel. One issue with this approach is that we do not know which of these thought up features alone or in combination with other imaginative features is optimal or even relevant for the characterisation of an image in the context of a particular application. A set of features that may be relevant to object recognition may have no relevance for predicting the aesthetic rating of abstract art images. The ability to discriminate between features of greater relevance from those of lesser or no relevance in the context of a particular application is desirable. In relation to computational aesthetics under study in this work, it is also desirable to ascertain how the relative values of features determined to be relevant, correlate to aesthetic appeal. Here we describe the traditional method and a novel method for approaching these questions:

Traditional Method

Collect many images that each have either high or low appeal and calculate some imaginative features for these labelled training images. Use this feature data to train a classifier and use the classifier to show that if we know these feature values for any image then the classifier can predict the level of appeal of that image. To determine which of these features is most important, apply attribute selection algorithms to rank the features by their contribution to the predictive power of the classifier.

A limitation of the traditional method is that a ranking of features gives little indication of their individual relevance. Even features that are weakly associated with aesthetic value will

be ranked highly in the absence of more relevant features. A more fundamental limitation of the traditional method is that it can offer no direct analysis of images but must rely on features which have been derived from those images. The derivation of features is likely to entail the loss of information encoded in the raw pixel data of images. Also there is the potential to distort the characteristics of the images by representing them as a set of values computed by algorithms of arbitrary design. The quality of the conclusions that can be drawn for a set of features is therefore completely dependent on the quality of the derived features themselves.

Novel Method

Collect many images that each have either high or low appeal. Apply a clustering algorithm to the raw pixel representations of the images and group together those images that are most similar. We determine similarity by a Euclidian distance measure computed from the images as high dimensional vectors of pixel intensities. The emphasis here is that clustering is performed on the raw pixel data and independently of explicit feature values and class labels. If a significant majority of the images in each of the clusters have similar aesthetic ratings, then we can say that the way we calculate the similarity of images from pixels, captures aesthetically discriminating properties in the images independently of any features. Now label each cluster as a high or low aesthetic cluster according to its majority class. Similarity can be computed not only within clusters but between clusters. The arrangement of clusters relative to each other can be viewed as a topology in which those clusters that are closer are more similar (on the basis of Euclidian distance between cluster centroids computed in the space of pixel intensities). In the novel method, this cluster topology is viewed in 1 dimension, ie all cluster centroids are projected onto a straight line and on this straight line adjacent clusters are considered to be more similar than clusters further away. The order of the majority class labels across the topology is termed a **semantic map**. If the semantic map shows a meaningful pattern of juxtaposition of clusters having the same class label, then we say that for the set of images under study, the topology of the semantic map is a projection of an aesthetically discriminating manifold in the high dimensional space of pixel intensities.

We now overlay feature values on this topology in order to make inferences about the features. We average the values of each feature for images within each cluster and visualize these feature averages across the cluster topology. We infer feature relevance from the pattern by which a feature average maps to the topology. If feature values can be shown to map to the cluster topology in a meaningful way, then there is a consistency between the way the clustering process expresses aesthetic discrimination and the way the feature value varies across the topology. We should consider such a feature to have relevance to the discrimination of images

by aesthetic value, whereas if feature values vary randomly or less meaningfully across the topology then there is an inconsistency between variation of feature values and the way the clustering process expresses aesthetic discrimination. We should question the relevance of such features to aesthetic discrimination.

As a metaphor for the novel method, imagine a pile of beads. Each bead is either black or white. We toss the beads into the air and at their apex we freeze time. Now we put on sunglasses so we cannot tell white from black beads and we look for clumps of beads in the stationary cloud. We place a number of red beads at positions in the cloud corresponding to the centres of gravity of clumps of beads. We then find the two red beads furthest apart, select one, weave a string through this bead and then through the next closest red bead and so on until the string connects all red beads. The order of the red beads on the string is what the self organising map gives us for a one dimensional feature map. Now we take off the sun glasses and change the color of each red bead to match the color of the majority color in its clump. If the pattern of white and black beads along the string shows a trend from mostly white to mostly black then we can say that the path of the string through the cloud is a path from white to black, ie a one dimensional manifold that discriminates white from black along its length. Now we measure some feature of all the beads. Imagine the beads have various diameters. We average the diameter of beads in each clump and visualize these averages along the string. If the average diameter along the string is from mostly big to mostly small in the direction of white to black then we can say that the diameter feature is discriminatory and that larger diameter correlates with whiteness.

In the novel method we exploit the cluster topology to analyse visual features of images by computing the average feature value in each cluster and mapping the cluster averages across the cluster topology. Mapping cluster averages of feature values to this topology allows us to do two things:

1. Identify which features have values that vary consistently with the semantic map across the cluster topology and hence are class discriminatory.

2. Identify how feature values correlate with class.

Since the SOM is at the heart of the novel method and a key idea motivating this work is the determination of image clusters independently of explicit features and class labels by clustering on the raw image data represented in pixel arrays, we refer to the novel method as the Pixel Array SOM (PASOM) method.

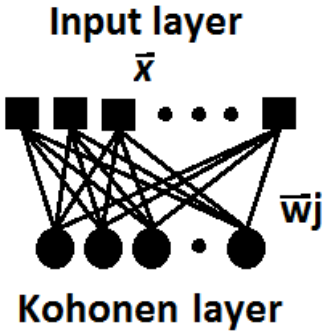


Figure 1.2: Self Organising Map Model

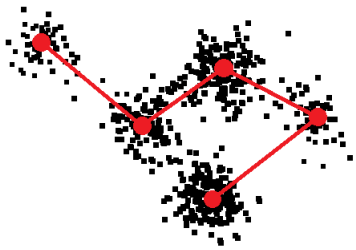


Figure 1.3: SOM Topology Preservation

1.1 The Self Organising Map (SOM)

The SOM algorithm falls within the unsupervised machine learning paradigm which is designed to find structure in input data without examples of known structure or class. The SOM model is an artificial neural network model comprised of a layer of input nodes and a layer of nodes called the feature map or the Kohonen layer after the algorithms inventor Teuvo Kohonen. Figure 1.2 shows a representation of the SOM model that takes an input vector $\mathbf{x}$ at the input layer. Each node in the Kohonen layer is connected to the input layer by a set of weighted links w_j. Competitive learning is a subcategory of unsupervised learning in which each input instance is associated with a single **winner-take-all** node in the network. In the SOM a learning rule updates the weight vector for this winning node and its neighbouring nodes in the Kohonen layer to more closely approximate the input instance. This results in a key characteristic of the SOM which is the preservation of the topology of the input data in the lower dimensional feature map. The model is inspired by examples from biology such as the formation in the auditory cortex of a tonotopic map of audible frequencies. The SOM is thus more than a clustering algorithm. High dimensional data can be projected onto a low dimensional feature map so that adjacency relationships between cluster centroids in the high dimensional space are preserved in the low dimension feature map. Figure 1.3 shows a representation of the projection of a 2 dimensional input space onto a 1 dimensional feature map and the preservation of the input data topology on the feature map.

Following is the conceptual operation of the algorithm:

1. Initialize: Set weight vectors w_j to small random values.
2. Stimulus: Apply input vector $\mathbf{x}$ selected at random from the training data.
3. Response: Compute the winning neuron.

4. Adaption: Update neuron weights.

5. Goto 2.

In Step 3, the winning neuron is selected by minimum Euclidian distance between the input vector $\mathbf{x}$ and the weight vectors w_j of neurons in the Kohonen layer:

$$c = argminj||\mathbf{x} - \mathbf{w_j}|| \tag{1.1}$$

c Is the index of the winning neuron in the Kohonen layer.

In Step 4, a generalized weight update rule is applied:

$$\mathbf{w_j}(t) = \mathbf{w_j}(t-1) + B(t)N(d,t)[\mathbf{x} - \mathbf{w_j}(t-1)] \tag{1.2}$$

B(t) is a scalar learning rate that decays as a function of the training epoch number. Examples of learning rate functions are:

$$B(t) = B_o(1 - t/T) \tag{1.3}$$

$$B(t) = B_o e^{-t/T} \tag{1.4}$$

- t is current epoch number
- T is total number of training epochs
- B_o is an initial learning rate

N(d,t) is a scalar neighbourhood learning rate function. An example of such is:

$$N(d,t) = e^{-d_{ij}^2/2s(t)^2} \tag{1.5}$$

- d_{ij} is the scalar distance in the Kohonen layer between the winning neuron and a neighbour neuron having updates to its weight vector.
- s(t) is a scalar neighbourhood size function that decays as a function of the training epoch number.

Examples of neighbourhood size functions are:

$$s(t) = S_o(1 - t/T) \tag{1.6}$$

$$s(t) = S_o e^{-t/T} \tag{1.7}$$

- S_o is an initial neighbourhood size

For this project we utilized the open source WEKA machine learning application together with the **SelfOrganisingMap** module (version 1.0.3) installable through the package manager [17]. We ran the SOM module with all standard default settings except the feature map dimensions.

Noted Weka SOM module defaults:

- Initial learning rate = 1.0
- Total epochs = 3000

1.2 Research Questions

For the purposes of this research we define image **aesthetics** as: the subjective rating of images for their emotional value which we refer to as visual appeal, allure or beauty. The aim of this thesis is the development of a method to analyse and understand features that distinguish between abstract images of high and low aesthetic value. We articulate this aim in the following research questions:

1. How can we identify features that have high relevance to abstract image aesthetics?
2. How do the values of high relevance features correlate to aesthetic appeal?

1.3 Our Contributions

The method proposed by this thesis makes the following contributions:

- Allow the pixel data of images to be analysed independently of any set of features.

- Allow the aesthetic separability of a set of images to be evaluated.
- Provide analysis of a number of features by visual inspection to determine their relevance to aesthetic discrimination.
- For features relevant to the aesthetic value of images, quantify the range of values that correlate with high and low aesthetic value.

1.4 Thesis Structure

Having described the PASOM method in the introduction and articulated the research questions, the following chapter will address the literature review. The body of the thesis will then expand upon the details of the method in three stages, applying the method to successive problem cases of increasing complexity. The first stage (Chapter 3) will be an illustration case. This is an orchestrated example designed to demonstrate the method and to show that the outcome expected for an artificial input dataset is achieved. The second stage (Chapter 4) is the application of the method to a problem of moderate complexity. This problem case does not have a deliberately biased input and hence we do not have an expected outcome. We use the method to investigate the relevance of a set of features for the task of discriminating between images of certain pattern types and in so doing we expand on the details of the method. The third stage (Chapter 5) is the application of the PASOM method to a complex problem that has been addressed by prior published work and in which we are primarily concerned by the outcome of the method rather than demonstrating the method. The closing chapter will address our conclusions.

Chapter 2

Literature Review

The literature review is structured around two themes: computational aesthetics and neural networks as a sub-field of machine learning applied to image mining.

