Assessing Systematic Distortions in Visuospatial Mental Representations with use of Non-Linear Dimensionality Reduction.

An explorative study

Lise Stork
Graduation Thesis
Media Technology MSc program, Leiden University
Thesis advisors: Maarten Lamers, University of Leiden
and Thomas Barkowsky, University of Bremen

August 2016

Abstract

The following research focusses on an explorative analysis of cognitive maps that are externalised via sketch maps. The information that is encoded in sketch maps possibly includes spatial information that is stored in memory. Participants were asked to walk through a fictional virtual environment after which they were required to draw a map of the route. To analyse these maps and possible distortions in spatial mental representations that are shared among participants, we explore a geometrical analysis of sketch maps. Assuming that, in pixel space, the sketch map data have intrinsic dimensionality, we will use machine learning algorithms to reduce the dimensionality non-linearly so that we are able to inspect and interpret the possible manifold underlying these sketch map data. By doing so we hope we will be able to retrieve more knowledge about the processing and storage of spatial knowledge and to possibly provide a tool for assessing in what way spatial information is stored in memory.

Math symbols:

$\mathbb{R}^n$ indicator for n-dimensional space of real numbers
M capital letters indicate subspaces
$\boldsymbol{A}$ bold capital letters indicate matrices
x indicates a scalar
$\boldsymbol{x}$ bold letters indicate vectors
$\hat{\boldsymbol{x}}$ a reconstruction of vector $\boldsymbol{x}$
$\tilde{\boldsymbol{x}}$ a stochastically corrupted version of the vector $\boldsymbol{x}$
$\overline{\boldsymbol{x}}$ the mean vector of all $\boldsymbol{x_i}$

Contents

1 Introduction **3**

2 Visuospatial mental representations **4**
2.1 Cognitive mapping . . . 5
2.2 Distortions in cognitive maps . . . 5
2.3 Assessing cognitive maps . . . 5

3 Dimensionality reduction **7**
3.1 Principal Component Analysis (PCA) . . . 8
3.2 Locally Linear Embeddings (LLE) . . . 8
3.3 Stacked Denoising Autoencoders (SDAE) . . . 9

4 Method **11**
4.1 Material . . . 13
4.2 Participants . . . 14
4.3 Sketch map data . . . 14

5 Experimental validation **15**
5.1 Dimensionality estimation . . . 16
5.2 Principal Component Analysis . . . 16
5.3 Locally Linear Embeddings . . . 18
5.4 Stacked denoising autoencoders . . . 19
5.4.1 Experimental setup - subset 1 . . . 19
5.4.2 Results - subset 1 . . . 20
5.4.3 Experimental setup - subset 2 . . . 22
5.4.4 Results - subset 2 . . . 22

6 Experiments **24**
6.1 Dimensionality estimation . . . 24
6.2 Principal component analysis . . . 24
6.3 Locally Linear Embeddings . . . 26
6.4 Stacked denoising autoencoders . . . 27
6.4.1 Experimental setup . . . 27
6.4.2 Results . . . 28

7 Conclusion **30**

References **31**

Appendix **34**

1 Introduction

When we study the human mind, most data that is collected is influenced by perception, judgement, memory and other cognitive faculties of the human mind. With mental imagery research we are for instance interested in encodings of visual information in the brain and thus, as we are not yet capable of extracting such knowledge directly, we need to trust participants to articulate the information as accurately as possible and subsequently study these products of the human mind. This proves problematic for the discovery of new knowledge in psychology research as validity of results can often be questioned. Additionally, it can be challenging to design an experiment that collects the aforementioned data without the data and eventually the results of the experiment being influenced by what are called demand characteristics [Rosnow, 2002] such as the bias of the researcher and the unconscious willingness of participants to help researchers.

In an effort to deal with these issues the following paper plays with the idea of using pattern recognition algorithms to search for meaningful structures in sketches; concepts and information originating from the mind, externalised to drawings. Specifically, it explores the effectiveness of a geometrical analysis of sketch maps[1] to gain insights into the subject of mental processing of environmental information.

