Exploring dimensionality reduction on semi-structured photos — a closer look at *Exactitudes*

S.D. (Sam) Verkoelen
Graduation Thesis
Media Technology MSc program, Leiden University
August 2015
Thesis advisors: *Maarten H. Lamers & Peter van Putten*

Abstract: Deep learning or dimensionality reduction has shown unprecedented results in contemporary researches. However, these systems are often referred to as black boxes. This exploratory research will look at dimensionality reduction via Restricted Boltzmann Machines in a set of small experiments. These experiments were performed on the semi structured photos of the *exactitudes* project, resulting in novel insights and visualisations on how to extract meaningful information from a deep networks. These insights do not only help to understand what such systems learns but it also provides a glimpse of the underlaying structure of the dataset.

1. Introduction

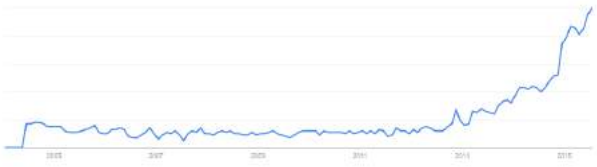


Figure 1.1:Google searches on "deep learning" from Januari 2004 to June 2015

Deep learning is a hot topic, as Figure 1.1 shows with the Google searches on this subject. Not surprising given the media attention and current developments on various problems, such as the classification of news stories[3] or music[1]. This makes it more striking that there is actually known surprisingly little about the learning process of these models[18].

Deep learning generally refers to the stacking of multiple layers in a neural network. By adding a layer, computational power is added and higher level information could be derived from a dataset[3]. To train a deep networks in practice, it requires, amongst other things, fast enough computers and large datasets. Since both requirements seem to have been fulfilled[3], the interest shown in Figure 1.1 could be explained.

These large datasets can be expressed by an arbitrarily number of dimensions, depending on the representation and accuracy. Dimensions can also be called features, this term is commonly used after a reduction of the number of dimensions[3]. The use of deep learning is also known as feature extraction or dimensionality reduction.

Features are useful because they have the potential to show us the building blocks of the data. A typical use for these extracted features is for example the classification of data[1, 2, 9, 10, 12].

Feature extraction in all these examples is unsupervised, meaning that the algorithm tries to find a good abstraction of itself in fewer dimensions based on similarities across the dataset. From this, it possible that features of which is not clear what they actually represent or yet unknown features emerge. In the classification of news articles[3] the researchers did not use the already available labels on types of articles. Yet the algorithm was able to determine eight classes that corresponded with the labels and separate the data in a meaningful way. This means that you cannot only search for a needle in a haystack, but feature extractors will also determine how many needles there are and what a needle even looks like. Meanwhile preserving the possibility for the researcher to bias the algorithm based on prior knowledge. For example by selecting input dimensions based on the presence of words in a carefully chosen list[3].

The need for automated feature extraction comes partially from the increasing amount of data, it simply becomes impractical to manually assign

features to large datasets. Additionally to the amount of data there is also the complexity of features that drive the need for automated feature extraction. There are cases where modern algorithms are able to find complex underlaying features, possibly not apparent to the human eye. Such as in the identification of Van Gogh paintings by extracting features on brush strokes[12].

In this particular case a noteworthy conclusion was that not only the classification was a result of their research, but also the insight in the complex features, or building blocks that are elemental to Van Goghs paintings. These results can help for example art historians to gain a better insight in the technique used by the painter[12] or the decisions made by the painter[10].

These examples show that features, that have been learned unsupervised, can provide new and important information on large datasets. However the neural networks used to get to these insights are often referred to as a 'black box', despite being built on well understood mathematical principles[18]. Regardless of the lack of knowledge about the inner workings, these networks do provide results that one can build upon.

Therefore it is not surprising that these networks are being used purely as a tool. A tool which produces results that are usable to researchers or a tool that can be integrated in commercial products. Only few attempts have been made to get a glimpse inside these 'black boxes', one of which is 'deep dreams' by Google research[18]. This research uses a method, explained in X, that produces visualisations of the learned information within a network. Such visualisations show a benefit to both the interpretation of results and the discovery of possible shortcomings and opportunities in the network or dataset.

This research will be an exploratory attempt to gain insight in what is learned in a dimensionality reducing neural network when applied to semi structured photos. The latter being photos that are individually unique, however show a great equality in visual appearance across them. This visual equality has the advantage that subtile differences across photos have a greater distinctive character and are presumably located in specific areas of the photos across the whole dataset.

The exploratory nature of this research is noticeable in the small experiments that are conducted on and with dimensionality reduction techniques. This method allows to define upcoming experiments based on previous results. Whilst simultaneously aiming to view the technique from multiple angles.

This paper will start with an exploration of related work in section 2. The semi-structured photos being used in the experiments are from the project *exactitudes*, which will be explained in section 3. These photos will undergo, amongst other things pre-processing and dimensionality reduction steps. These steps are combined in an encoder and described in section 4. In section 5 the experiments will be described individually, following the classic pattern of describing the method, results and observation.

2. Related work

The related works on which this study will be based is described in this section. It will start with an introduction into dimension reduction techniques followed by the subfield of feature extraction and ending in the application domain, specifically on semi structured photos.

2.1 Dimensionality reduction

2.1.1 Principal component analysis

A typical approach to dimensionality reduction is the linear form of Principal Component Analysis. The aim of Principal Component Analysis is to discover the importance of dimensions in describing the variance of the data. When applying this technique the goal is to discover a number of principal components (hopefully a smaller number than the original dimensions) that are still able to reproduce the original data in an acceptable manner. How to get to these principal components in described by Smith[8] and can be divided into four steps.

The first step is to loop though a multidimensional dataset and subtract the average of that dimension from every datapoint in the dimension. Resulting in a dataset where the mean of every dimension is 0. Secondly a covariance matrix is calculated from the normalised dimensions. In step three, eigenvectors and eigenvalues can be calculated from this covariance matrix. These eigenvectors are the principal components of the dataset and can now be sorted in descending order on their eigenvalue. Since the eigenvectors corresponding with the highest eigenvalues is the most significant in explaining the original dimensions, those are the principal components to keep. Discarding the eigenvectors with the lowest eigenvalue will result

in reduced dimensions having a loss as minimal as possible.

2.1.2 Auto encoding neural networks

Although there exist nonlinear variants of PCA, a more common method of nonlinear dimensionality reduction is the use of auto encoding neural network. These networks have been known for quite some time[4], but gained revived interest since the work of Hinton[3] and the research succeeding this paper. As Hinton shows in his 2006 paper, the dimensionality reduction using Restricted Boltzmann Machines (RBM) is possible. These reduced dimensions represent features of the original data.

RBM's are a type of logistic feed forward neural networks that have a visible layer (input layer) and a single hidden layer (feature layer). The layers are fully connected, however there are no connections between the units in a layer, hence the restrictedness. Training an RBM is surprisingly simple, data is presented at the visible layer, the probability of a unit in the hidden layer becoming a 1 is determined using an energy function and the weighted edges. This forms a representation of the original data in the hidden layer, this is called a high energy state. The representation in the hidden layer will then be reconstructed in the visible layer using the same edges and energy function, this is called a low energy state. The original data and the reconstruction can now be used to update the weights using an algorithm called contrastive divergence[4]. An example RBM can be seen in figure 1, notice the full connectivity between the visible and hidden layer:

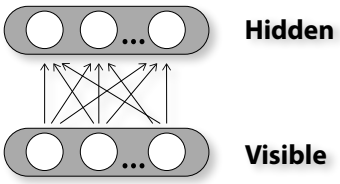


Figure 2.1: Structure of an RBM

The dimensionality reduction achieved by Hinton was realised by stacking these RBM's on top of each other. Where the hidden layer of the lower RBM provided the input for the visible layer of the next RBM. This deep network was trained layer by layer, while gradually reducing the number of units in the next hidden layer.

Lee has further developed RBM's and in 2009 he proposed his idea for Convolutional RBM's [5].

The idea of a convolutional RBM's is to take a subsample of the data also referred to as a *shift window* to form a representation in a single unit in the hidden layer. The structure of an convolutional RBM can be seen in figure 2, with characteristics such as the restricted connectivity between the visible and hidden layer and the addition of a pooling layer.

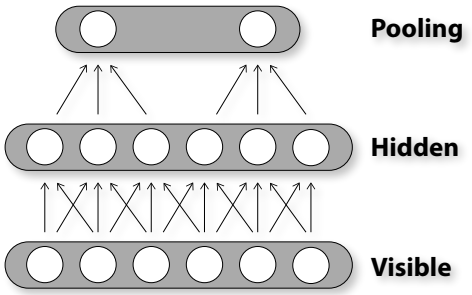


Figure 2.2: Convolutional RBM with a shift window of 3

Convolutional RBM's also consist of a pooling layer. The two most common pooling techniques are max-pooling and mean-pooling. Max pooling values the unit in the pooling layer at 1 if one of the connected hidden units is 1. Whereas mean pooling sets the pooling unit to 1 if the average of connected units exceeds a threshold.

The initial idea behind convolutional RBM's (C-RBM) is that because of the limited connectivity and the shrunken features in the pooling layer, it would make it more practical to feed high dimensional data into a neural network [5]. Besides this scalability advantage, it is also known that convolutional RBM's can outperform normal RBM's in some cases [1, 2, 5].

2.2 Feature extraction

Closely tied to dimensionality reduction is feature extraction. High dimensional data has become ubiquitous and in potential this should lead to more information with higher accuracy[11]. However, Kittler described an important paradox in feature extraction: The more features we have the more difficult information extraction becomes. This phenomena is also known as the *curse of dimensionality*. Where low dimensional data does not contain useful information and to high dimensional data makes it difficult to train for example a classifier. Because the more dimensions there are the more similar they appear[7].

This explains the close relationship of dimensionality reduction and feature extraction

because feature extraction is about making a lower dimensional representation of some higher dimensional data [11]. This means that a high dimensional space is transformed or mapped onto a lower dimensional space whilst staying informative.

Another approach is *feature selection,* whereby a subset is taken from the original high dimensional data. This procedure is without any transformation of the data[7, 11].

Both approached have in common that they try to reduce the dimensionality of data by discarding redundant or less important dimensions. Whilst making the assumption that the reduced dimensional space gives a reasonable approximation of the original space.

2.3 Dimension reduction as a generative model

When an auto encoding neural network is trained in a sufficient way, the network stores the shared information of the dataset, whilst the output nodes only have to represent the discrepancies[17]. This entails that the reduced dimensions not only describe the input data itself, but also the input data in the context of the whole dataset. Whilst the auto encoding neural network can also be seen used as a tool to discover the inherent dimensions of the data. One way to obtain these insights is to turn the network upside down and let it reconstruct images.

DeMers and Cottrell did this by taking two 5 dimensional output vectors and reconstructing from equally spaced points on the line joining them[4]. Resulting in two distinctive faces, pulling towards each other, while producing recognisable faces at every point in between. By using their network as a generative model, the underlaying features of the dataset could be visualised. In addition the network shows that it is capable of generalisation within the strict visual format of the used portraits. However, When an auto encoder is presented with deviating input, the difference between the input and reconstruction tells something about the novelty of the input[17].

A more contemporary example of image generation are the *deep dream* images from Google Research[18]. In an attempt to grasp what is learned in their image classification network, researchers at Google have also used their network in reverse to generate images. Starting with random noise being fed into a network that recognises for example dogs. Although the noise almost certainly contains no dog, the network is still looking for patterns that it associates with this animal. Probably with low confidence, the network will find a patterns and when a reconstruction is made based on these features, even minuscule dog patterns will be amplified. By repeating this process several times the resemblance of what the network has learned as a dog becomes more and more visible.

This produces esthetical interesting images to look at, but it might also expose learning mistakes. When this procedure was applied to the doorbell recognition network, they always became visible with an arm attached to it. Indicating a lack of trainings data of freestanding doorbells.