2.1 Computational Aesthetics

Computational aesthetics has two themes: content generation and content evaluation [16]. By far the greater body of the computational aesthetics literature comprises work focused on the generation of image, video and audio content mostly by evolutionary algorithms. These implicitly require evaluation by some fitness function, so the former theme subsumes the latter. However for this study we are not interested in the generation of content and focus only on the evaluation problem. It is necessary to make a distinction not always explicit in the literature that computational aesthetics is distinct from image quality assurance. The latter field is a well developed area that has sought to evaluate degradation due to lossy encoding and compression algorithms. There is the danger of confusion here since image compression is sometimes used in aesthetic evaluation to estimate image complexity [28]. Our focus is not quality degradation due to processing artefacts, but rather the subjective rating of images for their emotional value.

Approaches to Aesthetic Evaluation

Aesthetic evaluation of media by computational means is achieved by three methods, first by supplying algorithms with human rated examples, second by encoding rules according to some theory of aesthetics which an algorithm can apply automatically [29]. A third approach described in the literature is labelled as **emergent**, wherein computational algo-

rithms are allowed to invent their own aesthetics. Galanter reports that these products of **meta-aesthetics** explorations are often perceived by humans as alien [16], but ironically these programs are arguably truer artists than the programs that are designed to proxy the human aesthetic experience. This issue goes to the heart of a long held tension in computational aesthetics between the founding principles of an empirical, reducible science/technology and the conviction for some that art is a manifestation of a free expressive creativity that is fundamentally non empirical, irreducible and non automatable. See [30] and [7] for a discussion.

Human aesthetic evaluations are often applied in an on-line mode during content generation using a human-in-the-loop as an evolutionary computation fitness function. Dawkins in 1986 is an early example [10] and there are more recent examples [37]. Human ratings may also be applied in a crowd-source mode such as when content is uploaded to a website and subsequently rated [Eg http://www.dpchallenge.com/]. Likewise there are various examples of the method of encoding rules. Harold Cohen's AARON is an early and famous example [6].

Motivations and Limitations

The attraction of the rule based approach should be obvious. Human in the loop is labour intensive, slow, having limited throughput, can be expensive and is subject to problems of inconsistency. Humans have variable subjective standards and often the same person does not judge a given work the same way on subsequent evaluations. An approach to dealing with the variability in human aesthetic ratings is often to select only the highest and lowest rated images and formulate a machine learning task for binary classification [5] [26]. The crowd-sourced approach has its limitations also such as deliberate misrepresentation [36].

Motivations for developing automated aesthetic analysis however go beyond these factors and extend to such reasons as: advancing content based image retrieval technology and recommender systems, commercial evaluation of artistic intellectual property, marketing decision processes, personal library management, understanding more about aesthetics and even biological and physiological research in human vision [26]. For content based information retrieval in particular, it is desired to make available the aesthetic element of media as a parameter by which to select, filter and present information in accordance with the objectives of information consumers. To this end, research on the topic of computational aesthetics has been underway since at least 1992 [19] and since 2009 has been accelerating.

Datta in 2009 introduced the concept of the **aesthetics gap** as a parallel to the notion of **semantics gap** already in common usage by then [8]. Datta also introduced the approach

to bridging the aesthetics gap by the use of machine learning. Computational aesthetics as a field of study was defined comprehensively in 2005 by Hoenig [18] stressing three important factors: computation methods, the human aesthetic point of view and the pragmatic need to focus on objective approaches. The wisdom of these insights by Hoenig arguably laid the groundwork for the seminal work of Datta et al and those that have followed.

Photography and Art Aesthetics

By far the greater majority of work in the literature relates to the aesthetic evaluation of photographic images as opposed to evaluation of art images. There is crossover in terms of computed image features and machine learning techniques. Studies in art aesthetics borrow heavily from studies of photographic aesthetics, for example [5] and [9]. The latter bases features on the former work. There are three further example papers known to this author that invent image features that other researchers have used for further work: Datta et al [8], Reaves [32] and Ke et al [20].

In all examples we are aware of, research that does specifically address computational aesthetics with respect to art work, selects a very specific style of art as its focus. For example Z. Les [25] examines the post impressionist works of Mondrian, while Congcong Li and Tsuhan Chen [26] address selected impressionist and landscape paintings. A number of studies examine abstract art works generated by evolutionary algorithms, for example Ciesielski [5] and Heijer and Eiben [11].

Features as Foundations of Aesthetics

Aesthetic judgement is hypothesised to spring from two sources in art work: content and features [29]. The idea that aesthetic value is connected to low level features goes back at least to Birkhoff in the 1930's [35] [4]. Birkhoff presented a notion of aesthetic measure based on order and complexity. Birkhoff's ideas, groundbreaking at the time have been variously extended [34] and disputed. See [16] for a discussion. More recently the idea that low level features underpin aesthetics has been presented by Arnheim in his 1971 book, Art and visual perception: A psychology of the creative eye [1]. Machado and Cardoso propose that these two aspects (content and features) are independent of each other [29], and Cohen makes the point that content easily overrides features. "A formal colour evaluation that doesn't take account of the possibility that all the well-balanced colour harmonies may add up to a portrait of an oddly-dressed man making rude hand gestures..." [7, page 99]. Machado and Cardoso venture that the relationship between content and aesthetics is more a cultural dependency than a universal truth [28]. Reinecke directly tests this idea related to user interfaces and finds that the performance of users increases when the aesthetics of the user

interface is aligned with their own cultural norms [33].

Much work in the area of computational aesthetics rests implicitly on the notion that the aesthetic value of an image can be deduced from the feature level alone. It is impressive that this has been shown to be true.

2.2 Neural networks

Machine learning applied to image processing is one aspect of image mining. Two themes characterise this work: mining databases of images themselves and mining databases of images associated with textual information [15]. Qingyong Li, Siwei Luo and Zhongzhi Shi take the latter approach to the topic of image aesthetics [27]. For our study we focus only on the former theme, mining images themselves. Image datasets are mined using classification, clustering and association rule mining tasks. Decision trees and neural networks are the dominant algorithm families applied to these tasks [15].

Image Mining

We might think of image mining as a higher level problem category than image processing. There is a large body of literature on the use of neural networks for image processing. Two excellent survey papers written a decade apart and each approximately summarizing their preceding decade are [31] in 2012 and [12] in 2002. The 2012 paper gives a summary of 10 families of ANN architectures and the types of applications in which each tends to perform well. According to this study, the self organising feature map (SOM) architecture is applied primarily to medical image segmentation and color processing. Feed forward networks are applied primarily to biologic image processing and automotive traffic control and security applications. Reasons for these groupings are not identified. Based on their findings as at 2012, the authors predict that the focus in neural networks for image processing going forward will be self organising networks, cellular networks and pulsating and oscillatory networks.

Labelled Data

An obstacle with the traditional (supervised learning) approach to machine learning is the "human bottleneck". Supervised learning requires labelled data, ie human labour is required to interpret and categorize examples that can be presented to machine learning algorithms. Ever more complex patterns require vastly more data in order to learn more subtle distinctions. The age of "Big Data" and the "Internet of Things" promises to yield ever more data but humans are expensive to employ and also limited in their capacity to process data

in volume. It is this bottleneck together with notions of biological plausibility regarding unsupervised learning that has focused increasing attention towards unsupervised learning methods.

Unsupervised Machine Learning

The self organising feature map (SOM) is a well known unsupervised neural network architecture first introduced by Teuvo Kohonen in 1982 [22] and still under active research to the present, see [23] published in early 2013. The SOM is a popular unsupervised neural network architecture that has been subject to many extensions, elaborations and enhancements. An example is [24] which is a study of self organising algorithms and the introduction of a new family of SOTM (Self Organsing Tree Map) architectures. Another example is [21]. Abstract feature maps encode features where a feature has no direct correspondence to a particular neuron but is represented by a group of neurons or connection weights. The neocognitron by Fukushima in the 1980's combined supervised and unsupervised learning elements [14]. The neocognitron was a follow on development from the cognitron in 1975 [13] adding feature invariance in scale, translation and rotation. The neocognitron was followed notably by the work of Benke and Rojas [2], Rojas claiming the NAP (Neural Abstraction Pyramid) to be what the Neocognitron attempted and failed to be.

In summary, unsupervised neural network architectures for image mining using feature sets and raw pixel arrays for semantic and aesthetic analysis are active areas of research, anticipated to intensify as the volume of unlabelled data grows in the future.

Chapter 3

The PASOM Method: Low Complexity Images for Illustration

The Pixel Array Self Organising Map (PASOM) method is the central contribution of this thesis. An outline of the method was presented in chapter 1. In this chapter we present an illustrating example of the application of the method and show the operation of the method on a synthetic problem. The PASOM method was applied to a problem purposely constructed to show that the method operates as expected.

To illustrate the PASOM method, two classes of images were generated having an explicit difference based on some characteristic decided a-priori and a second set of two classes of images were generated lacking this explicit difference. The characteristic decided a-priori was pixel intensity. The PASOM method was applied to these data sets and a suite of 55 features originated in prior work [9] was analysed. The suite of 55 features is shown in table 3.1.

In the first set of images a feature that measures pixel intensity is demonstrated to be discriminatory, while in the second set the feature cannot discriminate. The two data sets are:

1. Bright vs dark gray-scale images
2. Bright red vs bright blue images

We expect that a feature that measures pixel intensity will be discriminatory for bright vs dark gray-scale images, but not discriminatory for bright red vs bright blue images since dark and light images have opposite levels of pixel intensities whereas bright images of any color have the same level of pixel intensity.

Feature	Description
F01	Colorfulness - Earth Mover Distance from unsaturated grey
F02, F03, F04	Average hue, saturation, brightness over the entire image
F05, F06, F07	Average hue, saturation, brightness in the centre of the image
F08, F09	Not Implemented
F10 - F21	Wavelet functions computing image smoothness at various scales
F22, F23	Image dimensions (width+height, width/height)
F24	The number of contiguous regions based on color similarity larger than 1/100th of the total number of pixels in the image
F25	Not Implemented
F26 - F40	Average hue, saturation and brightness for each of the 5 largest contiguous regions of similar colors
F41 - F45	Size in pixels of each of the 5 largest regions of similar contiguous colors divided by the total number of pixels in the image
F46, F47	Two variations on the measure of complimentary colors
F48 - F52	The location in the image of the centre of each of the 5 largest contiguous regions of similar colors
F53 - F55	Depth of field effect (emulating telephoto lens zoom) on each of the hue, saturation and brightness channels

Table 3.1: Suite of 55 Features originated in prior work

For the first data set, 50 dark gray images of 8 x 8 pixel resolution were generated having a random pixel intensity below 32. All three color components of all pixels in the image have the same value. Also, 50 light gray images of 8 x 8 pixel resolution were generated having a random pixel intensity above 255-32 = 223. For the second data set 50 bright red images of 8 x 8 pixel resolution were generated having a random pixel intensity above 255-32 = 223 for the red component only with the green and blue components set to 0. Also 50 bright blue images of 8 x 8 pixel resolution were generated having a random intensity above 233 for the blue component only with the red and green components set to 0. Figure 3.1 shows example images.