Although it is mentioned often that sketch maps prove useful to communicate spatial knowledge [Wang and Schwering, 2015], [Tversky, 2002b], little work has been done so far to explore the spatial information contained in sketch maps. In a study it was found that sketch mapping proves a reliable data collection method for retrieving insights into environmental knowledge processing [Blades, 1990] [Wang and Schwering, 2015]. Participants in the study seemed highly consistent in externalising the spatial information of a specific environment that they had stored in their memory via drawings of that environment, even from different perspectives. These findings suggest that in these sketch maps, some spatial information was encoded that participants had acquired earlier from perceiving that environment. Some studies also discuss sketching in general as a cognitive tool to augment memory and information processing [Tversky, 2002b]. A free mapping procedure generating these sketch maps is moreover desirable to obtain data as objectively as possible [Waterman and Gordon, 1984], without priming the participants too much by use of questionnaires or other methodologies to think into the direction of the hypotheses.

However, although sketch maps provide a reliable source for gathering knowledge about spatial cognition, it seems difficult to accurately extract knowledge from these drawings [Sanders and Porter, 1974] [Tobler, 1976] [Billinghurst and Weghorst, 1995]. One difficulty can be the manual analysis of these sketches. For such an analysis a team of researchers is required to discuss together what features to select and how to measure and evaluate the accuracy of these features. Instead of manual feature selection, we propose to outsource this task to *dimensionality reduction* (DR) methods to *learn* the data's most important features: their correlations and biggest variations. Such mathematical tools do not contribute meaning to the sketch maps, but see these maps as a pixel landscape to discover: they analyse these sketch maps in what is called *pixel space*[2].

We assume that, in pixel space, these sketch map data will have an intrinsic dimensionality much lower than their original input space. Thus, we assume that the sketch map data will reside on a lower dimensional manifold[3] in high dimensional space. Reducing the dimensionality of the data non-linearly to approximate the topology of the possible manifold will aid us in the feature selection process by providing us with a new coordinate space of meaningful features and, by inspection of the mapping to a new coordinate system, help us to interpret the topology of the possible manifold underlying these sketch map data. In general, the proposed method aims to detect the most important variations and commonalities between all sketch maps that can be compared to the real environment

[1] Maps drawn from observation rather than measurement and representing the main features of an area.

[2] In pixel space, every pixel is a feature and the amount of pixels is equal to the amount of axes in this feature space. The dimensionality is thus equal to the amount of pixels in an image: an image of size $m \times n$ has dimensionality $m \times n$

[3] A manifold is a space that locally appears to be a Euclidean plane, but which may have a different non-linear global structure

to see what tendencies people have when remembering an environment. By doing this we hope to retrieve more knowledge about mechanisms that are at work when processing, storing and retrieving spatial knowledge and to possibly provide a tool for assessing spatial information that is stored in memory.

In the coming sections we will explore the analysis of sketch maps by use of DR. If possible an attempt is made to assess some systematic errors regarding metrics such as relative distances. In section **2** the field of mental representation research is introduced and cognitive mapping, distortions in cognitive maps and methods used to analyse sketch maps are discussed. In section **3** we will discuss how DR could aid the analysis of sketch maps and we will subsequently mention the different DR methods used in this paper. The methodology that was used to gather sketch map data is explained in section **4**. Besides experimenting with the gathered sketch map data we have also tested the DR methods on a different image dataset in section **5** for a validation of our experiments, after which we have experimented with our own sketch map dataset in section **6**. Finally, conclusions about the results and possible improvements are discussed in section **7**.

2 Visuospatial mental representations

There has been much interest in mental imagery over the years as it seems to play a central role in many cognitive processes and provides mnemonic properties [Neisser and Kerr, 1973], but also as it is difficult to understand the nature of mental images. Imagine trying to remember the route to a place you have not visited for some time. Quite likely you are trying to remember what the streets looked like that brought you there, maybe you reason in a backwards manner, starting from your destination to mentally regress down the path you see before you, until you come across a street that links to your position in the real world. These sort of memories that you experience as if you are actually seeing images of the world appear to be different from other memories that are for instance stored sequentially. They give the impression of being real images, stored in the brain, which you can recall to look at at any given moment. We can not deny that, phenomenologically speaking, these memories are images. They give the experience of looking at real objects, real environments and real situations in what is mostly called the 'mind's eye'[4]. Thus, it is no surprise that the terminology relating to these mental processes consists of words related to the visual system. Much debate however has unfolded about the real nature of these mental images.