Prior to the *deep dreams* another study already came with a visual method of exposing the limitations of auto encoding neural networks[19]. This research also uses white noise as a starting point. However, the generation of images is not done within the network but separate using a evolutionary algorithm, whereby the classifying neural network serves as a fitness function to get for example an image that is recognised as a guitar. Since the termination condition for this evolutionary algorithm can be set at any desired confidence level of the neural network, images can be produces that are more guitar-like than a real guitar image according to the network. It is remarkable however unrecognisable shapes to humans can be well performing, classifiable images for a neural network. Again giving an insight in the shape and colours a neural network finds distinguishing.

2.4 Auto encoding neural networks and semi-structured natural images

Semi-structured are a typical application domain for feature extraction. Both Hinton[3] and DeMers[4] use portraits of faces with a strict format and visually normalised as a test-case for their bottleneck shape networks. Both studies have the aim of reducing the amount of information whilst still be able to reconstruct without noticeable errors. The semi-structured nature of the images on one hand makes it easier for a network to model, since the deviation within the dataset is not so large, whilst on the other hand small and specific details have an enhanced distinctive character that need to be captured in the model.

A well known benchmark dataset with semi-structured images is MNST, which consist of

handwritten digits in a highly normalised format. This dataset is widely used by scientist to train and compare handwriting recognition models, one of which is Hinton's Restricted Boltzmann Machine[3]. This model performed rather well, with an error rate of 1.2%.

However, the pixel intensities in the MNST dataset do not vary that much and are binary-like. Therefore, making a model of natural images is not just a matter of scaling up the network[6], since the Restricted Boltzmann Machine is a binary system. However, there are methods suggested to model natural images, such as making the visible layer gaussian[5]. As a downside this has a negative effect on the training time. Also pixel intensities are rarely independent in natural images, they most likely have a relationship with their surrounding, something that an RBM cannot take into account[6].

3. Exactitudes

The semi-structured images used in this research are those of the *exactitudes* series by Ari Versluis and Ellie Uytenbroek[13]. This project dates back to 1993 and aims to document subcultures omnipresent in society. A sample of every subculture is captured in a series which consist of 12 portraits, all systematically taken from the same angle, with a neutral background and similar body position. Examples of series can be found in Figures 3.1 until 3.4.

Ellie and Ari themselves appointed in an interview: “It should almost be a scientific anthropological record of people’s attempts to distinguish themselves from others by assuming a group identity”[16]. Although these portraits where intentionally made as an art project, they do hold value for researchers. For example in the investigation of sensory experience and affect in relation to denim clothing[14]. In this research the author compares different ways of wearing jeans, using *exactitudes* as one of the resources.

The systematic and accurate nature by which the photos where taken and curated also make them a useful dataset for this research on semi-structured images. Especially since the visual similarity is not only present within a serie but also across series.

The project is at the time of writing (august 2015) still ongoing and consist of 154 series with 12 photos each. After a manual evaluation of the series, five series where classified as outliers because the composition deviated to much from the majority of series. An example outlier are the ‘Gabbers’ in Figure 3.2. The photos in this serie only show the head and shoulders instead of from the upper legs such as in the majority of series.

The resulting dataset consists of 1788 colour photos (149 series, 12 photos each) of 600x600 pixels. This dataset will undergo pre-processing before being reduced in dimensions.

Figure 3.1: ‘Meuf’ serie 122 from exactitudes

Figure 3.2: ‘Gabbers’ serie 1 from exactitudes

Figure 3.3: ‘Annazaranina’ serie 143 from exactitudes

Figure 3.4: ‘United Americans’ serie 154 from exactitudes

4. Encoder

To explore the semi-structured photos first the dimensionality will be reduces. This reduction will be done using an encoder described in this section.

4.1 Pre-processing

Despite that the dataset is already in a strict format, there must still be some pre-processing steps applied. This pre-processing has multiple reasons which will be further elucidated in the following sections.

4.1.1 Normalising

In the first step of pre-processing the image size will be reduced to 50x50 pixels and converted to 8-bit greyscale. Although this process is irreversible, it allows the experiments to be executed within a feasible timespan. Furthermore the pixel intensities will be normalised between and including (0, 1) since the Restricted Boltzmann Machines require values within this range.

4.1.2 Enhancing sparsity

It is known that sparse representations have a number of benefits for energy based learning models such as Restricted Boltzmann Machines[15]. Due to the more abstract representation there is an increased likelihood of correct classification, perhaps even linearly separable when using sparse coding[15]. Sparse meaning that the input vector consist mainly of zeros, with the exception of a lower number of non-zero input dimensions.

The normalised dataset is not sparse, as shown in Figure 4.1. To achieve this sparsity within the normalised dataset the contrast of the photos is enhanced using the sigmoid like contrast function in Formula 1. The formula forces small pixel intensities to approach zero and large ones to approach 1.

$$contrast(x) = \frac{1}{1 + e^{g(c-x)}} \tag{1}$$

Every input pixel value (x) is updated individually. The 'steepness' of the sigmoid is determined by constant g (gain) which is set to 40 and the c constant determines the cut off between large and small pixel intensities, which is set to 0.25. Both values where experimentally determined based on the resulting distribution of pixel intensities over the dataset, which can be found in Figure 4.2

The contrast enhancement creates a binary-like vector that can be reversed without any loss of information with formula 2.

$$contrast^{-1}(x) = \frac{-ln(\frac{1}{x} - 1)}{g} + c \tag{2}$$

The resulting input images however have a majority of non-zero (information) values due to the light background and darker foreground. To enable sparsity and prevent learning the background, the inverse of pixel intensities will serve as input for the dimensionality reduction.

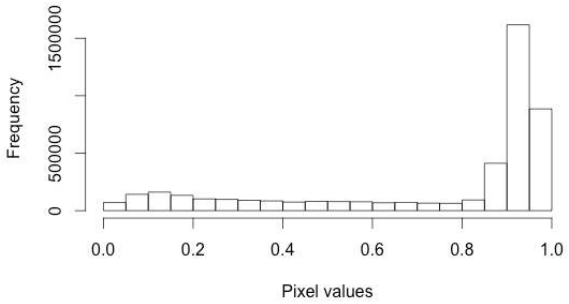


Figure 4.1: Histogram of pixel values before contrast enhancement and inverse

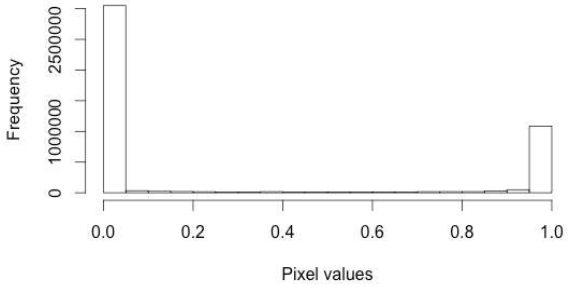


Figure 4.2: Histogram of pixel values after contrast enhancement and inverse

4.2 Dimension reduction

What remains after pre-processing is a 2500 dimensional sparse image space containing all 1788 photos. In order to extract features or do classification the dimensionality of this space will be reduced by stacking RBM's on top of each other. Every hidden layer will serve as the visible layer for the next RBM, whilst gradually decreasing the number of hidden units.

The full structure of the RBM starts with the original 2500 (50x50) dimensions and goes to a broader feature space of 4000, as shown in Figure 4.3. The reason being that the binary like input contains values that approach 0 and 1 but are not exactly. These subtle deviations need to be captured requiring a broader binary layer in comparison to the input, a strategy that is previously been applied by Hinton[3]

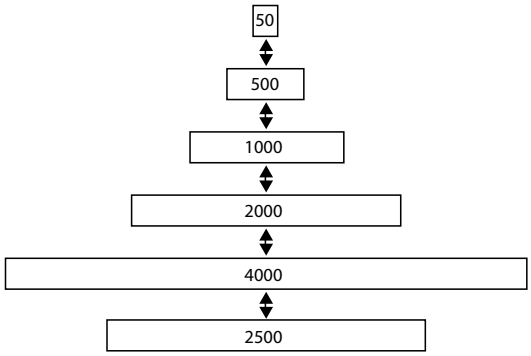


Figure 4.3: Flow of dimensionality within the RBM stack

Furthermore the stack gradually narrows down, again following the same structure as Hinton[3] until it reaches a final funnel of 50 dimensional feature space. This bottleneck is determined by assessing the reconstruction quality when features in the bottleneck layer propagate back, from top to bottom through the RBM stack. Different size bottlenecks have been tried (90, 70, 50 & 30). However, 50 dimensions was the smallest feature space, still able to give a reasonable reconstruction when eyeballing the results.

The cascaded methods from pre-processing through the RBM stack will be referred to as the *encoder*. As previously indicated, the encoder can be used in both directions, with the exception of the reduction to 50x50 pixels and the conversion to grayscale. A selection of these results, made by the author, based on visual variation and reconstruction quality can be seen in Figure 4.4.

There is a clear loss of information from the encoder and reconstructing back, as to be expected from a 50 dimensional funnel. However, the human shape is still visible as well as posture, body position and colour intensity of clothing.

Figure 4.4: From left to right: Normalised photos, contrast enhanced photos, reconstruction from RBM with 50 feature funnel and the same reconstruction after the contrast is reversed

4.3 Decorrelating

When an original photo is put through the encoder, the result is a 50 dimensional feature vector. Yet, RBM's do not necessarily make an ordering in these features and it is probable that these features are related to one another.

The last (optional) segment of the method is going from the feature space to a decorrelated feature space by means of principal component analysis. From the 50 features of every photo in the dataset, the 50 principal components, or decorrelated features are calculated. This decorrelated feature vector is now ordered based on the variance every feature explains and is used without the removal of the least significant ones. This allows a lossless transformation back to the previous state. The encoder with the additional decorrelation step will be referred to as the *decorrelated encoder.*

5. Experiments

The exploratory nature of this research will emerge in this section. Every experiment is described following the structure: method, results and observations. In this workflow, the results of one experiment may lead to an hypothesis that will be tested in a following experiment.

5.1 Straight path between two feature vectors

5.1.1 Method

Every feature vector out of the (decorrelated) encoder is a point within (decorrelated) feature space. Meaning that there is a straight path between two points that can be visualised, since the encoder is able to reconstruct.

In this experiment the euclidean distance between every possible combination of feature vector pairs is calculated, to retrieve the 50 most distant vectors pairs. A reconstruction of the straight path between these feature vectors is made in 10 equidistant steps. It is noteworthy to mention that the two most distant pairs in feature space do not necessarily have a great distance in decorrelated feature space

5.1.2 Results

From the resulting 50 pairs, a subset is chosen based on visual disparity and can be found in Figures 5.1 and 5.2. The complete set of results can be found in Appendix 1.

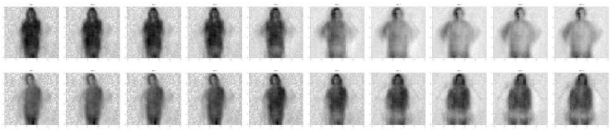


Figure 5.1: Two straight paths between two distant feature vector pairs in feature space

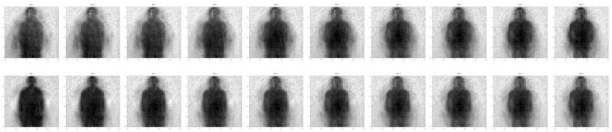


Figure 5.2: Two straight paths between two distant feature vectors in decorrelated feature space

5.1.3 Observations

The most left and right reconstructions in Figure 5.1 show a clear visual difference in colour intensity, posture and body position. Whilst the outer reconstructions from the decorrelated feature space (Figure 5.2) have less visual differences, despite their relative high euclidean distance. Indicating that the distance in feature space is more related to visual differences in contrast to the distance in decorrelated feature space.

Regardless of the position in (decorrelated) feature space, every vector on the straight line reconstructs as an image that could be assessed as human-shaped. An observation which is being further exploited in the next experiment: Random points in (decorrelated) feature space.