We used the SelfOrganisingMap module in the open source WEKA machine learning application [17] to compute the SOM neural network model for these two cases. We ran the module with all standard default settings except the feature map dimensions which where set to width = 1 and height = 16 for a one dimensional 16 neuron feature map. Initial learning

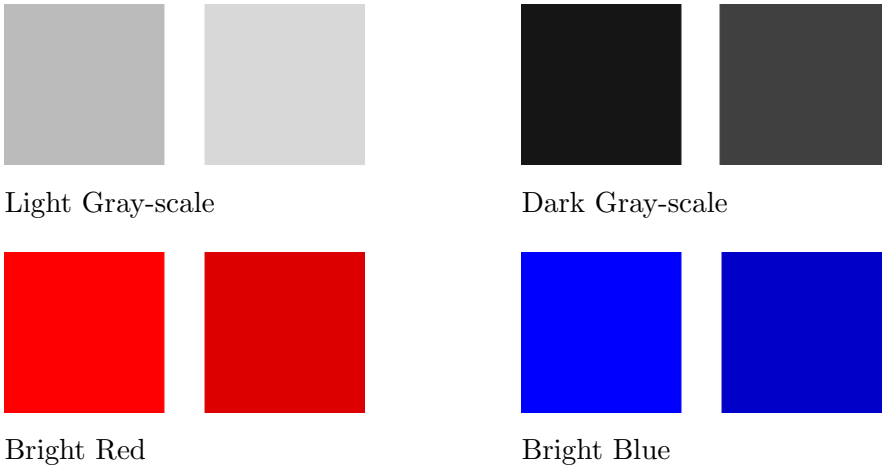


Figure 3.1: Example Low Complexity Images as Input

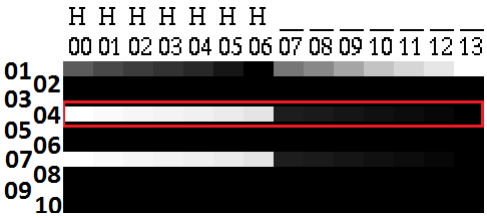


Figure 3.2: Output Heat-map for Grey-scale Images

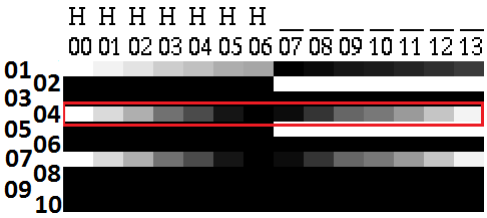


Figure 3.3: Output Heat-map for Red-Blue Images

rate was 1.0 and total training epochs was 3000.

The output of the PASOM method for each case is shown in figures 3.2 and 3.3. These figures show the results of clustering and the mapping of feature values across the cluster topology, depicted visually using grey-scale heat-maps. Each column corresponds to a cluster. Each row corresponds to a feature. Feature values are averaged in each cluster and cluster averages are mapped to a gray-scale intensity where black corresponds to the minimum value of all cluster averages for that feature and white corresponds to the maximum of all cluster averages for that feature. A row that is all black indicates that all cluster averages are equal, not necessarily zero. The symbols in the row above the cluster numbers indicate the majority class in each cluster. Features 1 to 10 (as described in table 3.1) for both problem cases are shown in figures 3.2 and 3.3.

In figure 3.2, H indicates clusters where the majority class is **bright** gray-scale images and _ indicates **dark** grey-scale. In figure 3.3, H indicates majority **red** clusters and _ indicates majority **blue**. The following 3 features in the suite of 55 (See table 3.1) relate to pixel intensity:

- F01 - Distance from unsaturated gray, ie distance from (r,g,b) = (128,128,128).
- F04 - Average brightness over the entire image.
- F07 - Average brightness in the centre of the image.

For images where every pixel has the same intensity, the average intensity over the entire image will be identical to the intensity in any region of the image. Unsurprisingly, in both figures 3.2 and 3.3, the row for feature F02 is identical to the row for feature F05 and similarly for F03 and F06 and F04 and F07. We will ignore the last three. All clusters in both figures have a majority class proportion of 100%. Also, the method has juxtaposed all clusters having the same class labels. We therefore infer the existence of a discriminating manifold for this set of images. For the purpose of illustration, we limit our focus to F04 - Average brightness over the entire image. Consider row 04 in both figures (high-lighted). In figure 3.2, there is contrast between the lighter gradient across clusters (columns) 00 to 06 and the darker gradient across clusters (columns) 07 to 13. This contrast does not exist in figure 3.3. The entire spectrum of the feature values is represented across both classes in figure 3.3. This is indicated by the fact that row 04 in figure 3.3 has the same shades of gray in clusters of both classes. Therefore as expected F04 is shown to be discriminatory for the gray-scale images but not for the red-blue images. Furthermore it can be seen how the values of F04 relate to the class of the image. In figure 3.2, F04 has higher values (lighter shades) for the brighter gray-scale image clusters labelled H than for the darker gray-scale image clusters labelled _.

In summary, the PASOM method has produced a visualization of the semantic map which in this example is strongly grouped and polarised allowing us to infer the existence of a discriminatory manifold in the high dimensional space of pixel intensities. Also produced is a visualization of how feature values vary across the topology of this manifold so that features which permit class discrimination can be identified by visual inspection. In addition, the method has shown qualitatively how feature values correlate to the class of images by contrasting shades across the cluster topology. Features that measure pixel intensity were expected to be shown to be discriminatory between classes in the first problem case but not in the second problem case. The method found that F04 (Average brightness over the entire image) showed a strong contrast between classes in the first case but not the second, as expected.

Chapter 4

The PASOM Method: Moderate Complexity Images

The purpose of this chapter is to show the application of the PASOM method to a problem case of moderate difficulty and to use this example to elaborate on considerations that were omitted in the illustrating example.

Images were generated in four pattern categories (solid color, gingham, squiggle, rainbow) without the explicit a-priori bias designed to high-light particular features as was done previously. Example images are shown in Figure 4.1. The class labels now identify the type of image pattern (SC=solid color, GH=gingham, SQ=squiggle, RB=rainbow). We are interested in discriminating between these pattern types and we wish to know which features, if any, in the suite of 55 shown in table 3.1 are discriminatory and for those that are, how their values correlate with the pattern type.

100 images of 8 x 8 pixel resolution for each pattern type were generated having random colors and 6 balanced problem cases comprising 200 training images in each case were formulated

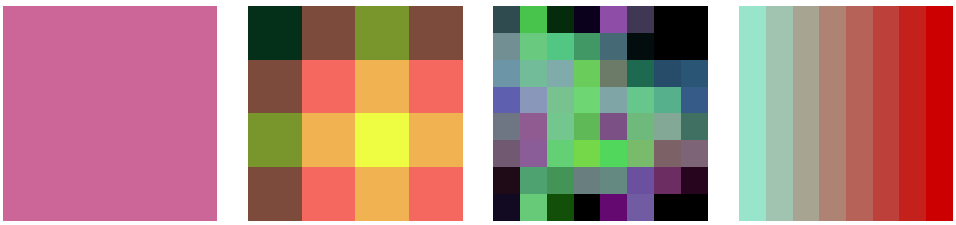


Figure 4.1: Example Moderate Complexity Images as Input

SC-SQ	Solid Color - Squiggle
SC-RB	Solid Color - Rainbow
SC-GH	Solid Color - Gingham
SQ-RB	Squiggle - Rainbow
SQ-GH	Squiggle - Gingham
RB-GH	Rainbow - Gingham

Table 4.1: Moderate Complexity Image Problem Cases

for each binary combination of classes shown in table 4.1

It was noted in the introduction of the PASOM method on page 4 that if a significant majority of the images in each of the clusters determined by the SOM have the same class label, then we can draw further inferences. We think of the majority class proportion as clustering accuracy, determined in the following manner:

The training images are input to the SOM algorithm. The weight vectors for neurons in the Kohonen layer come to represent cluster centroids in the high dimensional space of image pixel intensities. The same training images are then used to determine a class label for each neuron in the Kohonen layer. Each training image is presented to the network and is assigned to its nearest neuron. When all training images have been presented, each neuron is labelled with the majority class of all instances assigned to it. Clustering accuracy is determined over the entire network by the total proportion of training images that have the same class label as the neuron to which it is assigned.

Note that this procedure does not adhere to the standard classification procedure in which the model is tested against a separate validation dataset during training. The purpose of a validation set is to prevent over fitting the model to the training data by the determination of an early stopping point, where the error rate against the validation dataset begins to increase while the error rate against the training data continues to show improvement. The purpose of the clustering step in the PASOM method is not to compute a classification model that generalizes well to future unseen data but rather to obtain a projection of the high dimensional topology of the training images represented as points in a high dimensional space of pixel intensities. To this end, validation is unnecessary.

The PASOM method was applied using a 1 dimensional SOM in 3 experimental configurations: For each of the 6 problem cases described above, a SOM model was configured with either 16, 36 or 64 neurons in the Kohonen layer. The number of neurons in each configuration was

decided on the basis of judgement. The clustering accuracy (%) resulting for each experiment is shown in table 4.2.

Neurons	**SC-SQ**	**SC-RB**	**SC-GH**	**SQ-RB**	**SQ-GH**	**RB-GH**
16	97	69	81	93	97	81
36	98	75	90	95	97	87
64	99.5	81	97	96	97	97

Table 4.2: Moderate Complexity Image Clustering Accuracy (%)

Three observations are noted:

1. Clustering accuracy tends to improve with increasing feature map size.
2. SQ-GH (Squiggle versus Gingham) shows no improvement for increasing feature map size.
3. SC-RB (Solid color versus Rainbow) shows notably poorer clustering accuracy than the other problem cases for each given cluster size.

It is expected that cluster accuracy should improve with increasing feature map size since increasing the number of cluster centroids allows increasing compartmentalization of the training data, the limit being a number of cluster centroids equal to the number of training examples and each training example assigned to its own centroid. This is a mechanism by which we can remedy a lower than desired clustering accuracy. We simply add neurons to the Kohonen layer. The trade-off will be that aggregation within each cluster becomes less meaningful since there are fewer instances assigned to clusters.