Stephen Kosslyn, a key researcher in the field, researched the structure of these mental representations, calling them mental images. His research compares mental images to real images, and argues that these mental images can be viewed and scanned in a similar manner [Kosslyn, 1973], that they are pictorial, and laid out somewhere in the brain in what he calls a *functional space* [Kosslyn, 1994]. Pylyshyn however rightly points out that the properties that are observed which would prove the existence of imagery in the mind are properties that are intrinsic to the structure of mental representation, the architecture of the mind, and reasoning in general [Pylyshyn, 2002], and that this idea of mental representations being images laid out in some space in the brain is perhaps not so self-evident after all. As a thought experiment, try to capture a visual mental representation and attempt to inspects its details, can you fix your mind's eye on some part of a picture? Do these thoughts really appear to you in the same way as images that you perceive via the visual system?

What we can agree upon is the idea that visuospatial mental representations are incomplete, distorted, simplified and idiosyncratic [Kaplan, 1973]. Thus, these representations are abstractions of the real world. What seems most interesting to research is how these mental representations are constructed from perception and recalled from memory. What information do we pick up from our environment, which features function as the constituents of mental representations, for example in constructing a model of the environment in our mind?

[4]The Mind's eye refers to one's ability to "see" things with the mind.

A goal in the study of mental imagery is to describe the functional similarities and differences between mental images and the perceptual events from which they were formed. Researching the mechanisms that store and create mental representations constructed from environments will provide us with more insight in spatial decision making. In this study we would like to research what spatial information humans remember when moving through an environment, what metric knowledge they use to construct internal representations and how this metric knowledge is possibly distorted by intentions in the environment and schematisation of the gathered knowledge.

2.1 Cognitive mapping

To be able to effectively walk around in and reason about three dimensional space, we construct representations of that space in our mind. Mental representations of space (e.g. an environment), or often called 'cognitive maps' [Kitchin, 1994].

What cognitive maps essentially consist of are a series of knowledge structures that hierarchically differ in level of detail and are formed for specific tasks [Golledge et al., 1985]. This means that cognitive maps are not actual maps that are stored in memory, but rather dynamical constructs that arise from cognitive processes interacting with memory structures. We retrieve this knowledge in different ways by for example seeing, touching, hearing and reading. The elements that we select from the environment and the spatial relations that we use to base these representations on are dependent on the space itself and the purpose that the representation serves [Meneghetti et al., 2015].

Barbara Tversky calls this view a constructionist view: we gather separate pieces of knowledge from an environment that we later use to reconstruct mental representations of new or encountered environments with [Tversky, 1993]. She also poses a new term for these mental constructs that better represents their natural structure: cognitive collages. This thesis will hold the same assumptions as Tversky and as the following theses by Thomas Barkowsky [Barkowsky, 2002] about human geographical knowledge processing: (1) geographic knowledge representations in the human mind are constructed on demand, (2) this construction is based on knowledge from long-term memory, (3) this knowledge is stored in a fragmentary, hierarchically structured form, and (4) the resulting representation in working memory is a visual mental image.

2.2 Distortions in cognitive maps

Researchers show that cognitive maps or cognitive collages contain systematic errors concerning relative distances and shapes of routes or environments [Stevens and Coupe, 1978] [Waterman and Gordon, 1984] [Tversky, 1992] . Distortions that arise are amongst others over- or underestimating distances, simplifying shapes and rotating locations relative to each other. A possible reason these distortions exist in mental representations is due to *schematisation* of the geographic and environmental information of which these representations are constructed [Tversky, 2002a]. This schematisation could arise due to many reasons, one of which being the limited capacity of working memory: schematisation of spatial information reduces the working memory load.

Furthermore, some research done in the field of cognitive mapping shows that knowledge structures seem to adapt to intentions with and in the environment such as navigating to work, home or other locations [Tversky, 2002a] [Lynch, 1960]. Thus, this schematisation of spatial information could also be an effect of the purposes one has in an environment.

2.3 Assessing cognitive maps

One way to retrieve access to some of the spatial information that is stored in memory is by use of a *direct mapping* or *sketch mapping* exercise which produces sketch maps such as the one in figure 1.

Figure 1: Sketch map example [Wang and Schwering, 2015]

This method allows for minimal involvement of the researcher by allowing the participant to draw a map, preferably without any restrictions, of how they remember a specific environment. As mentioned priorly, these sketch maps are thought to contain spatial information that is encoded in memory [Blades, 1990] [Wang and Schwering, 2015].