5.2 Random points in feature spaces

5.2.1 Method

From the previous experiment one might hypothesise that every arbitrary point in (decorrelated) feature space can reconstruct into an image with clear human shapes. In this experiment 100 random point in both feature and decorrelated feature space are taken, based on the uniform distribution. These random points are reconstructed into images for further review.

5.2.2 Results

Out of the 100 reconstructions in feature space and 100 reconstructions in decorrelated feature space, a random selection is presented in respectively Figures 5.3 and 5.4. The full two times 100 reconstructions can be found in Appendix 2.

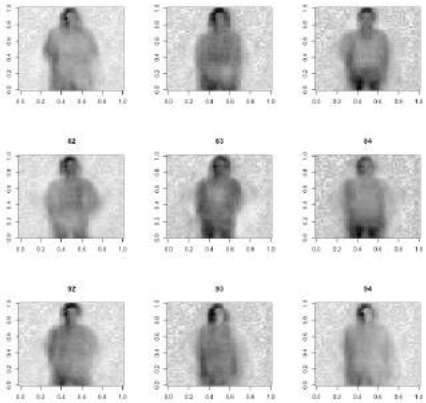


Figure 5.3:Reconstruction from random points in feature space

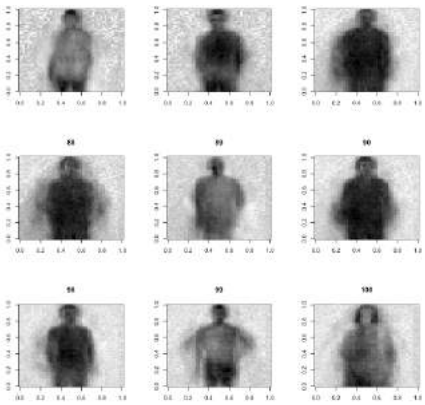


Figure 5.4: Reconstructions from random points in decorrelated feature space

5.2.3 Observations

The hypothesis that every arbitrary point in (decorrelated) feature space can reconstruct in images with recognisable human shapes is supported by the results. This is not limited to the random subset shown in Figures 5.3 and 5.4, but across all the 200 reconstructions that can be found in Appendix 2.

Furthermore, it is notable that the colour intensity differs when comparing the reconstruction from feature space with those of the decorrelated feature space. The latter being noticeably darker in contrast to the reconstructions from feature space.

Visual differences can be expected, since the additional PCA transformation in the decorrelated encoder changes the feature space in a linear way. Nonetheless, it does indicate that the distribution of feature vectors across the decorrelated feature space has strongly changed. The transformation placed the darker photos throughout the space whilst the lighter ones are closer together. This observation might also explain the lack of visual differences in Figure 5.2.

5.3 Activation distributions in feature spaces

5.3.1 Method

The following experiments will be a more in-depth exploration of the individual features in both the feature space and the decorrelated feature space. This experiment will be a box plot visualisation of the feature activations per feature based on the entire dataset.

5.3.2 Results

The features in decorrelated feature space are ordered based on the variance each one does explain. Thus the first is the most important, subsequently decreasing until the least explanatory feature. Figure 5.6 shows the first and last three features in decorrelated feature space and their activations.

In the feature space there is no particular ordering in features, therefore a random subset of features and their activation distribution is shown in Figure 5.5. The complete results can be found in Appendix 3.

Activations are considered an outlier if the value is outside 1.5 times the interquartile range above the upper quartile or bellow the lower quartile. Outlier positions are marked with a small circle in both box plots.

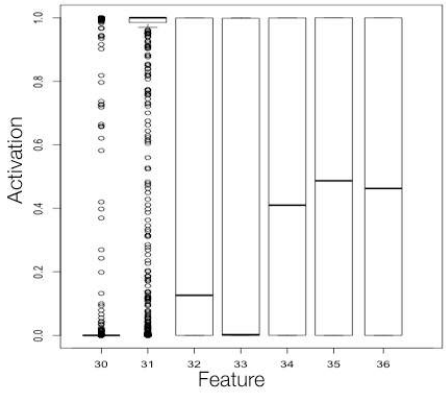


Figure 5.5: A box plot of activations in random features in feature space across the whole dataset

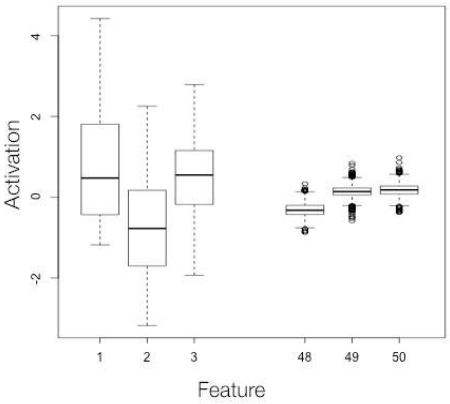


Figure 5.6: A box plot of activations in the first and last three features in decorrelated feature space across the whole dataset

5.3.3 Observations

As to be expected, due to the order in the decorrelated feature space, the first features in Figure 5.6 show more variance in comparison to the last. The decreasing of explained variance is clearly visible in this box plot.

The activations in feature space are different because the order is arbitrary. Also, there are many differences across the features, as is visualised in

Figure 5.5. For example feature 30 is fairly outspoken at zero with a few exceptions. Whilst feature 34 is about similar in high and low activations.

A noteworthy difference between the two spaces is the distribution of the activations. The activations in the decorrelated space is mostly around the centre whilst the activations in the feature space are mostly located on the outer edges. For the latter, a potential explanation is the probabilistic nature of Restricted Boltzmann Machines[3].

5.4 Highest and lowest activating photos per feature

5.4.1 Method

It is known from experiment 5.3 that feature activations in feature space are disordered and can be outspoken. This is in contrast with the decorrelated feature space where features are ordered and congregate around the centre. In this experiment all photos are ordered on their activation for every feature individually. Resulting in the highest and lowest activating photos per feature for both the feature space and decorrelated feature space.

5.4.2 Results

A subset of the highest and lowest activating photos per feature can be found in Figures 5.7 and 5.8. This subset is based on the observations of the previous experiment. Moreover, the complete results can be found in appendix 4.

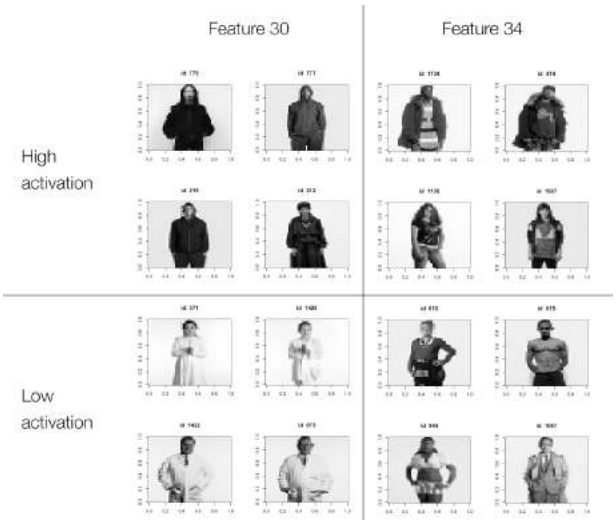


Figure 5.7: Four high and low activating photos on features 30 and 34

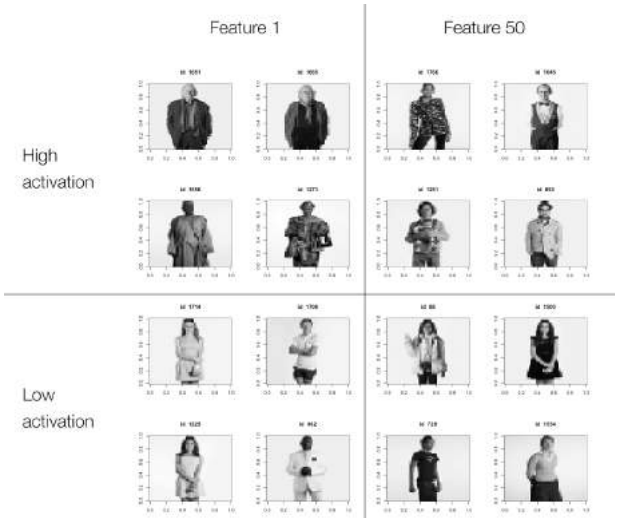


Figure 5.8: Four high and low activating photos on decorrelated features 1 and 50

5.4.3 Observations

The previous experiment showed that over the complete dataset, the activation of feature 30 in feature space is generally situated on the outside of the space, a lower value in this case. Figure 5.7 shows that these outspoken activations have a tendency to explain visual differences of photos. In this example the feature explains the colour intensity of the photo. In contrast, a less outspoken feature such as 34 is also less outspoken in the visual differences it explains.

The ordering in the decorrelated feature space gives reason to have a closer look at the first feature. Figure 5.8 shows the high and low activating photos on the first feature and expose a difference in both colour intensity as well as posture. Again the less outspoken activating photo's, explain differences visually less obvious.
Arguably, the partition that is made by the feature activations is more evident in the feature space compared with the decorrelated feature space. The merging of visual differences in the decorrelated feature space, such as colour intensity and posture could adversely affect the subdivision. This is also evident in the complete results in Appendix 4.

5.5 Single feature variations in context

5.5.1 Method

In order to visualise the effect of each feature on the reconstruction, this experiment will set every feature individually to the highest and lowest boundary of the (decorrelated) feature space. This single feature will be varied in a existing photo in (decorrelated) feature space. The reason being that the existing photo can provide context in the visual assessment of the reconstructions.

5.5.2 Results

The full results consist of the high and low feature variations, for every feature on a randomly selected subset of 10 photos. There are results for the correlated as well as the decorrelated feature space. A subset based on the features used in previous experiments are to be found in Figures 5.9 and 5.10, starting at (decorrelated) feature 30 and showing low on the left and high on the right. Appendix 5 will provide a random selection of the results for this experiment

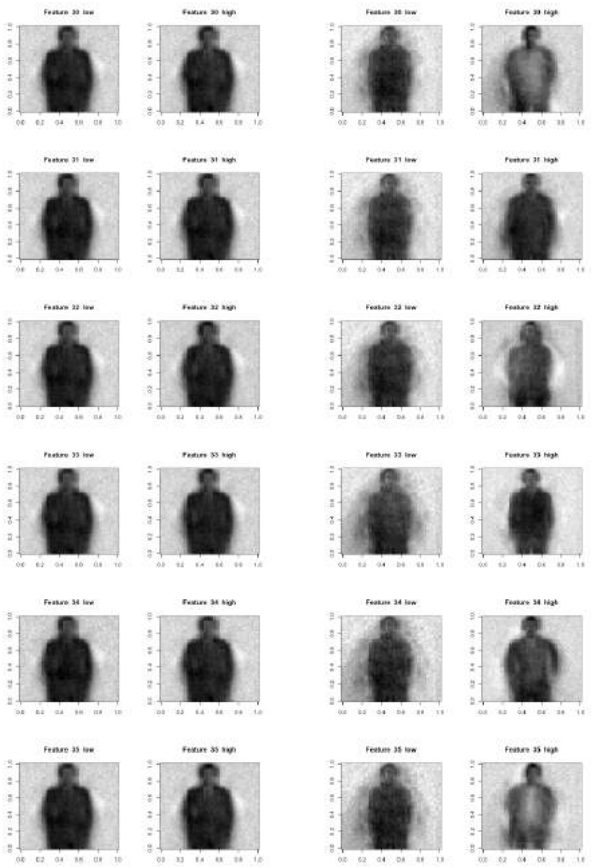


Figure 5.9: Single feature variations in feature space

Figure 5.10: Single feature variations in decorrelated feature space

5.5.3 Observations

When observing Figure 5.9 it is immediately apparent that it is lacking visual diversity. Although there is a measurable difference in pixel intensities, the results have hardly any advantage to evaluation by the human eye.