Since increasing feature map size should increase clustering accuracy, observation 2 is counter to our expectations. A possible explanation for this is the curse of dimensionality. One aspect to the curse of dimensionality is the loss of significance of variance between Euclidian distance measures on vectors with increasing dimensionality of those vectors. This can be stated for two vectors **Lmax** and **Lmin** of differing length and having dimension {d} as [3]:

$$limd->inf(||\mathbf{Lmax}|| - ||\mathbf{Lmin}||)/||\mathbf{Lmin}||->0 \tag{4.1}$$

The SOM makes its decisions by comparing Euclidian distances and if the variance in difference vectors between a coordinate point and various centroids becomes insignificant ie beyond

Image Pattern Combination	Accuracy (%)	Semantic Map
Solid-Color - Squiggle (SC-SQ)	97	_ _ H _ _ _ _ _ H H H H H _ _ _
Solid-Color - Rainbow (SC-RB)	69	_ H _ H _ H _ H _ H _ H _ H _ _
Solid-Color - Gingham (SC-GH)	81	H H _ _ _ _ _ _ H H H _ _ H H _
Squiggle - Rainbow (SQ-RB)	93	H _ _ H _ H H H H _ _ _ _ H H H
Squiggle - Gingham (SQ-GH)	97	_ _ _ _ H H H H H H H H H H _ _
Rainbow - Gingham (RB-GH)	81	_ H _ H H _ _ _ H H _ _ H H H _

Table 4.3: Semantic Maps for 16 neuron experiments

Class Label	SC-SQ	SC-RB	SC-GH	SQ-RB	SQ-GH	RB-GH
H	SQ	RB	GH	RB	GH	GH
_	SC	SC	SC	SQ	SQ	RB

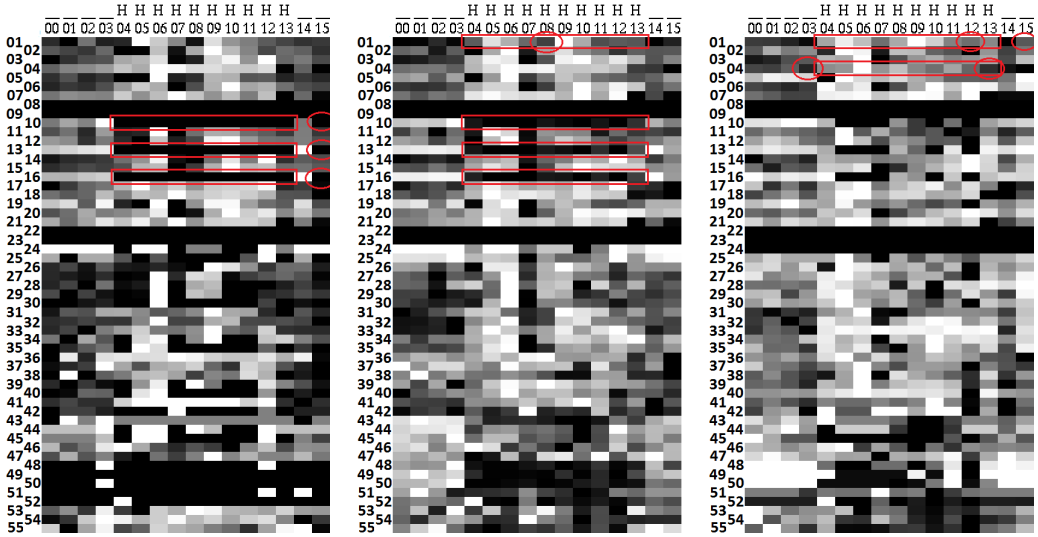
Table 4.4: Class label key for 16 neuron experiments

the numerical resolving power of the number of significant digits of the machine then the co-ordinate may be assigned to the incorrect centroid. Another possible explanation is that the topology of the manifold of class labels across the cloud of coordinate points representing images in the high dimensional space of pixel intensities is such that the trend to increased compartmentalization with increasing feature map size exhibits clustering accuracy plateaus. The conclusion is that we should be mindful of subtleties that may be associated with high dimensional input data.

In both the introduction chapter and the chapter presenting the illustration example it was stated that the **average** values of each feature for images within each cluster are computed and then visualized using grey-scale heat-maps. Other metrics, such as **minimum** and **maximum** feature value can be similarly visualized. It is quite possible that the average value of a discriminatory feature shows no significant variation across the topology, but the minimum or maximum value does. Therefore we should examine all three.

Table 4.3 shows the semantic maps that result for each of the 16 neuron experiments described above. The relevant clustering accuracies in table 4.2 are repeated in column 2 of table 4.3 for ease of reference. Symbols H and _ are used as generic class labels to distinguish whichever two image patterns are being analysed. Table 4.4 shows the image pattern corresponding to these symbols for each experiment.

Notice that for the SC-RB case in table 4.3, the semantic map shows alternating class



a. Minimum Values b. Mean Values c. Maximum Values

Figure 4.2: Output Heat-map for Moderate Complexity Images

labels. There is no grouping of like with like. Therefore we cannot infer the projection of a discriminating manifold for these two image patterns. This conclusion is supported by the noticeably lesser clustering accuracy (69%) in comparison to the other cases (81% - 97%). The PASOM method under the given experimental configuration performs poorly at distinguishing the solid color from the rainbow image pattern types. This precludes further analysis of features. Increasing the number of neurons in the Kohonen layer may increase clustering accuracy and yield an improved semantic map. Instead, consider the comparison of the SC-RB case to the SQ-GH case. The SQ-GH case shows a strong grouping of H class labels in the centre of the map and the clustering accuracy is encouraging (97%). We can infer the projection of a discriminating manifold and proceed to the next step of visualizing how features of interest map to this topology. The heat maps for minimum, mean and maximum feature values for the SQ-GH case are shown in figure 4.2, with highlights on patterns of interest.[1]

In figure 4.2 we refer to the central region as clusters 04 to 13 and the outer regions as clusters 00 to 03 and 14 to 15. Minimum feature values (Figure 4.2a) might suggest that features 10, 13 and 16 are relevant since in these rows the central region is uniformly dark in contrast to the outer regions, with the exception of cluster 15 (circled). We could not claim these features to be discriminatory solely on the evidence of minimum feature values due to this

[1]The heat maps for all 6 problem cases are shown in appendix figures 1 to 6 for reference.

	Image Type H (GH)	**Image Type _ (SQ)**
Average minimum value	0.0	15.0
Average mean value	1.7	27.4
Average maximum value	12.4	42.9

Table 4.5: Averaged measures of feature 10 for Squiggle - Gingham image patterns

exception. Mean feature values (Figure 4.2b) show the same contrasting regions for these 3 features, but without the exception. In figure 4.2b, features 10, 13 and 16 show starkly contrasting mean values between the two pattern types (squiggle and gingham). Feature 01 in figure 4.2b and features 01 and 04 in figure 4.2c show patterns that suggest these features have relevance, but again with exceptions (circled) that preclude stronger conclusions.

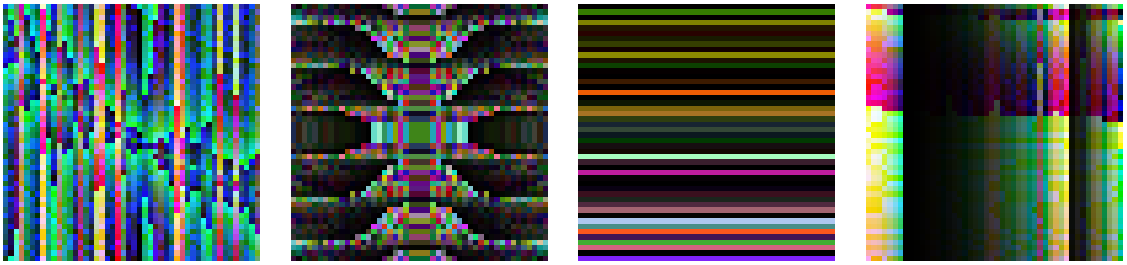
Heat maps provide a qualitative indication of feature value variation across the cluster topology. The data underlying the heat maps can be used to quantify this variation and identify the range of feature values that correlate with either class. As an example, table 4.5 shows the averages of minimum, mean and maximum values of feature 10 contrasted between the central and outer regions visualized in figure 4.2. Our conclusion is that for this set of images, Mean value of feature 10 is discriminatory between squiggle and gingham pattern types: values (0.0 to 12.4) correlate with the gingham pattern type while values (15.0 to 42.9) correlate with the squiggle pattern type.

In summary the PASOM method has been shown to produce a visualization, which allows us to infer a discriminatory manifold in the high dimensional space of pixel intensities for some combinations of image pattern types of interest. Just as importantly, problem cases have been identified for which a discriminatory manifold is not evident and further analysis of features is precluded. Discriminatory features have been identified by visual inspection and an example of a quantitative correlation of feature values with image pattern type has been shown.

Chapter 5

The PASOM Method: IMAGENE Images

To apply the PASOM method to the abstract art images of primary interest at full resolution (320x320 pixels) proved to be impractical given available computing resources. Instead a new set of image miniatures was generated at 48x48 pixel resolution using the same genetic programming application (IMAGENE) [37] used to generate the original 320x320 images. These image miniatures were subjected to the aesthetic judgement of the author and a balanced dataset comprising 100 images deemed to have high aesthetic value and 100 images deemed to have low aesthetic value was formulated. Figure 5.1 shows example images.[1]



High Aesthetic High Aesthetic Low Aesthetic Low Aesthetic

Figure 5.1: Example IMAGENE Miniature Images as Input

The PASOM method was applied to this dataset using a SOM having 16 neurons in the Kohonen layer. Clustering accuracy is 73% which is weak, representing a kappa value of only

[1]The entire set of images in each class is shown in Appendix figures 7 and 8 for reference.

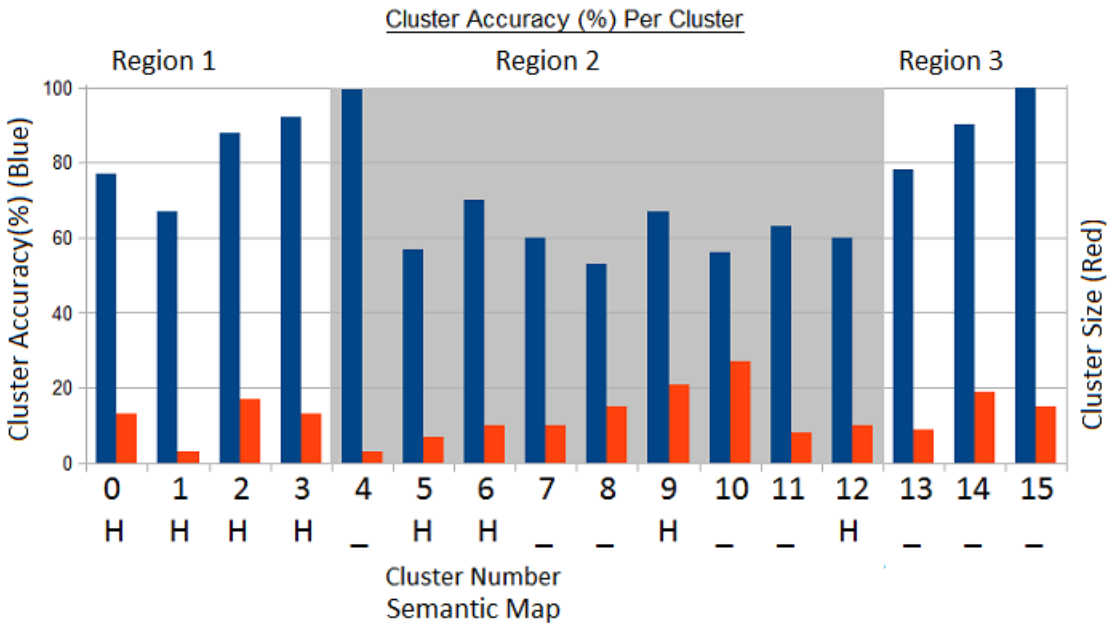


Figure 5.2: IMAGENE Image Cluster Accuracy Per Cluster (%)

0.46, less than half the available margin for improvement over 50% expected from the random model. [2] Clustering accuracy could be increased by the addition of some number of neurons in the Kohonen layer. However, uninterrupted run time to build the 16 neuron SOM model for this dataset extended over 2 weeks on a single node of a high performance computing platform so the decision was made to not pursue that avenue.