Underneath we will provide the reader with an overview of some example methods that are often used to analyse sketch maps.

- 1995: One study immersed participants in virtual worlds [Billinghurst and Weghorst, 1995] after which they were asked to produce a sketch map of the virtual environment. Participants were also required, posteriorly, to answer questions about their understanding of the virtual world; if they knew where everything was and if they could orient themselves in the virtual world. A goodness score was developed by the researchers that measured how accurately the sketch map representation resembled the virtual environment and a score was developed that expressed their ability to orient in the virtual world. The researchers hypothesised that in case the sketch maps were accurate external representations, this would correspond to the subject's survey scores for orientation in the virtual world. To obtain results these two measures were compared to see if they were indeed correlated.

- 1999: In research done by Tversky and Lee [Tversky and Lee, 1999], route maps were gathered by asking bypassers near a campus dorm to show the researchers the way to a specific restaurant. These route maps, both lists of street names and sketches of the route, were analysed through observation by the two researchers. The researchers hypothesised from these sketch maps that participants used similar structures when conveying the route directions as they noticed similar semantic content and a similar use of schematisation when eyeballing all route descriptions.

- A study done in 2009 [Wang and Schwering, 2015] collected sketch maps from participants by asking them to draw a map of an environment they were familiar with. To analyse the sketches they created evaluation metrics such as 'orientation of landmarks with respect to the street segment', 'orientation of street segment' and 'topology of street segment'. To prove the accuracy and reliability of these metrics, they evaluated them by use of a human study.

Many studies that analyse sketch maps thus rely on a subjective evaluation of metrics, created to evaluate sketch maps, based on observation and measurement.

3 Dimensionality reduction

In the following research we deviate from traditional methods in psychology research by exploring the effectiveness of using pattern recognition for the analysis of sketch maps. A method that allows for an objective analysis as we purely view the sketch map images as data and not as visual references to worldly concepts. This way the prior assumptions about our data are kept minimal.

What we do assume is that when many participants walk through the same environment, some of the environmental information they store in memory will be similar. Thus, we assume that when a group of participants are asked to externalise these mental constructs via sketch maps, these sketch maps will share some features. A possibility could be that all participants overestimate the horizontal distances and underestimate the vertical distances. The underlying shared structure could however be more complex and hard to extract from eyeballing these sketch maps only.

When we represent the sketch map images in pixel space or vector space where the dimensions correspond to pixels' grey values, we can analyse the features of these sketches geometrically by analysing the topology of the space they cover in high dimensional space. When images have similar feature values they will likely be neighbours in geometrical space. Thus, although the dimensionality of an image is often very high (an image of 64×64 pixels is a point in 4096-dimensional space, or a vector in $\mathbb{R}^{4096}$), sketch images in $\mathbb{R}^d$ all representing the same environment will most likely be positioned close together in a much lower dimensional space on or near what is called a non-linear manifold residing in the high dimensional euclidean space. This means they will lie entirely in an M-dimensional subspace of $\mathbb{R}^d$, M being smaller than d. The real dimensionality of this data is called its *intrinsic dimensionality*, its essential or natural dimensionality [Lee and Verleysen, 2007].

DR focusses on the visualisation and extraction of this natural dimensionality. A lot of different techniques exist that can reduce the dimensionality of a dataset to fewer dimensions while preserving much of its intrinsic structure. One technique to estimate, visualise and analyse the topology of the data *linearly* is called the Principal Component Analysis (PCA) [Abdi and Williams, 2010], of which we will discuss the essentials below. As data however often reside on or near a manifold with a non-linear structure, and not on a line, plane, cube or hypercube, the dimensionality of the data can also be reduced using a *non-linear* technique to provide more interesting insights about the data. The non-linear techniques that will be used in this exploratory study to learn the underlying structure of the possibly embedded manifold are *locally linear embeddings* (LLE) and a type of neural network: the *stacked denoising autoencoder* (SDAE). We will use these methods to reduce the dimensionality of the sketch map data for interpretation of their underlying structure.

It is nice to use PCA and SDAE on image data as these methods in particular provide a visual way to test the quality of the mapping to a lower dimensional space. That is to say, it is possible to reconstruct the images after compression by these DR methods. An encoding function is generated that maps each datapoint from high dimensional space to its low dimensional representative, and a decoding function that reconstructs each low dimensional representative back to its original dimensionality. With image data it is possible to visualise characteristics of the encoding and decoding functions and to create new images with these decoding functions to visually show the degrees of freedom of the possible manifold.