In the reconstructions from the decorrelated feature space there are differences visible (Figure 5.10). For example, when feature 30 is set to a high value, the clothing in the upper body becomes lighter. This result can be placed in addition to the results in Figure 5.7, where the activations on the 30th feature also predicts the colour intensity of the clothing. However, it is noteworthy that in Figure 5.7 concerns with the feature space whereas Figure 5.10 shows reconstructions from the decorrelated feature space. Furthermore, the light colour intensity is in Figure 5.7 is associated with a low activation whereas the colour intensity in Figure 5.10 becomes lighter when feature 30 is set to a high value.

Notwithstanding, feature 34 shows, in lesser extend, a similar pattern with arm positions. In the 34th feature of Figure 5.10, both arms are partially distant from the body when the feature is set to a high value. Figure 5.7 shows again a similar pattern on the low activating photos on feature 34 in feature space.

Although this pattern emerged at two features it is highly unlikely that these two are actually related since the linear transformation of the decorrelated encoder not only transforms the space but also reorders the features. It is therefore likely that these observations are based on coincidence.

5.6 All features low except a single feature

5.6.1 Method

Encouraged by the results of the previous experiment, the question arose if single feature variations could be exploited further. The experiment showed reconstructions with a variation in a single features, while the remaining features where inherited from an existing photo: the context. This experiment differs in that reconstructions will be taken out of their context. All features will be set to zero with the exception of a single feature, that will be set to one. Reconstructions will be made for all features. In addition, the mean will be calculated on all reconstructions and subtracted from each individual reconstruction to enhance the differences between every reconstruction.

5.6.2 Results

A single, well performing, feature is chosen by the author and presented in Figure 5.11 for both the feature space and the decorrelated feature space. The complete results can be found in Appendix 6.

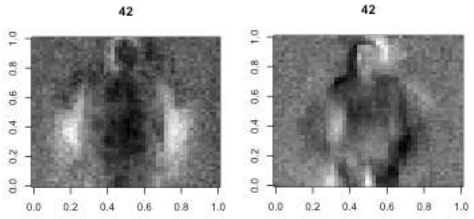


Figure 5.11: Reconstructions of all zeros except feature 42 which is set to 1.
Left: feature space, right: decorrelated feature space

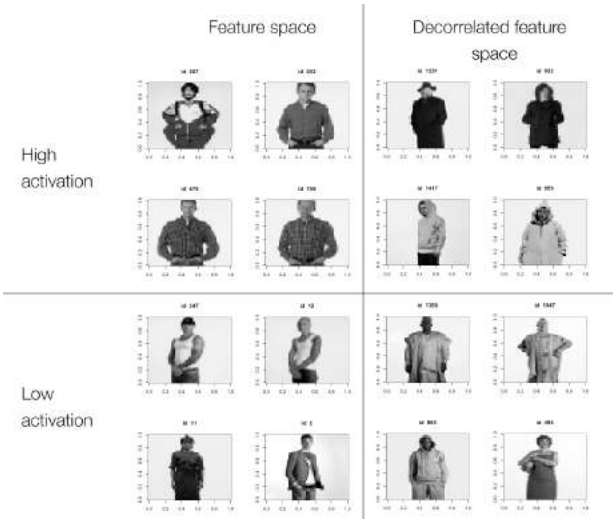


Figure 5.12: High and low activations on feature 42 in feature space and decorrelated feature space

5.6.3 Observations

The reconstructions on all features show high and low scoring areas on various body parts, which is visible in Figure 5.11. However, a single, randomly selected feature will be discussed. The remaining features can be found in appendix 6. High scoring areas are visible in these reconstructions as lighter pixel intensities and a low scoring is represented by dark pixel intensities.

What stands out in left reconstruction of Figure 5.11, is the high scoring area on the arms. Whilst the right reconstruction from the decorrelated feature space does not show a high scoring on a particular area, but is more spread among small patches.

If these observations are compared with the high and low activating photos on feature 42 (Figure 5.12), a link can be found. This Figure shows that the arm positions are different, an observation that feature 42 in feature space seems to explain. This is not the case with the decorrelated feature space, where there is, arguably and in general, a lesser visual relationship between the high scoring areas and the high and low activating photos.

5.7 Shape of the feature spaces

5.7.1 Method

Visualising the complete 50 dimensional (decorrelated) feature space is hardly possible in a meaningful way. However, the distribution of feature vectors over the feature space can be essential information when, for example, a classifier is trained. Although there is a loss of information, by making a histogram of the distance from the centre of the space for all 1788 photos, a rough impression of the space becomes visible.

5.7.2 Results

Figure 5.13 shows a histogram the euclidean distance of every photo from the centre of the feature space. Whilst the euclidean distance of every photo in decorrelated feature space can be found in Figure 5.14.

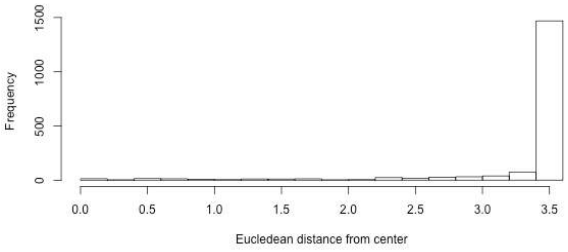


Figure 5.13: Distributions of euclidean distances from centre in feature space

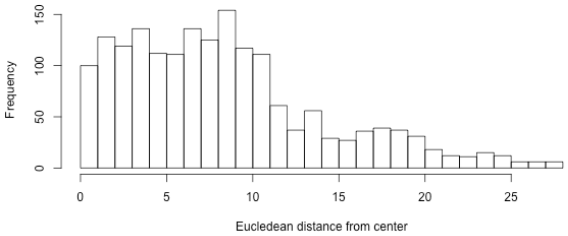


Figure 5.14: Distributions of euclidean distances from centre in decorrelated feature space

5.7.3 Observations

The distribution of photos across the two spaces is quite different, as to be expected after applying a linear transformation in the decorrelated encoder. In addition, the observations from Experiment 5.2 is strengthened by these results. Which settles that the photos in feature space are mostly situated on the outer edges of space. In contrast to the decorrelated feature space where the photos are situated close to the centre.

5.8 Feature pair visualisations

5.8.1 Method

This experiment will proceed on the results of the previous experiment and attempt to visualise the (decorrelated) feature space. A pair of features is plotted, starting with the first two, followed by *{3, 4}, {5, 6}, {…}*. Furthermore, a distinction between series will be made using a unique colour.

5.8.2 Results

The resulting plots of feature pairs in feature space and decorrelated feature space can be found respectively in Figure 5.15 and 5.16, starting in the top left corner, proceeding row by row whilst the first feature is plotted on the horizontal axis. Larger versions can be found in Appendix 8.

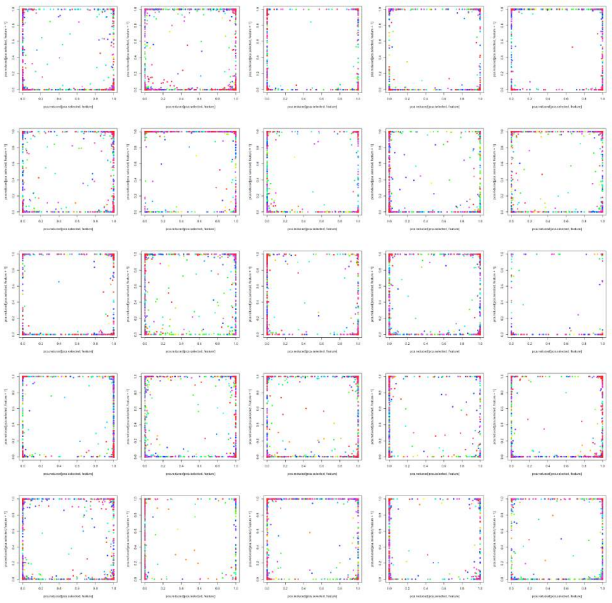


Figure 5.15:Feature pairs in feature space

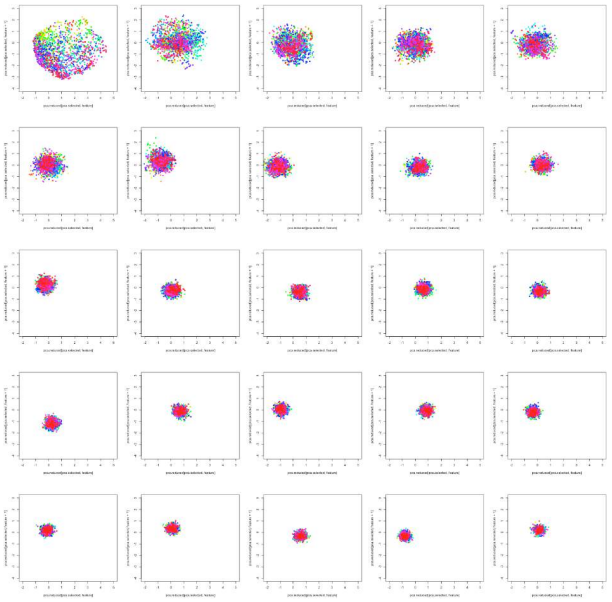


Figure 5.16: Feature pairs in decorrelated feature space

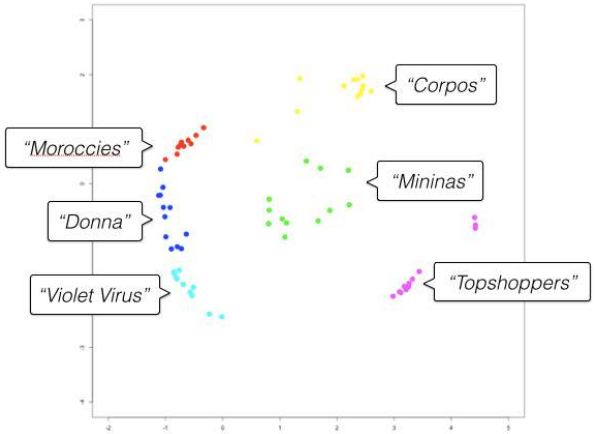


Figure 5.17: A subset of series on the first two features in decorrelated feature space

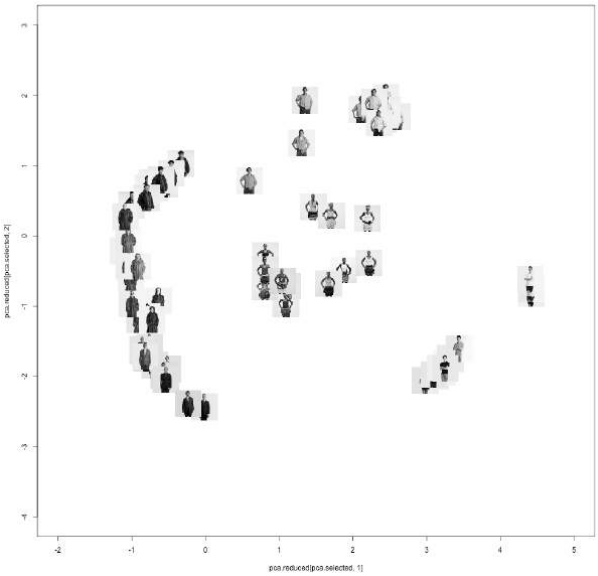


Figure 5.18: A subset of series on the first two features in decorrelated feature space, presented as photos

5.8.3 Observations

Features in feature space have no particular ordering, but they show a conspicuous pattern where the features are distinctively on the edges of the space. Whilst the features in decorrelated feature space form a circular shape and clustering together. Also in Figure 5.16 the arrangement of features is clearly visible, wherein a ordering is made on the amount of variance explained by each feature. Since this visualisation is mades with feature pairs, a minimal oval shape is visible because the two features explain different proportions of variation.

The series are marked individually with a different colour. It should be pointed that there are 149 series, which reduces the distinctive character of the colours for the human eye. However, when a subset of the first two features in decorrelated feature space is taken such as in Figure 5.17, the

distinct series become more visible. When the photos are laid across the plot (Figure 5.18), the arrangement of photos also seems meaningful considering the visual characteristics. This arrangement is able to separate series from each other but also works within series such as the 'Topshoppers' where light and dark photos are places at a distance from each other.