Due to the weak aggregated clustering accuracy we examine the clustering accuracy per cluster. Figure 5.2 charts the accuracy per cluster (Blue) and also the number of images in each cluster (Red). See also table 5.1 for the number of images in each cluster. Columns labelled H in Figure 5.2 have a high aesthetic majority class and columns labelled _ have a low aesthetic majority class. The horizontal axis of this chart corresponds to the same topological arrangement of clusters produced by the SOM that is presented across the rows in the heat map visualizations.

[2]For a balanced two class problem as we have here, 50% is the majority class proportion and hence the expected accuracy of the random model.

Cluster	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Size	13	3	17	13	3	7	10	10	15	21	27	8	10	9	19	15

Table 5.1: Number of images per cluster

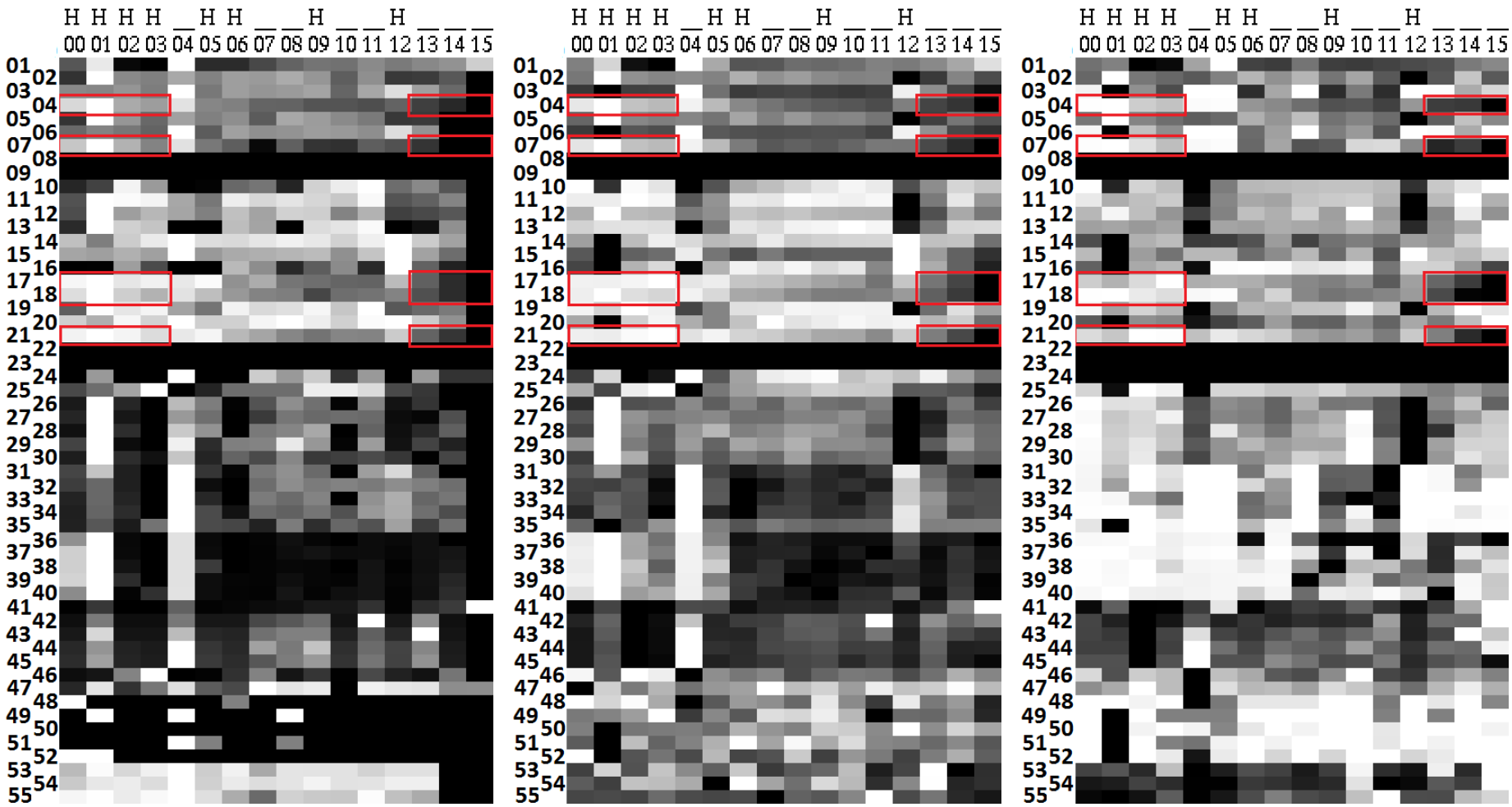
We identify 3 regions in this topology: region 1 - clusters 0 to 3, region 2 - clusters 4 to 12 (shaded) and region 3 - clusters 13, 14 and 15. The outer regions (1 and 3) are strongly representative of their respective aesthetic labels both in terms of juxtaposition of similarly labelled clusters and in terms of majority class dominance within these clusters, whereas the central region is less so. Region 2 of the topology tends to show juxtaposition of clusters with different class labels and most clusters in this region show markedly lower (approx 50% to 70%) majority class dominance than in the outer regions (approx 80% to 100%).

Average cluster accuracy in the 3 regions is shown in table 5.2. For the purpose of identifying features with high discriminatory power we will ignore the central region and focus only on the strongly representative regions 1 and 3.

Region 1	**Region 2**	**Region 3**
81	65	89

Table 5.2: Average Clustering Accuracy Per Region (%)

The heat maps for minimum, mean and maximum feature values across the 16 cluster topology are shown in figure 5.3, with highlights on patterns of interest in regions 1 and 3. The suite of 55 features being analysed in figure 5.3 is listed in table 3.1. The heat maps offer a rich source of patterns that together with the semantic map has a wide scope for analysis. We identify a selection of features that differ starkly between regions 1 and 3 as highlighted in figure 5.3. We can quantify the range of feature values in each region and show the correlation to aesthetic value. Table 5.3 shows a comparison of minimum, mean and maximum values between high and low aesthetic regions.



Minimum Values | Mean Values | Maximum Values

Figure 5.3: Output Heat-map for IMAGENE Images

	High Aesthetic			Low Aesthetic		
	Min	Mean	Max	Min	Mean	Max
F04	0.77	0.84	0.89	0.13	0.20	0.27
F07	0.73	0.84	0.92	0.07	0.20	0.37
F17	39.8	42.7	47.2	7.6	14.0	23.9
F18	58.6	63.9	67.4	11.3	18.9	30.3
F21	104.8	115.3	122.2	22.1	40.5	67.7

Table 5.3: Quantified feature values

Higher brightness over the entire image (Feature 04) corresponds to higher aesthetic value. Feature 04 has an average value that ranges between 0.77 and 0.89 for high aesthetic images and between 0.13 and 0.27 for low aesthetic images. Higher brightness in the centre of the image (Feature 07) corresponds to higher aesthetic value. Feature 07 has an average value that ranges between 0.73 and 0.92 for high aesthetic images and between 0.07 and 0.37 for low aesthetic images. Higher smoothness at certain scales corresponds to higher aesthetic value. Feature 17 as an example has an average value that ranges between 39.8 and 47.2 for high aesthetic images and between 7.6 and 23.9 for low aesthetic images. In each of these cases the average maximum feature value for low aesthetic images is less than the average minimum value for high aesthetic images. This suggests that it may be possible in future work to determine for certain features, threshold values at which high aesthetic value is lost.

We can verify the sensibility of the results above by visual inspection of sample images from the clusters computed by the SOM. Figure 5.4 shows example images from clusters 00 to 03 (High aesthetic) and figure 5.5 shows example images from clusters 13 to 15 (Low aesthetic). All images (except the first image shown for cluster 01) have an aesthetic rating consistent with the majority class of their resident cluster. By visual inspection it should be reasonably obvious that the higher aesthetic images do indeed have higher brightness in comparison to the lower aesthetic images. Smoothness is somewhat a misnomer. Smoothness features are often associated with the character of texture within an image. The low aesthetic images in figure 5.5 tend to show uniform (dark) regions lacking in texture and hence have low smoothness feature values, whereas the high aesthetic images in figure 5.4 have rich textures and hence high smoothness feature values. Feedback from a professional artist/photographer[3] corroborates these conclusions as being consistent with principles for attaining aesthetic value in visual media.

[3]Karen Trist, School of Media and Communication, RMIT University, Design & Social Context.

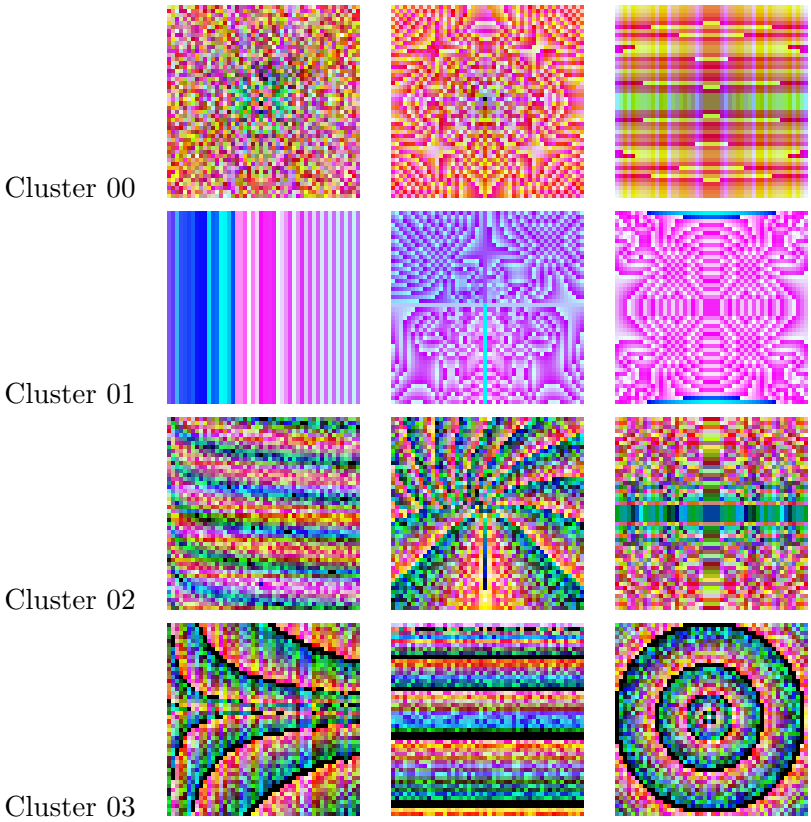


Figure 5.4: Sample IMAGENE images in high aesthetic clusters.