5.9 Classification of series

5.9.1 Method

Based on separation of series in Figure 5.17, the question arose how well a classifier would perform after being trained on the (decorrelated) features in order to predict the series.

The classifier used in this experiment will be random forest[7]. Whilst the dataset is randomly divided into a trainings set (n=1500) and a test set (n=288). The classifier is trained in three runs to get an idea of the repeatability of the training process. In each run the test and trainings set will be randomly repopulated.

The percentage of correct classifications on the test set is reported in the results section. For comparison a basic principal component analysis is performed on the normalised photos. The first 50 principal components will be used to train the classifier. Again, the three runs are divided into a test and trainings set similar to the previous procedures.

5.9.2 Results

	Features	Decorrelated features	First 50 principal components
Run I	62.2%	51.7%	64.2%
Run II	61.5%	53.1%	67.7%
Run III	62.8%	53.5%	66.7%
μ	62,2%	52,8%	66,2%

Figure 5.19: Percentage of correct classifications in three runs.

5.9.3 Observations

All three methods of classification have a reasonable success rate that clearly outperforms random classification *(1/149 ≈ 0.7%)*. Nevertheless it is noteworthy that the transformation and ordering of the decorated features seem to have a negative effect on the ability to classify series within the data, despite the visual distinctiveness that the PCA transformation adds in previous experiments. This makes it even more remarkable that the PCA alone produces the best classifiable data of the three. Although only a minor but consistent improvement compared to the features.

5.10 Best and worst classifiable photos

5.10.1 Method

This experiment will elaborate on the notable results of experiment 5.8. The relatively weak performance of the decorrelated features in classification were not in accordance with the expectations, therefore the best and worst classifiable series will be retrieved to get a better insight in the classification process.

5.10.2 Results

The resulting series where the same for both the feature space and decorrelated feature space. The best classifiable series are presented in Figure 5.20, whilst Figure 5.21 shows the least performing series in classification.

Figure 5.20: A subset of the five best classifiable series, series are displayed horizontally

Figure 5.21: A subset of the least classifiable series, series are displayed horizontally

5.10.3 Observations

The fact that the results were the same in feature space and decorrelated feature space indicates that the classifier is likely to learn the same differences across the series. However, the nuances are still able to make a difference, perhaps straightened by the linear transformation from the decorrelated encoder.

Also notable is the homogeneity within the series in Figure 5.20. Homogeneity is a characteristic of the photo project, however this is to a greater or lesser degree present in individual series. This is in contrast with the least classifiable series in Figure 5.21, which are more heterogeneous.

6. Discussion

Despite their popularity and interest from researchers, dimensionality reduction methods such as deep learning are commonly referred to as black boxes. This exploratory research aimed at giving an insight in the inner workings of these networks when applied to the semi-structured photos of the *exactitudes* photo project[13]. This understanding is obtained using an exploratory approach, one might not consider typical.

Given a dimensionality reducing network, referred to as the (decorrelated) encoder, a serie of small experiments was conducted. These experiments were of a manageable size, making it possible to develop new experiments based on previously obtained results and explore the (decorrelated) encoder from different viewpoints.

This approach has a number of noteworthy results that will be reflected on. However, the encoder itself has also a notable aspect. Namely, the use of a sigmoid-shaped contrast enhancement function to force the input dimensions into a binary-like vector.

Prior research did not indicate a similar approach to handle real valued data in Restricted Boltzmann Machines. In this research, the results showed that the network could learn the binary-like data, whilst still able to reverse the transformation in order to reconstruct the photos. Although this research only used photos as input, there is no reason to suspect that this approach is limited to this type of input.

Novel results also emerged from the experiments. For example experiments 5.3 and 5.4 together indicate that there is a relationship between the outspokenness of a feature and the visual differences that feature is able to explain. Wherein outspoken means that the average feature activation on the entire dataset is close to one or zero. This relationship shows that the more outspoken a feature is, the more visual discrepancy it is able to explain.

Furthermore the additional decorrelation step showed varying results. Experiment 5.3, 5.7 and 5.8 clearly visualise a transformation of feature space, as to be expected. However, due to this transformation, features that were distilled in feature space are intwined in the decorrelated feature space, as experiment 5.6 indicated. In this experiment a division is made on both posture as well as colour intensity, making it more difficult to select photos on an individual feature.

However, decorrelation also showed an advantage, by visually enhancing the separation between series such as in Figure 5.17. Whilst simultaneously discriminate on visual differences as Figure 5.18 shows. These results make it worthwhile to use decorrelation when eyeballing the visualisation for clusters or groups within the dataset.

However, a distinction should be made between the human eye and classification algorithms. Since the results of experiment 5.9 show that the classification of series is inferior on decorrelated features in comparison with the features. A result that is difficult to explain from a computational viewpoint.

The same experiment also shows that classification on the first 50 principal components of the normalised photos, outperforms the previous methods. All classifications are doing much better than a toss with a 149 sided die. However, it is still noteworthy that a relatively simple technique such as Principal Component Analysis can still compete with contemporary dimension reduction algorithms. Despite the enormous popularity of dimension reduction or deep learning, it remains worthwhile to make a comparison with well established techniques.

Furthermore, what this research has yielded in a broader sense it the approach itself. Because of its exploratory nature, the path to follow is less clear in comparison with a more traditional research question. Thus, there is an increased need for structure and guidance. Performing small and manageable experiments is, to our best knowledge, novel for exploratory research into deep learning. The method allows new insights to be fitted directly into new experiments. Thus adjusting the direction of the research based on preliminary results and insights. Although one must ensure in a exploratory research that the problem is viewed from multiple angles. Something we have done to our very best in this study.

When looking back at the exploratory goal, we hope this research has made a positive contribution to the understanding of deep neural networks. By means of various visualisations, a better understanding of what is learned it these networks could be obtained. This understanding could not only help in the selection of relevant features but also bring light to possible shortcomings in the dataset or previously unknown features that might be beneficial. Therefore a good understanding of what is learned adds considerable value in exploring and exploiting large datasets.

References

[1] Lee, H., Pham, P., Largman, Y., & Ng, A. Y. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems* (pp. 1096-1104).

[2] Tran, S. N., Wolff, D., Weyde, T., & Garcez, A. (2014). Feature Preprocessing with RBMs for Music Similarity Learning. *learning*, *9*(8), (pp. 16-16).

[3] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), (pp. 504-507).

[4] DeMers, D., & Cottrell, G. (1993). Non-linear dimensionality reduction. *Advances in neural information processing systems, 5,* (pp. 580-580).

[5] Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009, June). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 609-616).

[6] Krizhevsky, A., & Hinton, G. E. (2010). Factored 3-way restricted boltzmann machines for modeling natural images. In *International Conference on Artificial Intelligence and Statistics* (pp. 621-628).

[7] Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. Elsevier.

[8] Smith, L. I. (2002). A tutorial on principal components analysis. *Available from: http://www.sccg.sk/~haladova/principal_components.pdf. Retrieved on: 12 May 2015*

[9] Wolff, J., Martens, M., Jafarpour, S., Daubechies, I., & Calderbank, R. (2011). Uncovering elements of style. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1017-1020).

[10] Wu, T., Polatkan, G., Steel, D., Brown, W., Daubechies, I., & Calderbank, R. (2014). Painting Analysis Using Wavelets and Probabilistic Topic Models. *arXiv preprint arXiv:1401.6638*.

[11] Kittler, J. (1986). Feature selection and extraction. *Handbook of pattern recognition and image processing*, (pp. 59-83).

[12] Johnson Jr, C. R., Hendriks, E., Berezhnoy, I. J., Brevdo, E., Hughes, S. M., Daubechies, I., ... & Wang, J. Z. (2008). Image processing for artist identification. *Signal Processing Magazine*, *25*(4), (pp. 37-48).

[13] Versluis, A., & Uyttenbroek, E. (2002). Exactitudes. *010 Publishers*.

[14] Candy, F. J. (2005). The fabric of society: an investigation of the emotional and sensory experience of wearing denim clothing. *Sociological Research Online*, *10*(1).

[15] Poultney, C., Chopra, S., & Cun, Y. L. (2006). Efficient learning of sparse representations with an energy-based model. In *Advances in neural information processing systems* (pp. 1137-1144).

[16] Huffington Post. (2012) Artists Create Anthropological Photo Series. *Available from: http://www.huffingtonpost.com/2012/06/25/exactitudes-interview-art_n_1619483.html. Retrieved on: 25 July 2015*

[17] Gluck, M. A., & Myers, C. E. (2001). Gateway to memory: An introduction to neural network modeling of the hippocampus and learning. *MIT Press*.

[18] Google research blog. (2015). Inceptionism: Going Deeper into Neural Networks. *Available from: http://googleresearch.blogspot.ch/2015/06/inceptionism-going-deeper-into-neural.html. Retrieved on: 20 June 2015*

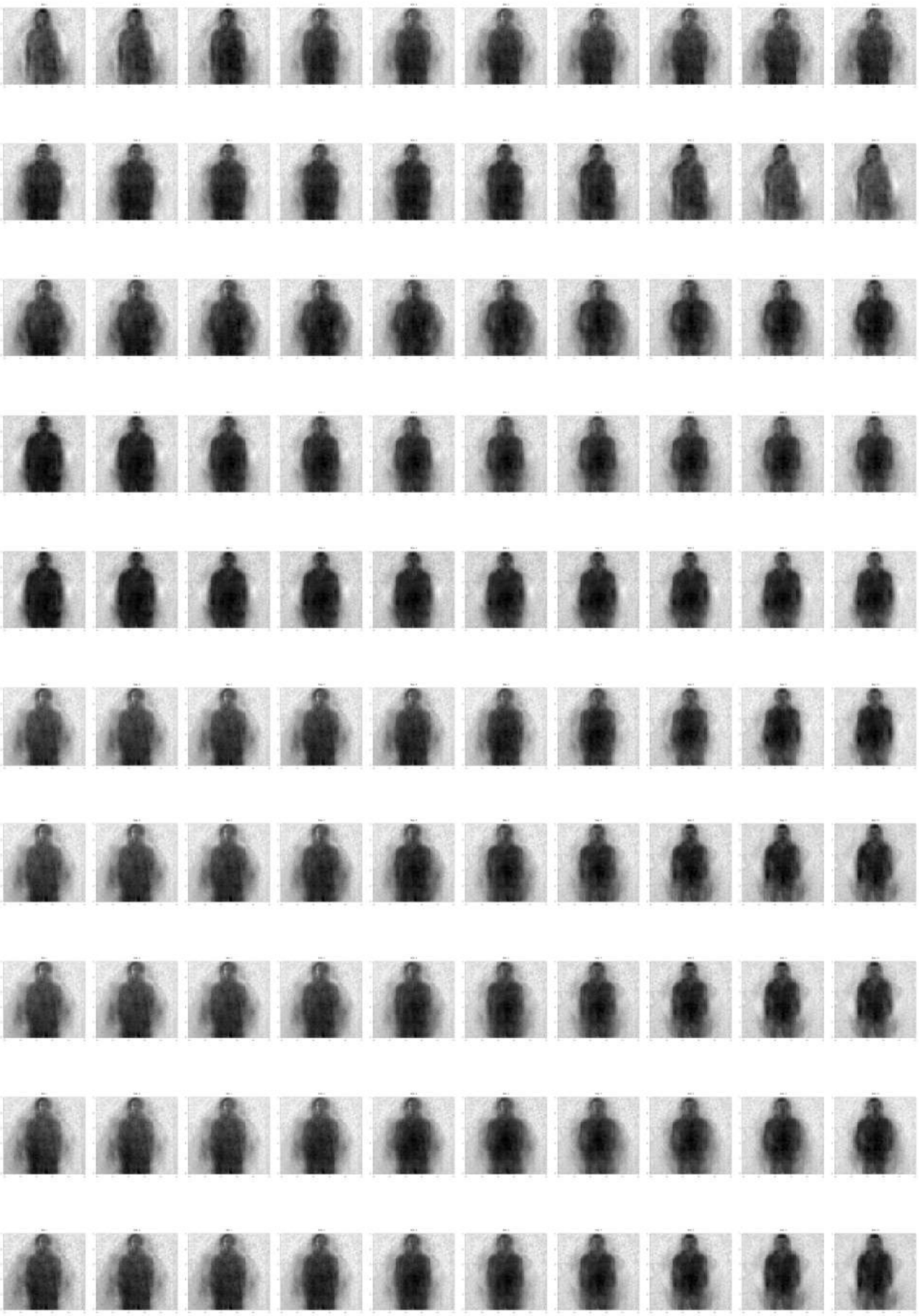
[19] Nguyen, A., Yosinski, J., & Clune, J. (2014). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *arXiv preprint arXiv:1412.1897*.

Appendix 1

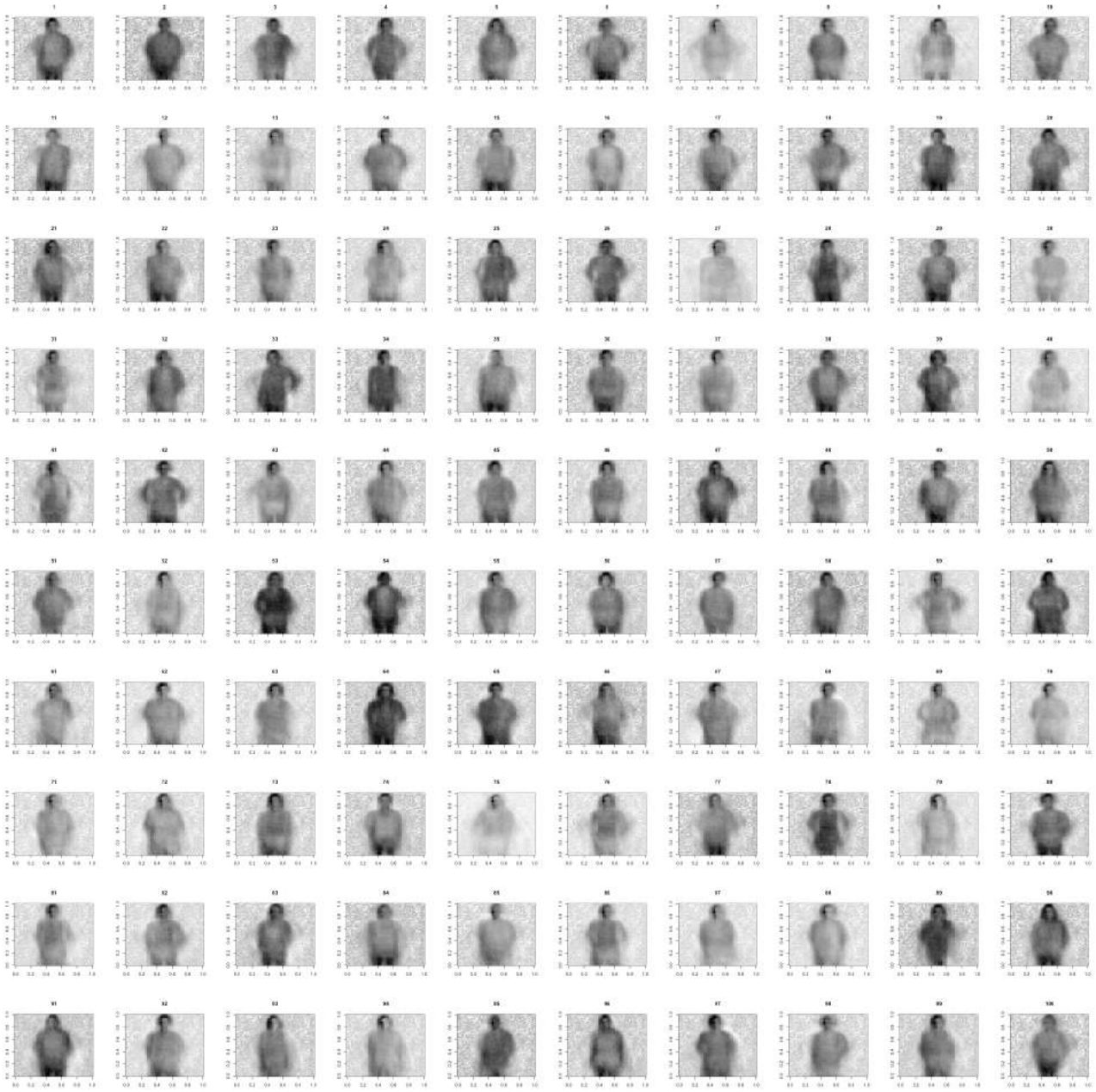
Paths in feature space

Pairs are horizontally displayed

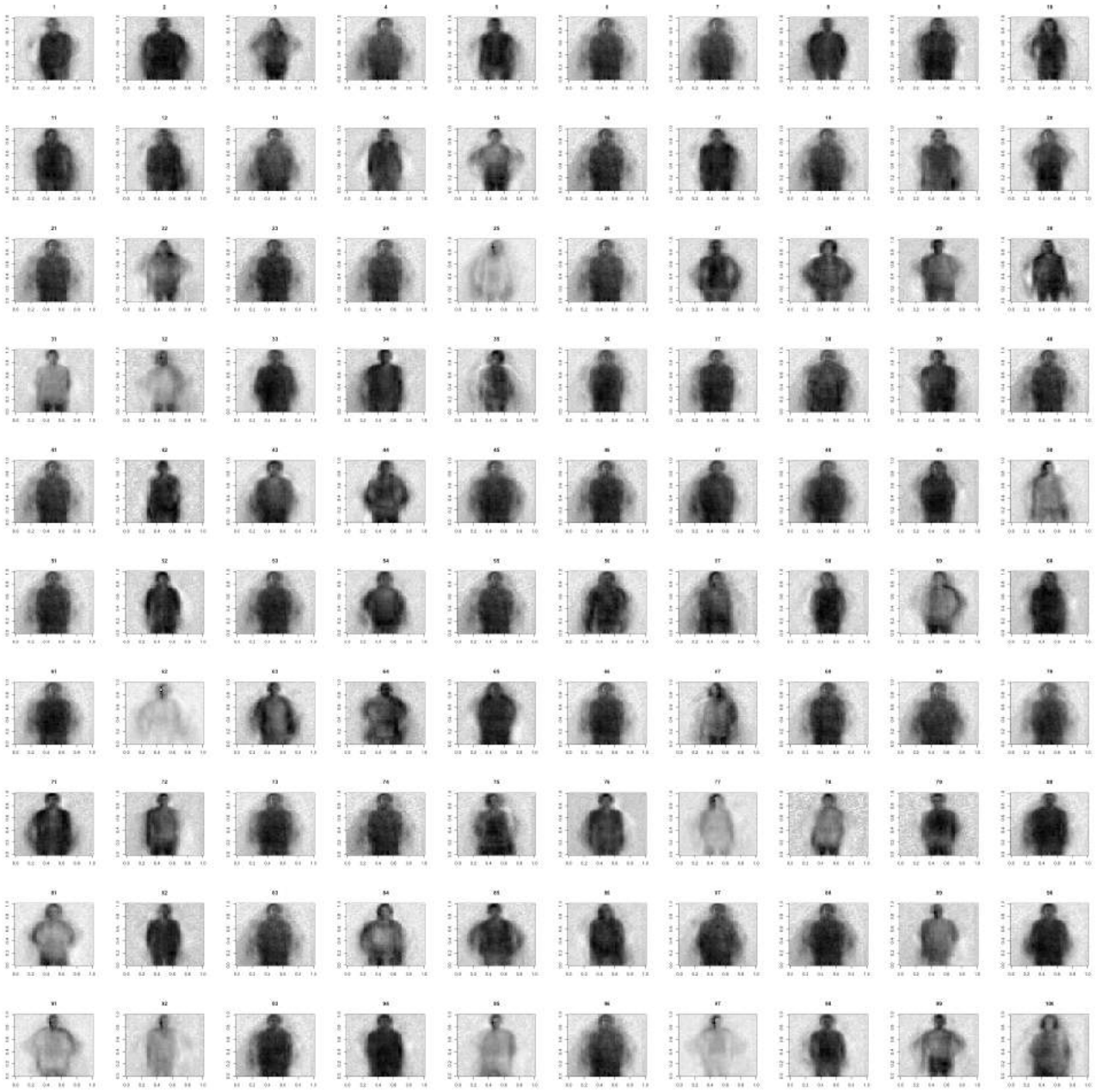
Paths in decorrelated feature space
Pairs are horizontally displayed



Appendix 2
Random points in feature space

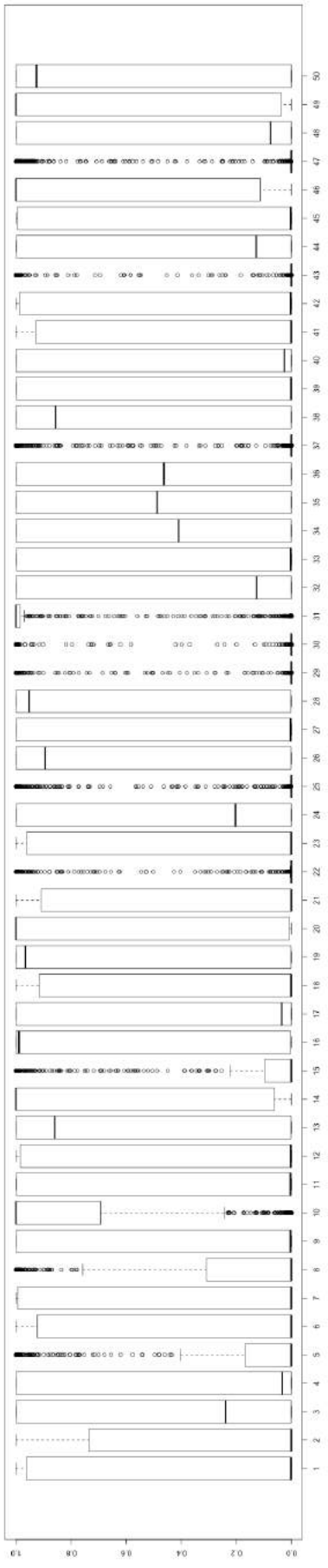


Random points in decorrelated feature space

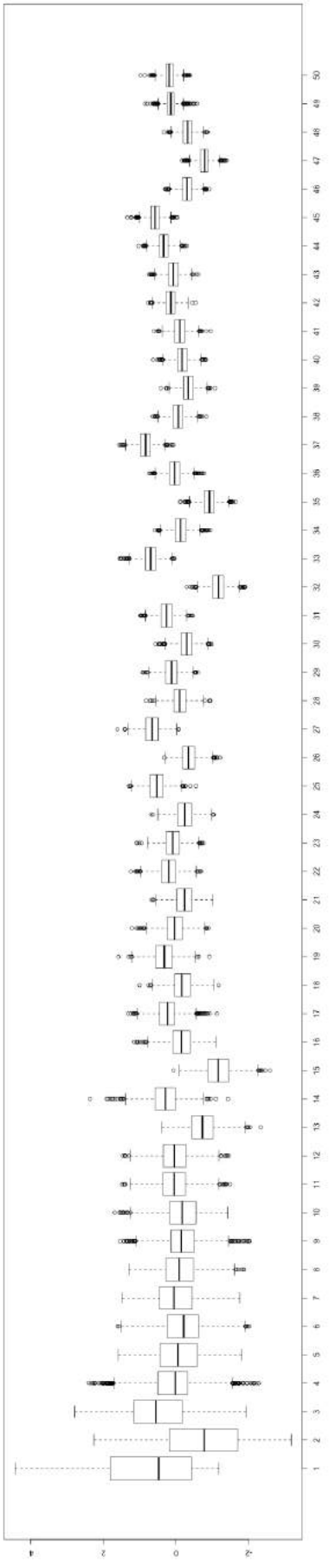


Appendix 3

Activations in feature space



Activations in decorrelated feature space



Appendix 4

The top two rows are high activating photos whilst the bottom two rows are the low activating photos