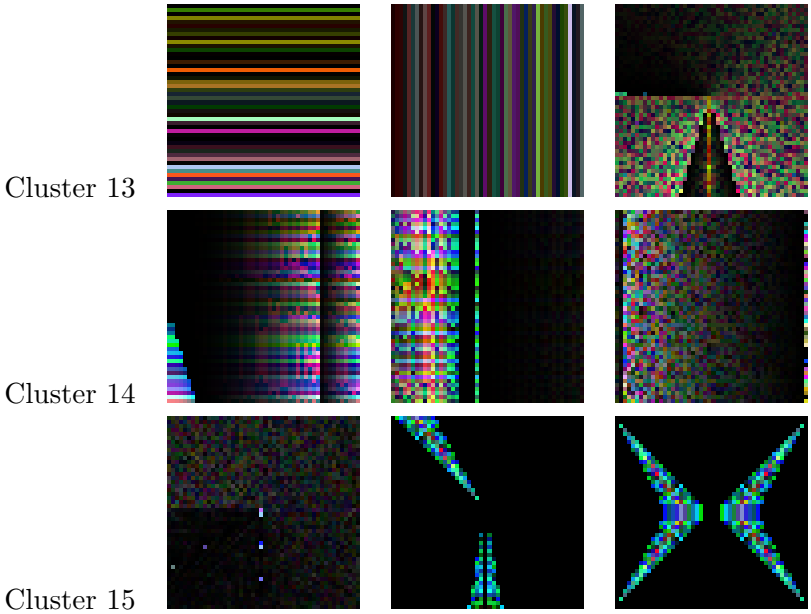


Figure 5.5: Sample IMAGENE images in low aesthetic clusters.

In summary the PASOM method has allowed an analysis of a set of abstract images in a manner independent of image features. The method allows us first to gauge aesthetic class separability of the images by visual inspection of the semantic map. For a semantic map that can be interpreted as the projection of a discriminating manifold a heat map visualization technique then allows a large group of features to be analysed by visual inspection, a task that is difficult to visualize using charts. The heat map visualization allows complex patterns of feature variation across the cluster topology and how these relate to the semantic map to be analysed. Aesthetically discriminating features can be identified and quantified.

Chapter 6

Conclusions

This work was concerned with the development of a method to analyse and understand features that distinguish between abstract images of high and low aesthetic value. We have proposed a method based on clustering with the Self Organising Map where the raw pixel data of images is input to the algorithm as high dimensional pixel arrays. We found that the method was successful on a difficult problem involving images of limited size.

6.1 Research Questions

1. How can we identify features that have high relevance to abstract image aesthetics?

The PASOM method produces a one dimensional semantic map from the raw pixel data of images, independently of any features. We have argued that if there is a juxtaposition of similarly labelled clusters and the clustering accuracy within clusters is judged to be significant then the pattern of labelled clusters across the semantic map can be interpreted as the projection of an aesthetically discriminating manifold and features that have high relevance to image aesthetics can be identified by a contrast of values between clusters of high aesthetic majority class and clusters of low aesthetic majority class. For the set of IMAGENE images analysed using a suite of 55 features originated in prior work and in the context of the aesthetic judgement of the author, we showed the following features to exhibit this contrast:

- Feature 04 - Average brightness over the entire image
- Feature 07 - Average brightness in the centre of the image

Rank	1	2	3	4	5	6	7	8	9	10
Feature	F04	F07	F01	F39	F43	F03	F06	F34	F40	F45

Table 6.1: Feature ranking shown by previous work

- Features 17, 18, 21 - Wavelet functions computing image smoothness at various scales

Prior research with the same suite of features applied to similar abstract art images generated by the IMAGENE program, established a ranking of those features by their contribution to the predictive power of an accurate (91%) classifier ensemble [5]. That ranking is shown in table 6.1.

There is strong agreement between our findings and the findings of prior work in regard to features 04 and 07. Prior work found features 04 and 07 to be the top 2 ranking features that predict aesthetic rating for IMAGENE images. The PASOM method identified strongly contrasting values of these features between high and low aesthetic IMAGENE images. This agreement corroborates the ability of the PASOM method to identify features of high relevance to aesthetic discrimination.

Features 17, 18 and 21 were identified as relevant by the PASOM method, but do not appear in the ranking established by prior work. A possible explanation for this is that different aesthetic judgements were applied. Perhaps features 17,18 and 21 are measures of visual properties favoured by the author, but not by the critics consulted in the prior work.

Some features appearing in the ranking of prior work are not identified as clearly discriminatory by the PASOM method. A limitation of the traditional method applied in prior work is that a ranking of features gives no indication of the relevance of those features. Even features that are weakly associated with aesthetic value will be ranked highly in the absence of more relevant features. The PASOM method makes a strong contribution to the field of computational aesthetics because features are able to be shown to be independently relevant to aesthetic value by contrasting values of those features between different aesthetic regions in the semantic map.

2. How do the values of high relevance features correlate with aesthetic appeal?

For the images under study, in the context of the aesthetic judgement of the author, we have shown for each of the features identified as relevant, the range of values that correlate with high aesthetic value and the range of values that correlate with low aesthetic value, see table

5.3. In each case the average maximum feature value for low aesthetic images is less than the average minimum value for high aesthetic images, suggesting the existence of threshold values at which high aesthetic value is lost.

6.2 Other Findings

- It was necessary to consider metrics of feature values for both centre and spread to avoid erroneously discounting features on the basis of insignificant variation in a single metric.
- Run time to build the 16 neuron SOM model for the dataset of images shown in figure 7 and figure 8 extended over 2 weeks on a single node of a high performance computing platform.

6.3 Further Work

Relating back to figure 5.2, we make the following observation: The central region (region 2) of the topology shows interleaved clusters of high and low aesthetic majority class with a variation from left to right that is suggestive to us of a concentration gradient. We might interpret this concentration gradient as an aesthetic gradient from high aesthetic to low aesthetic, left to right across the topology for this set of images. Further work with higher resolution feature maps may contribute to the idea that not only can aesthetically discriminating manifolds be projected, but also aesthetic gradients in the high dimensional space.

Features 17, 18 and 21 are only 3 of 11 features analysed that quantify image smoothness at various scales. There is scope for further research to enhance understanding of how the scale of smoothness features and texture relates to image aesthetics.

Also worthy of further investigation is the potential for using the PASOM method to identify feature value thresholds at which high aesthetic appeal is lost.

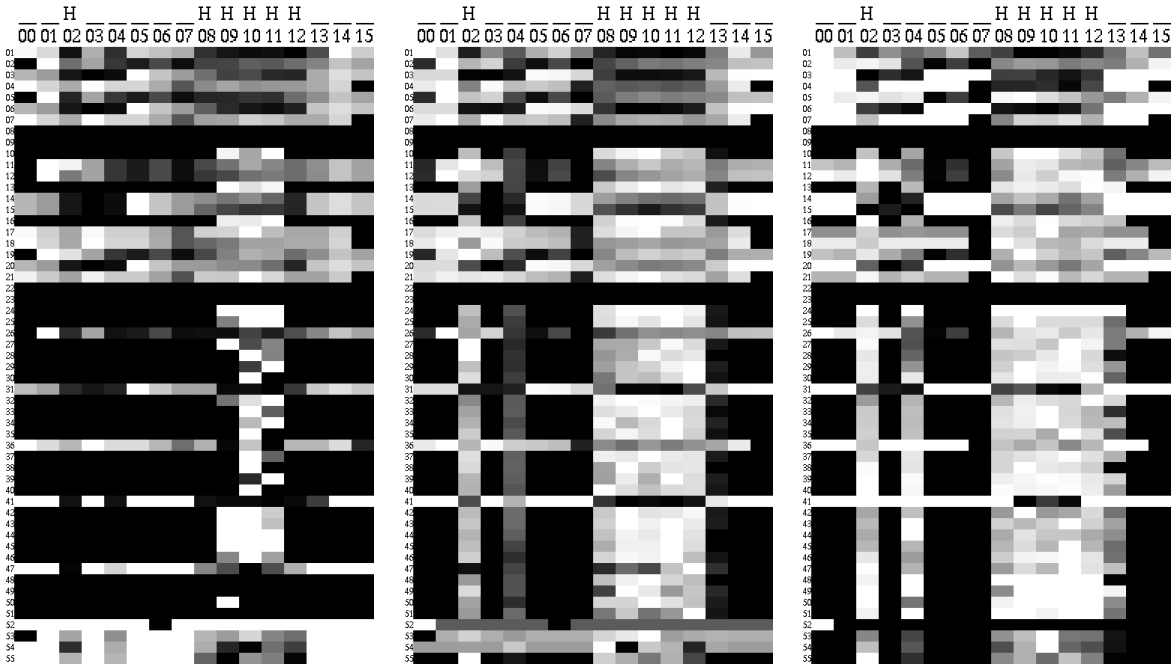
The Kohonen layer feature maps employed in this work are all one dimensional. Many applications of the Kohonen self organising map employ two dimensional feature maps and the idea has occurred to us that the PASOM method could be extended to feature maps of two and possibly higher dimensions. We speculate on the possibly of second order correlations between feature value variation and the landscape of a two dimensional cluster topology. Some features may be found to be dominated by others and in a one dimensional feature map the

dominated feature would appear to map less meaningfully or not at all to the topology and be dismissed as irrelevant whereas given additional degrees of freedom in higher dimension feature maps, a less dominant feature might show meaningful variation orthogonal to a feature that dominates it.

To apply the PASOM method to the abstract art images of primary interest at high resolution proved to be impractical. Instead a set of image miniatures was analysed. Further work is necessary to investigate scaling the method to higher resolution images.

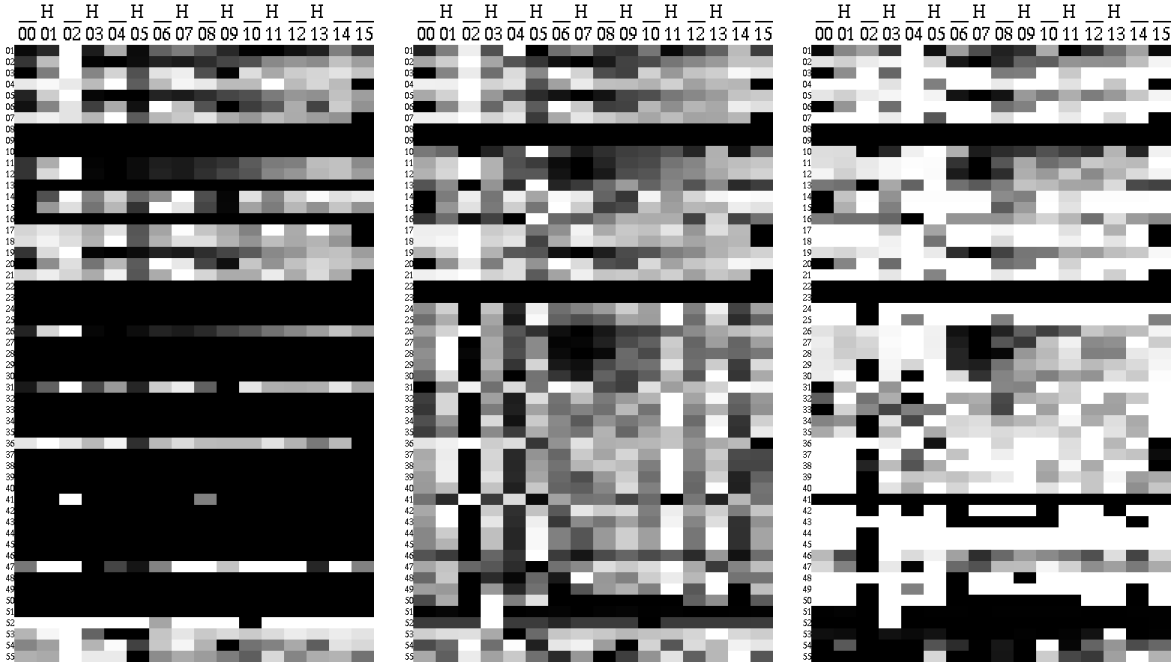
Finally, recent developments in deep architectures suggest they have some potential for contributing to the field of computational aesthetics. Deep learning has shown the capability to discover a set of features that encode an abstraction of a set of images in terms of learned rather than engineered features. Furthermore this discovery is achieved in an unsupervised machine learning mode and is hierarchical, ie higher layers of abstraction can be learned from lower layers of features. We suggest that the application of the deep learning approach to the discovery of a hierarchy of features that underpin aesthetic value in images has some potential for advancing the field of computational aesthetics.

Appendices



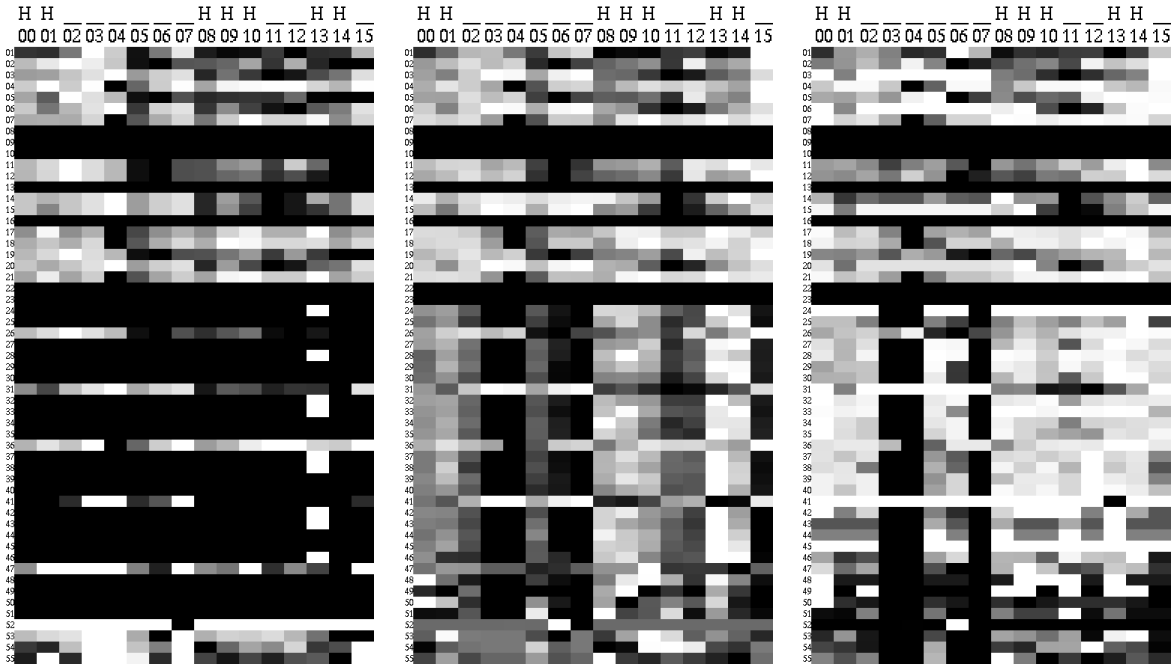
a. Minimum Values b. Mean Values c. Maximum Values

Figure 1: Heat Maps: Solid Color - Squiggle



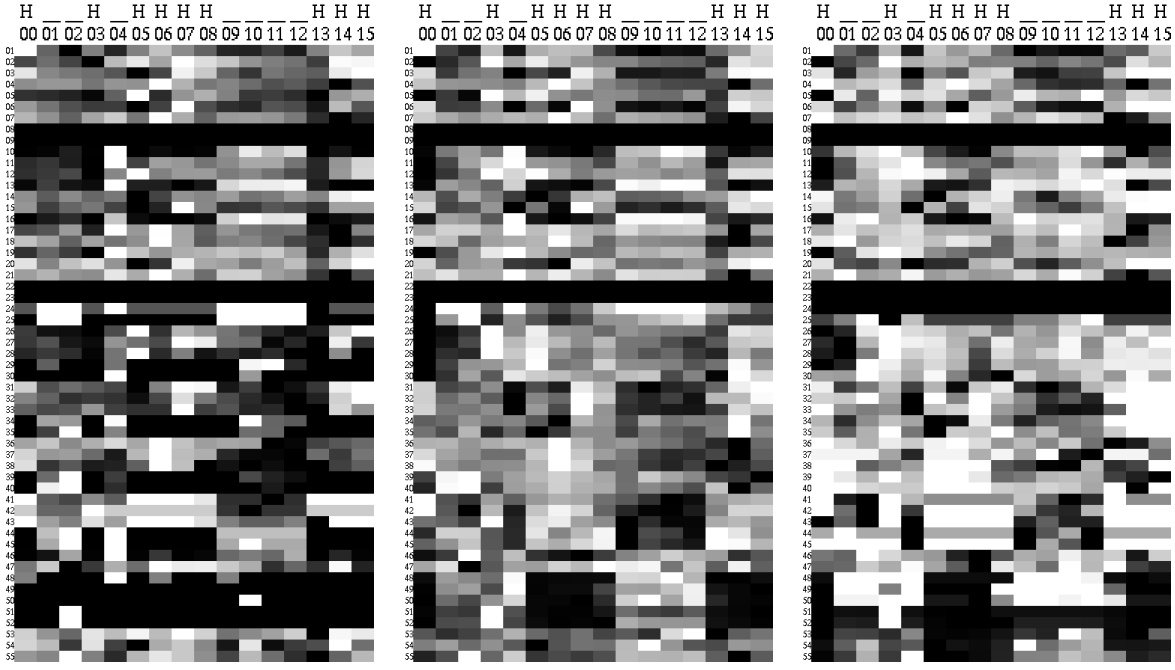
a. Minimum Values b. Mean Values c. Maximum Values

Figure 2: Heat Maps: Solid Color - Rainbow



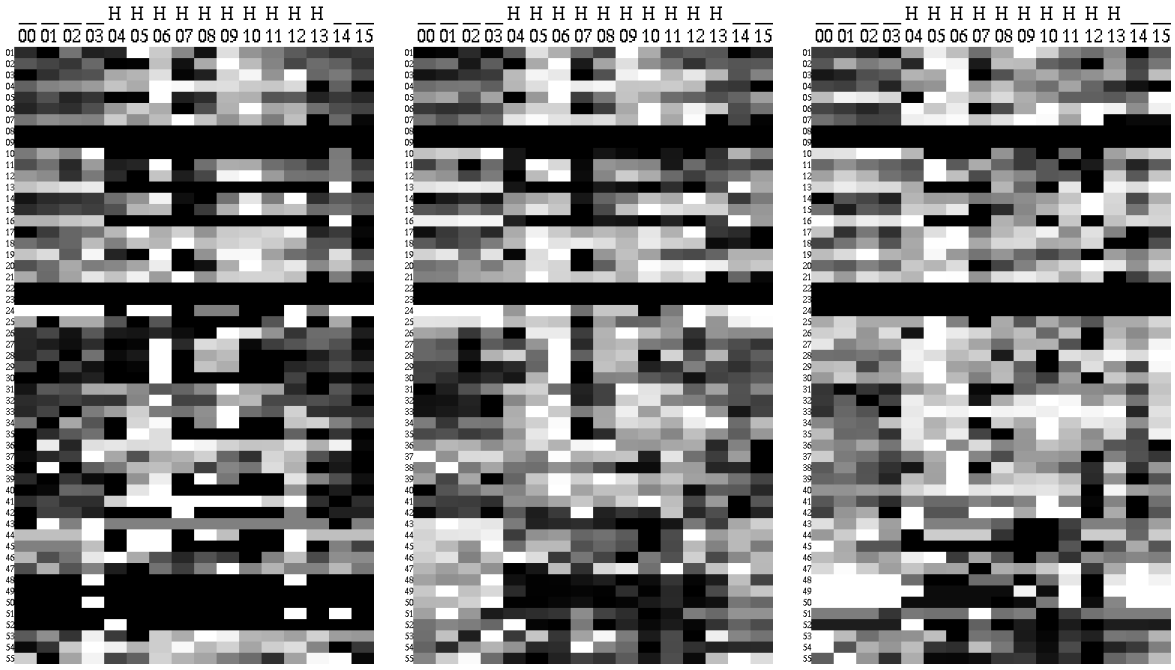
a. Minimum Values b. Mean Values c. Maximum Values