Feature 1

Feature 2

Feature 3

Feature 4

Feature 5

Feature 6

Feature 7

Feature 8

Feature 9

Feature 10

Feature 11

Feature 12

Feature 13

Feature 14

Feature 15

Feature 16

Feature 17

Feature 18

Feature 19

Feature 20

Feature 21

Feature 22

Feature 23

Feature 24

Feature 25

Feature 26

Feature 27

Feature 28

Feature 29

Feature 30

Feature 31

Feature 32

Feature 33

Feature 34

Feature 35

Feature 36

Feature 37

Feature 38

Feature 39

Feature 40

Feature 41

Feature 42

Feature 43

Feature 44

Feature 45

Feature 46

Feature 47

Feature 48

Feature 49

Feature 50

Decorrelated feature 1

Decorrelated feature 2

Decorrelated feature 3

Decorrelated feature 4

Decorrelated feature 5

Decorrelated feature 6

Decorrelated feature 7

Decorrelated feature 8

Decorrelated feature 9

Decorrelated feature 10

Decorrelated feature 11

Decorrelated feature 12

Decorrelated feature 13

Decorrelated feature 14

Decorrelated feature 15

Decorrelated feature 16

Decorrelated feature 17

Decorrelated feature 18

Decorrelated feature 19

Decorrelated feature 20

Decorrelated feature 21

Decorrelated feature 22

Decorrelated feature 23

Decorrelated feature 24

Decorrelated feature 25

Decorrelated feature 26

Decorrelated feature 27

Decorrelated feature 28