Figure 3: Heat Maps: Solid Color - Gingham



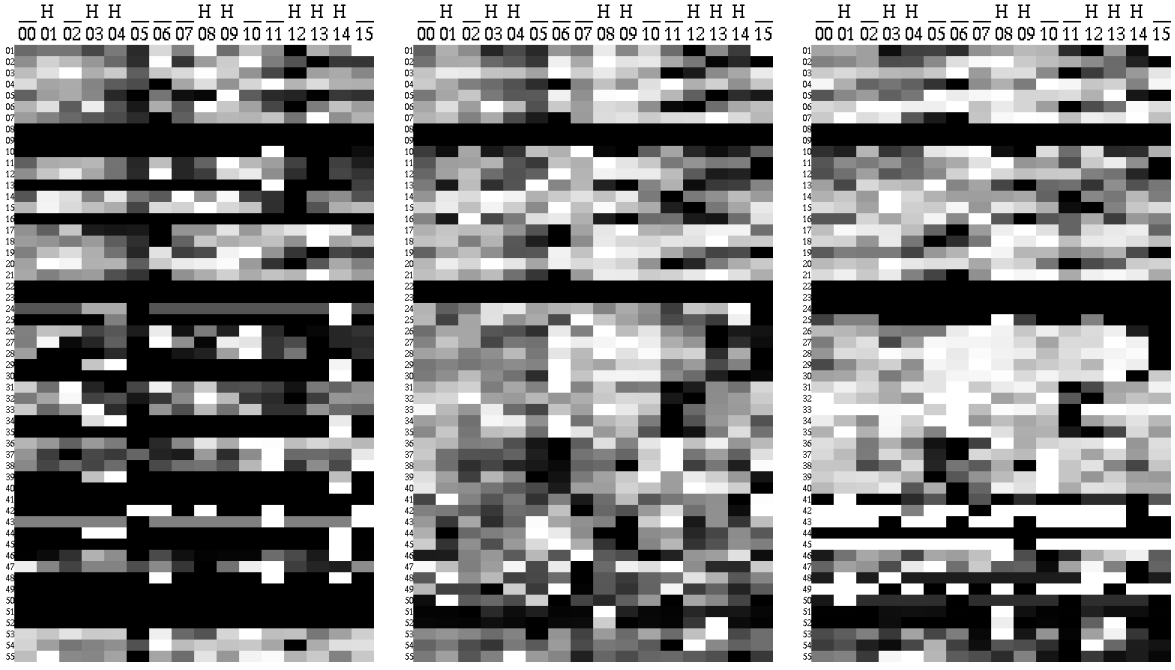
a. Minimum Values b. Mean Values c. Maximum Values

Figure 4: Heat Maps: Squiggle - Rainbow



a. Minimum Values b. Mean Values c. Maximum Values

Figure 5: Heat Maps: Squiggle - Gingham



a. Minimum Values b. Mean Values c. Maximum Values

Figure 6: Heat Maps: Rainbow - Gingham

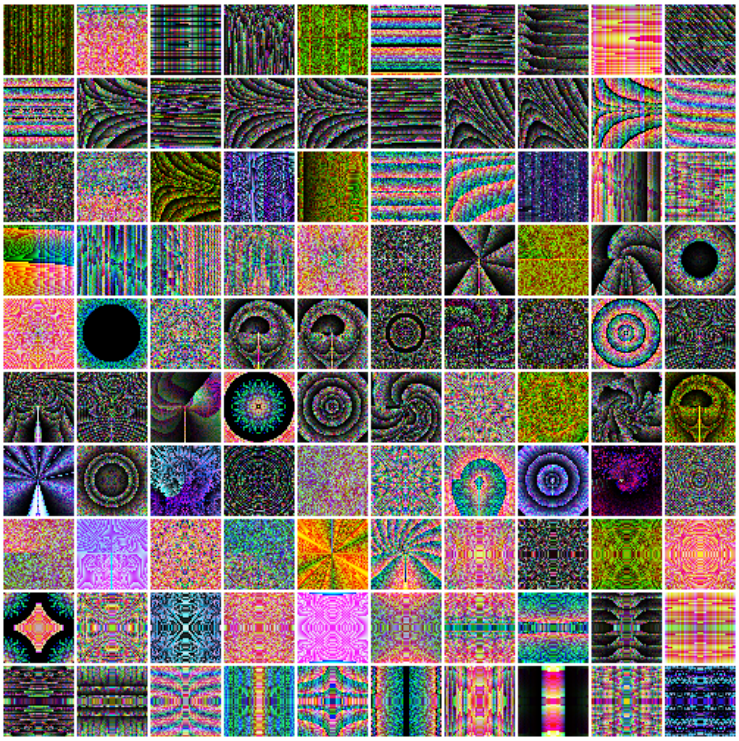


Figure 7: All High aesthetic IMAGENE miniatures

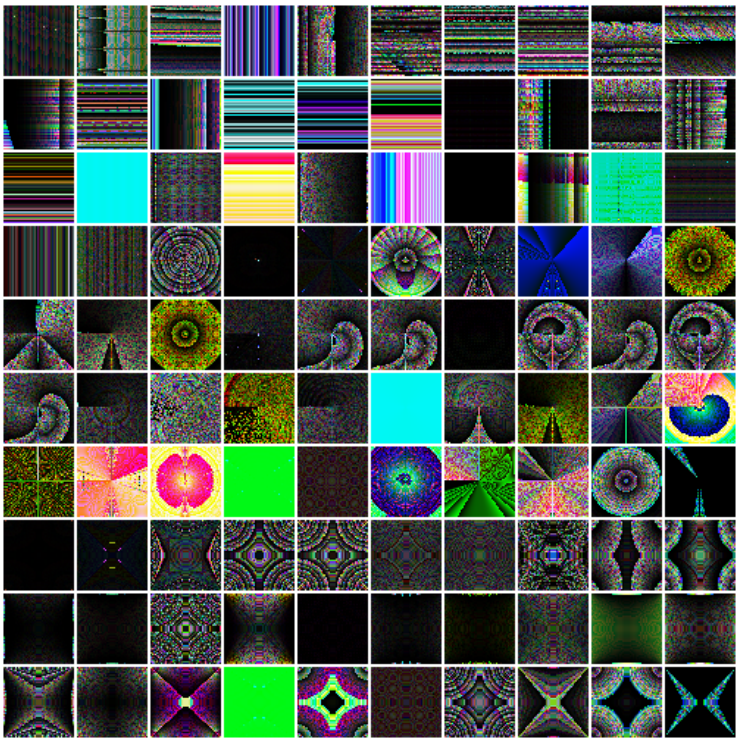


Figure 8: All Low aesthetic IMAGENE miniatures

Bibliography

[1] Rudolf Arnheim. *Art and visual perception: A psychology of the creative eye*. Univ of California Press, 1971.

[2] Sven Behnke. *Hierarchical neural networks for image interpretation*, volume 2766. Springer-Verlag New York Incorporated, 2003.

[3] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is nearest neighbor meaningful? In *Database TheoryICDT99*, pages 217–235. Springer, 1999.

[4] George David Birkhoff. *Aesthetic measure*. Cambridge, Mass., 1933.

[5] Vic Ciesielski, Perry Barile, and Karen Trist. Finding image features associated with high aesthetic value by machine learning. In *Evolutionary and Biologically Inspired Music, Sound, Art and Design*, pages 47–58. Springer, 2013.

[6] Harold Cohen. The further exploits of aaron, painter. 1995.

[7] Harold Cohen, Frieder Nake, David C Brown, Paul Brown, Philip Galanter, Jon McCormack, and Mark dInverno. Evaluation of creative aesthetics. In *Computers and Creativity*, pages 95–111. Springer, 2012.

[8] Ritendra Datta. *Semantics and aesthetics inference for image search: statistical learning approaches*. Pennsylvania State University, 2009.

[9] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006*, pages 288–301. Springer, 2006.

[10] Richard Dawkins. The blind watchmaker, harlow, 1986.

[11] E Den Heijer and AE Eiben. Comparing aesthetic measures for evolutionary art. In *Applications of Evolutionary Computation*, pages 311–320. Springer, 2010.

[12] Michael Egmont-Petersen, Dick de Ridder, and Heinz Handels. Image processing with neural networksa review. *Pattern recognition*, 35(10):2279–2301, 2002.

[13] Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological cybernetics*, 20(3-4):121–136, 1975.

[14] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[15] TY Gajjar and NC Chauhan. A review on image mining frameworks and techniques. *International journal of computer science and information technologies, vol3 (3)*, 2012.

[16] Philip Galanter. Computational aesthetic evaluation: past and future. In *Computers and Creativity*, pages 255–293. Springer, 2012.

[17] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[18] Florian Hoenig. Defining computational aesthetics. In László Neumann, Mateu Sbert, Bruce Gooch, and Werner Purgathofer, editors, *Computational Aesthetics*, pages 13–18. Eurographics Association, 2005.

[19] Toshikazu Kato. Database architecture for content-based image retrieval. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, pages 112–123. International Society for Optics and Photonics, 1992.

[20] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 419–426. IEEE, 2006.

[21] Anthony Knittel and Alan D Blair. An abstract deep network for image classification. In *AI 2012: Advances in Artificial Intelligence*, pages 156–169. Springer, 2012.

[22] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

[23] Teuvo Kohonen. Essentials of the self-organizing map. *Neural Networks*, 2012.

[24] Matthew Kyan, J Jarrah, Paisarn Muneesawang, and Ling Guan. Strategies for unsupervised multimedia processing: self-organizing trees and forests. *Computational Intelligence Magazine, IEEE*, 1(2):27–40, 2006.

[25] Zbigniew Les. *Image Understanding and Aesthetic Evaluation.* University of Melbourne, 1996.

[26] Congcong Li and Tsuhan Chen. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):236–252, 2009.

[27] Qingyong Li, Siwei Luo, and Zhongzhi Shi. Semantics-based art image retrieval using linguistic variable. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, volume 2, pages 406–410. IEEE, 2007.

[28] Penousal Machado and Amílcar Cardoso. Computing aesthetics. *Advances in Artificial Intelligence*, pages 105–119, 1998.

[29] Penousal Machado and Amílcar Cardoso. Generation and evaluation of artworks. In *Proc. of the 1st European Workshop on Cognitive Modeling, CM'96*, pages 96–39, 2010.

[30] John Pearson. The computer: Liberator or jailer of the creative spirit. *Leonardo. Supplemental Issue*, pages 73–80, 1988.

[31] Juan A Ramirez-Quintana, Mario I Chacon-Murguia, and Jose F Chacon-Hinojos. Artificial neural image processing applications: A survey. *Engineering Letters*, 20(1):68, 2012.

[32] David Reaves. Aesthetic image rating (air) algorithm. 2008.

[33] Katharina Reinecke and Abraham Bernstein. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2):8, 2011.

[34] Jaume Rigau, Miguel Feixas, and Mateu Sbert. Informational aesthetics measures. *Computer Graphics and Applications, IEEE*, 28(2):24–34, 2008.

[35] Jaume Rigau, Miquel Feixas, and Mateu Sbert. Conceptualizing birkhoff's aesthetic measure using shannon entropy and kolmogorov complexity. In *Proceedings of the Third Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, pages 105–112. Eurographics Association, 2007.

[36] Yaowen Wu, Christian Bauckhage, and Christian Thurau. The good, the bad, and the ugly: Predicting aesthetic image labels. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1586–1589. IEEE, 2010.

[37] Qinying Xu, Daryl D'Souza, and Vic Ciesielski. Evolving images for entertainment. In *Proceedings of the 4th Australasian conference on Interactive entertainment*, page 26. RMIT University, 2007.