Decorrelated feature 29

Decorrelated feature 30

Decorrelated feature 31

Decorrelated feature 32

Decorrelated feature 33

Decorrelated feature 34

Decorrelated feature 35

Decorrelated feature 36

Decorrelated feature 37

Decorrelated feature 38

Decorrelated feature 39

Decorrelated feature 40

Decorrelated feature 41

Decorrelated feature 42

Decorrelated feature 43

Decorrelated feature 44

Decorrelated feature 45

Decorrelated feature 46

Decorrelated feature 47

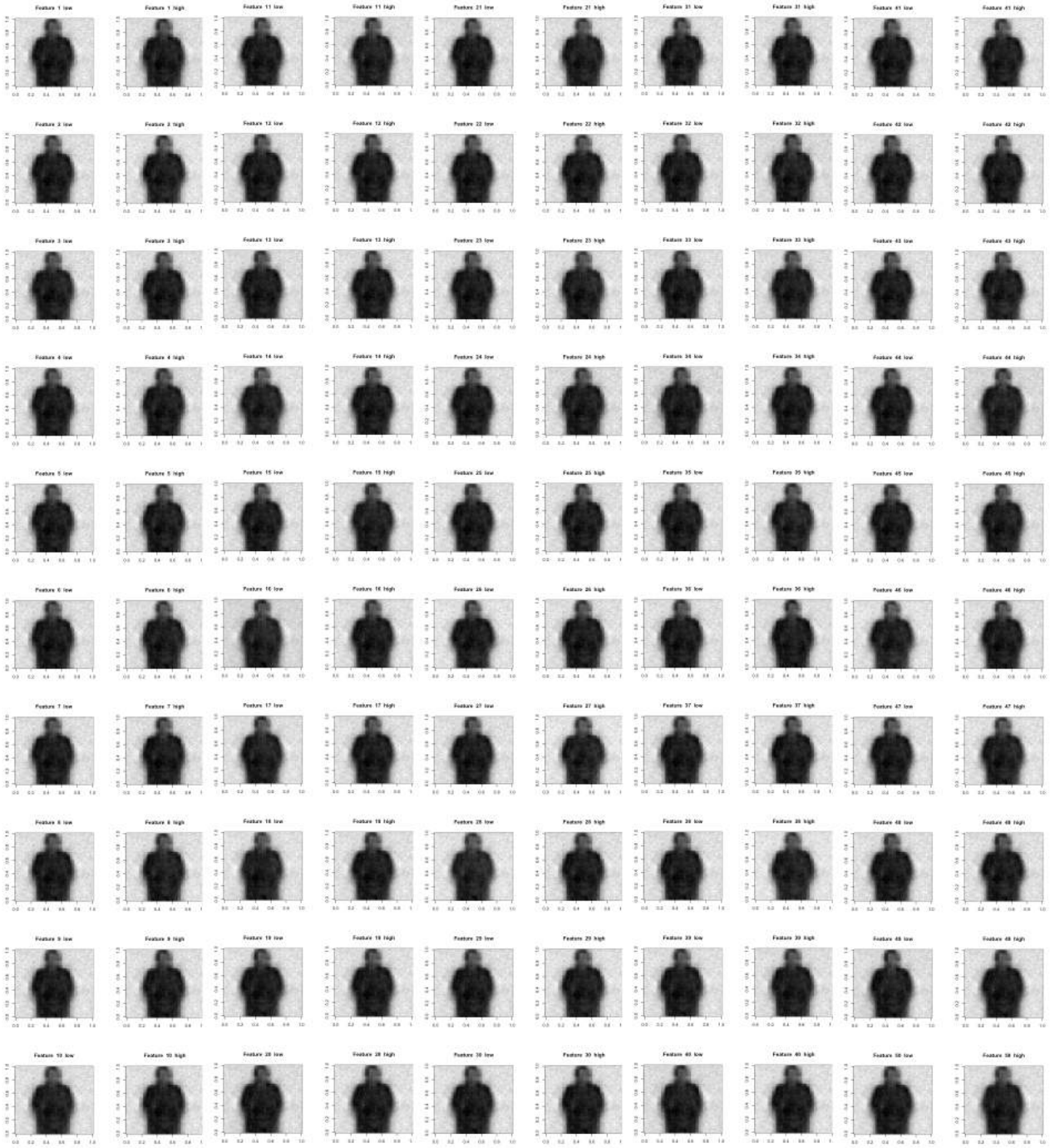
Decorrelated feature 48

Decorrelated feature 49

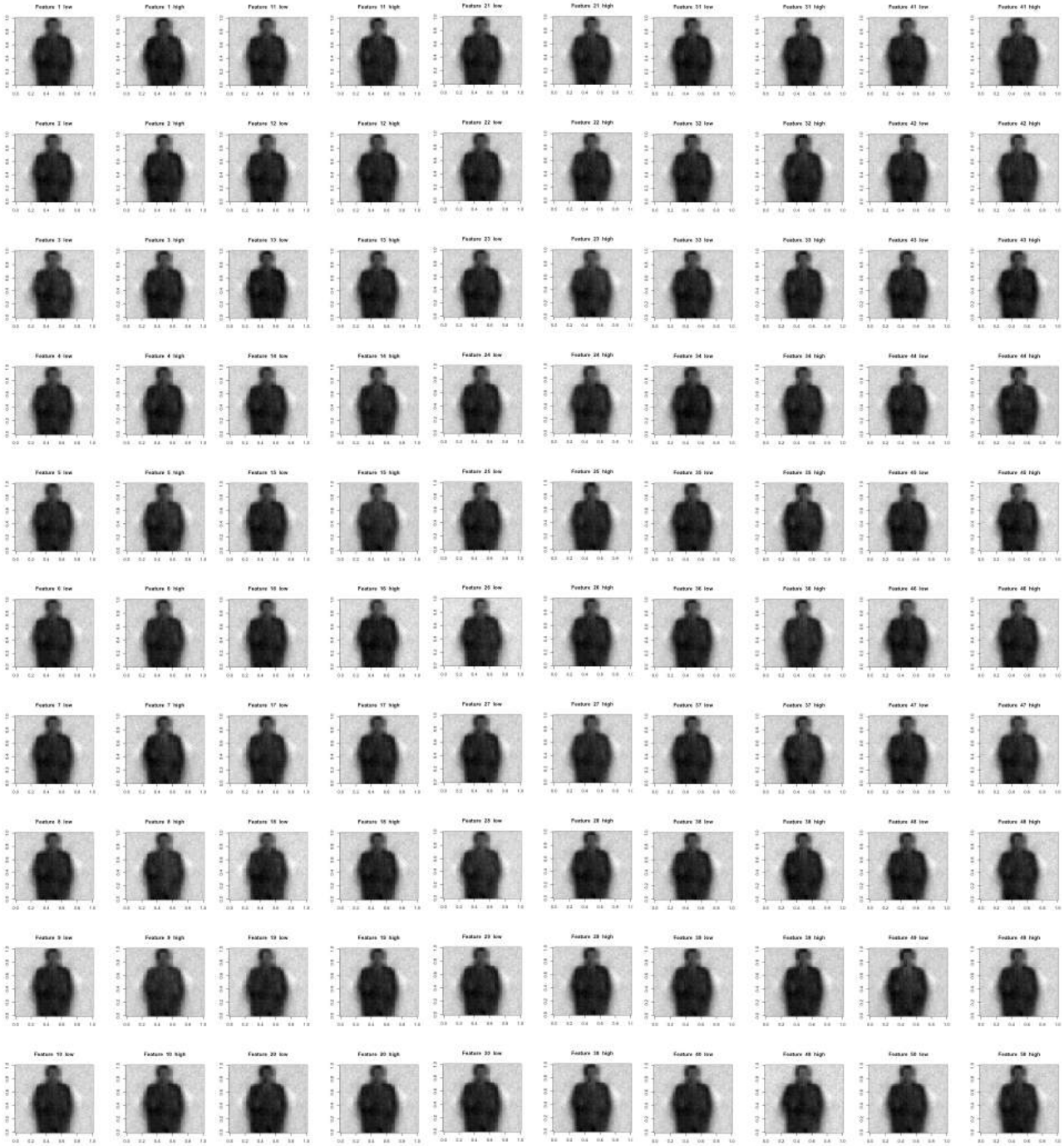
Decorrelated feature 50

Appendix 5

Single feature variations in feature space



Single feature variations in decorrelated feature space

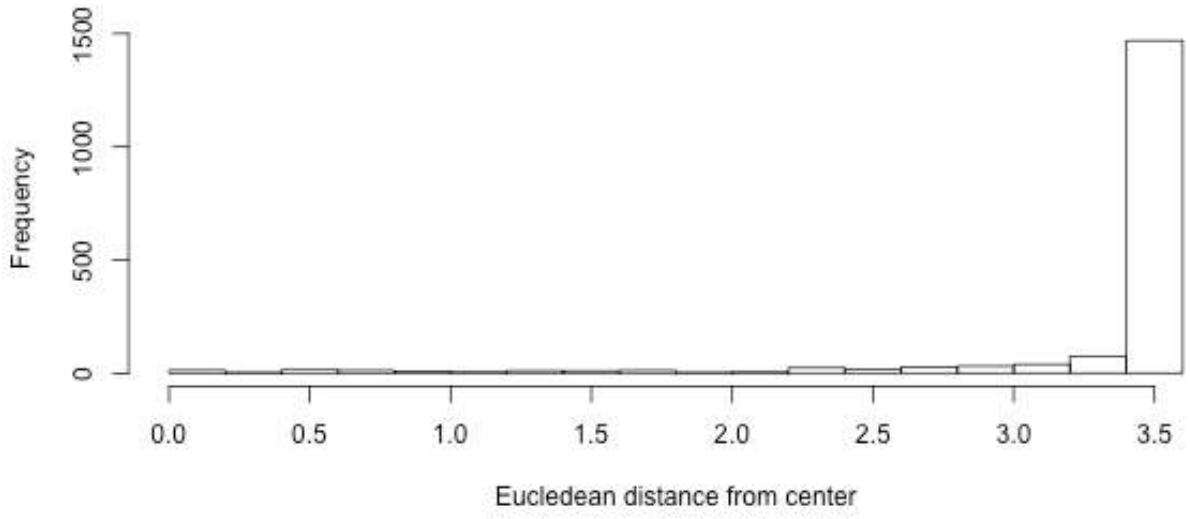


Appendix 6
All features low except a single feature

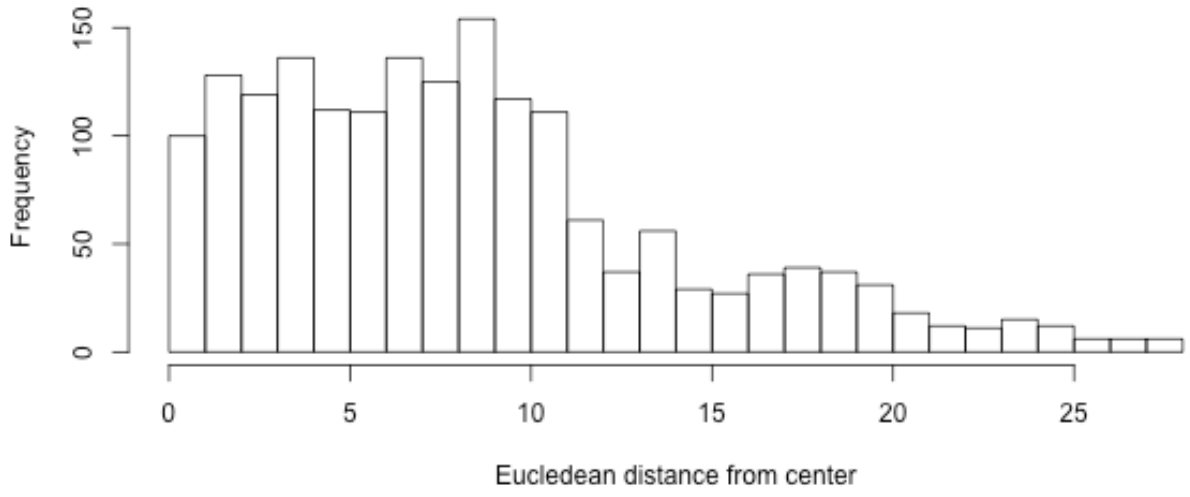
All decorrelated features low except a single feature

Appendix 7

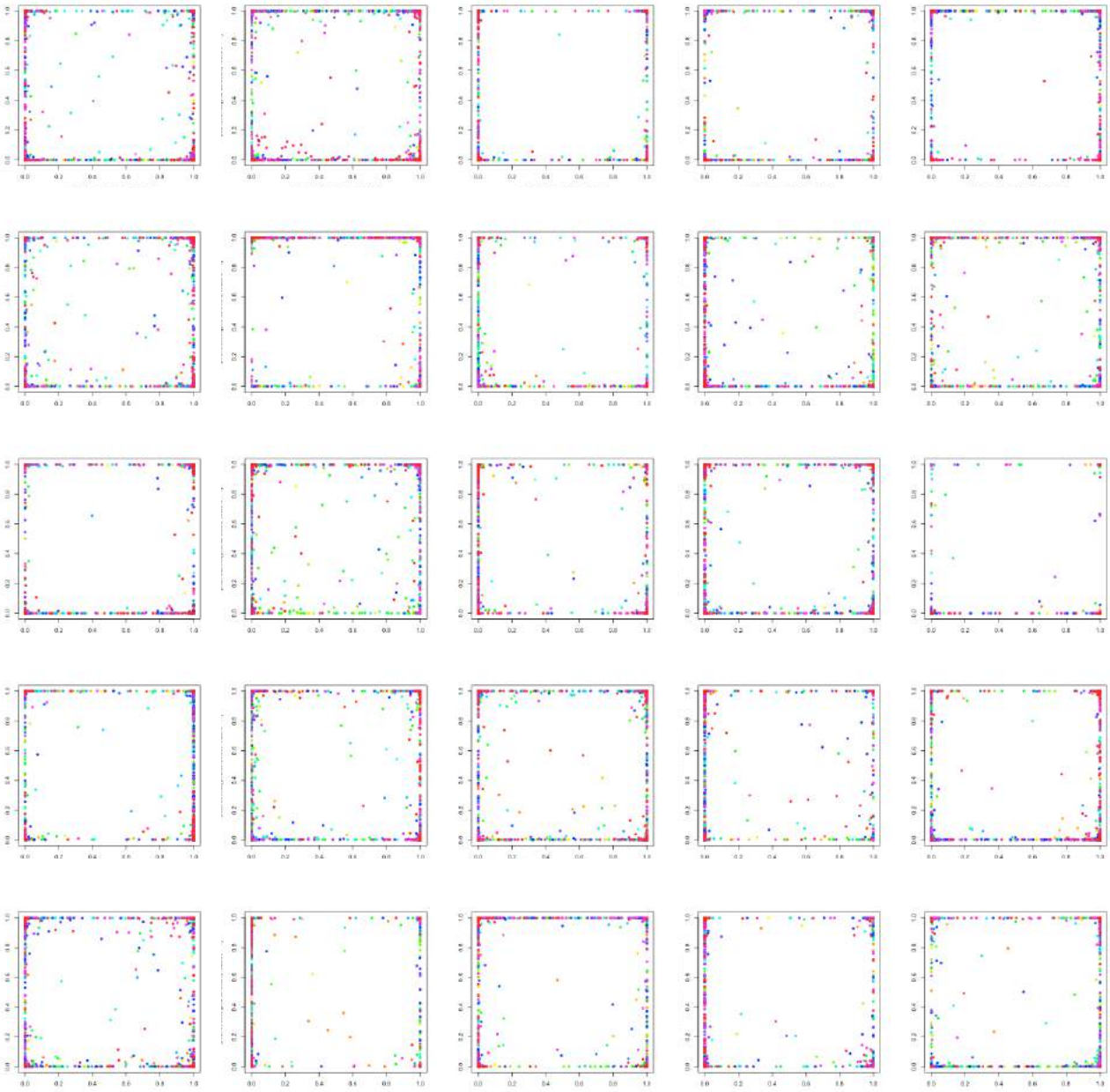
Histogram of feature activations



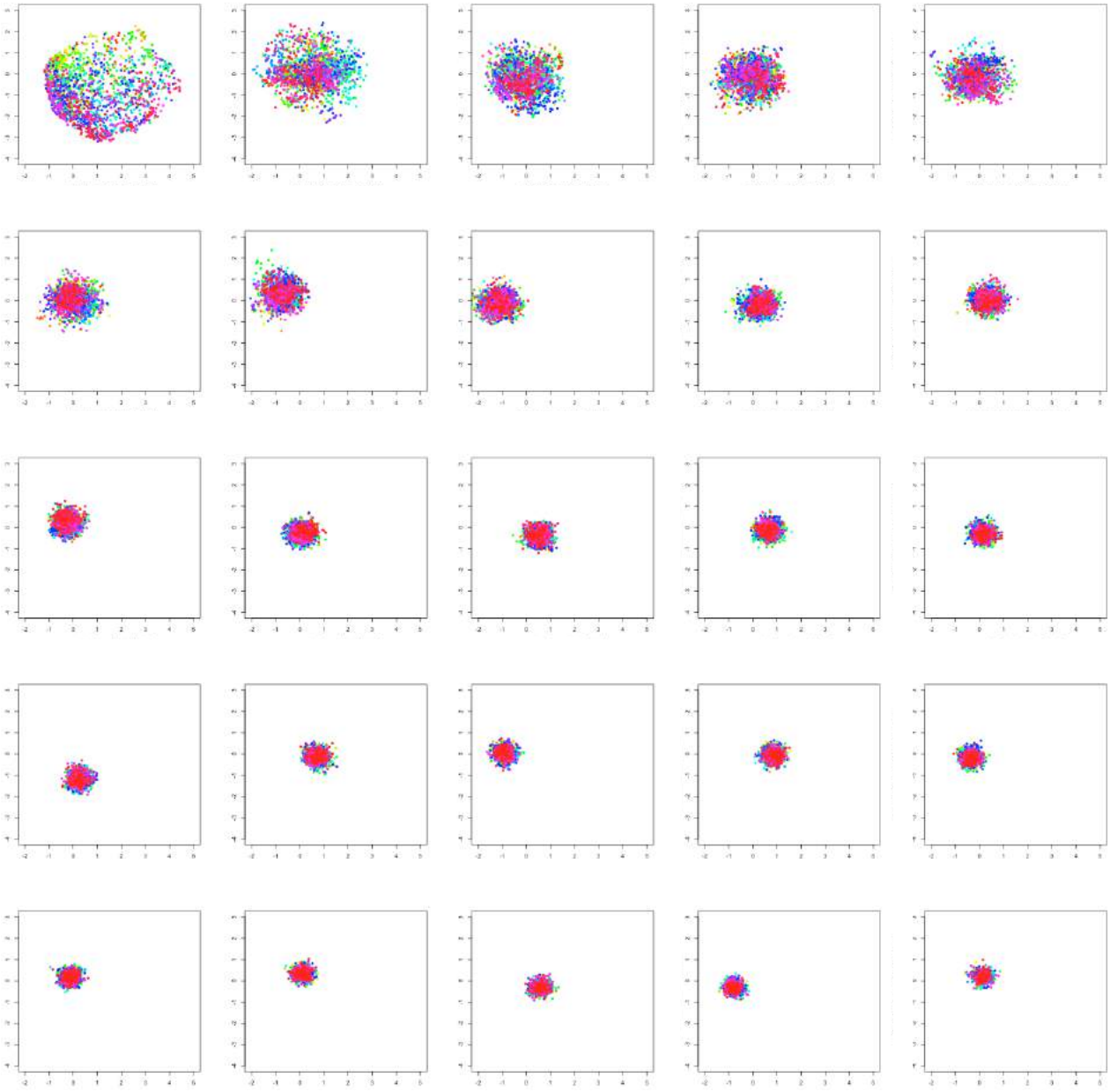
Histogram of decorrelated feature activations



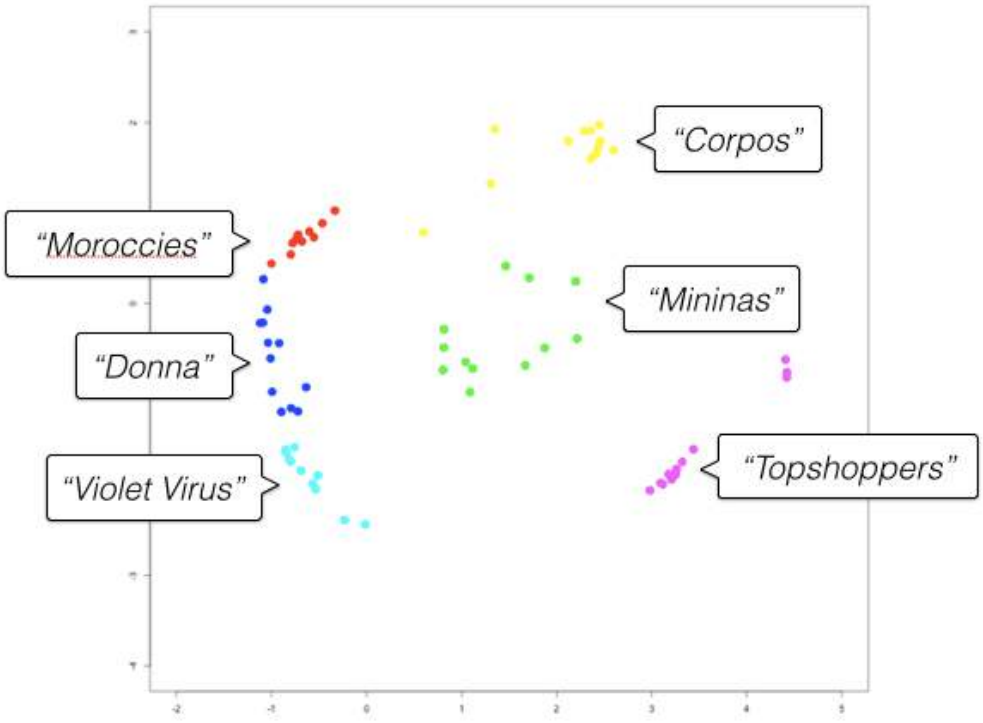
Appendix 8
Feature pairs



Decorrelated feature pairs



Subset of decorrelated features



Subset of decorrelated features with photos