Other methods such as the LLE algorithm transform the data to a lower dimensional space while trying to maintain the topology of the manifold. Although this is also a minimisation problem, no encoding function is constructed; an effort is made to maintain the distances between the data samples as accurately as possible.

In the coming section we provide a short explanation of the pattern recognition methods used to find useful representations in our data. Firstly, we describe a widely used method for *linear dimensionality reduction* (LDR) that we mentioned priorly: PCA. After, we explain both *non-linear dimensionality reduction* (NLDR) techniques discussed earlier in this section: we will start by discussing the LLE algorithm and finish by introducing the workings of the SDAE.

3.1 Principal Component Analysis (PCA)

PCA is a LDR technique that relies heavily on linear algebra, especially on eigenvector and eigenvalue decomposition. It aims to find a linear subspace S on which the d-dimensional image data can be projected with minimal information loss. It first searches for a basis $B = \{ \boldsymbol{v_1}, \boldsymbol{v_2}, \ldots \boldsymbol{v_n} \}$ that spans the subspace in which most data points fall. The eigenvectors $\{ \boldsymbol{e_1}, \boldsymbol{e_2}, \ldots \boldsymbol{e_n} \}$ of a dataset serve as a good set of basis vectors, as they are orthogonal to one another. The eigenvectors can be seen as the principal components of the data as they span the directions of greatest variance, see figure 2. In the case of this ellipsoid in $\mathbb{R}^3$, reducing the dimensionality to the first two components would lead to a mapping of the data onto the first and the second direction of greatest variance and thus projecting the data onto an ellipse in $\mathbb{R}^2$.

The eigenvalues $\{ \lambda_1, \lambda_2, \ldots \lambda_n \}$ corresponding to the eigenvectors indicate the magnitude of the variance in the direction that each eigenvector spans. The eigenvector of the direction with the highest variance, the first principal component, will thus correspond to the highest eigenvalue.

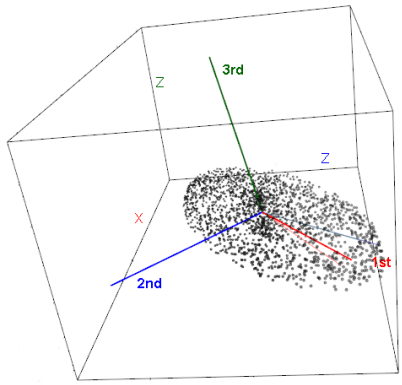


Figure 2: Principal component analysis on ellipsoid data in $\mathbb{R}^3$: the first, second and third principal component indicate the directions of greatest variance in the data.

By retrieving the eigenvectors and eigenvalues of an image, we can figure out which dimensions contain the highest variance and thus the most data, and thus which dimensions are redundant. Eigenvalues of 0 indicate that the dimension that is spanned by the corresponding eigenvector does not contain any data, low values indicate sparsity. The space that we are left with on which we can project all our data tells us something about our data. For images this mapping can be analysed very intuitively. The positions of the images in our dataset relative to our basis B tell us something about the variations between different images. For example, images that are furthest away from each other in this subspace are linearly least similar, while images that are closer together are linearly more resemblant.

3.2 Locally Linear Embeddings (LLE)

While PCA is a good method for LDR, other methods can reduce the dimensionality of the data *non-linearly*. An example of such a method is the LLE algorithm, an unsupervised learning algorithm that computes low-dimensional, neighbourhood-preserving embeddings of high-dimensional inputs [Roweis and Saul, 2000]. If we have a dataset with n vectors $\boldsymbol{x}_i$ which are sampled from an underlying manifold, the LLE algorithm can transform this embedding to lower dimensional space by calculating linear coefficients that can reconstruct every data point from its nearest neighbours, see figure 3. First, k neighbours are assigned to each datapoint by use of for example the k-nearest neighbours algorithm. k is thus one of the parameters to be optimised. Secondly, weights are computed that are best capable of reconstructing each data point from its k neighbours. Finally, a low-dimensional

embedding is computed that best reconstructs each datapoint from its neighbours with the priorly obtained weight vectors.

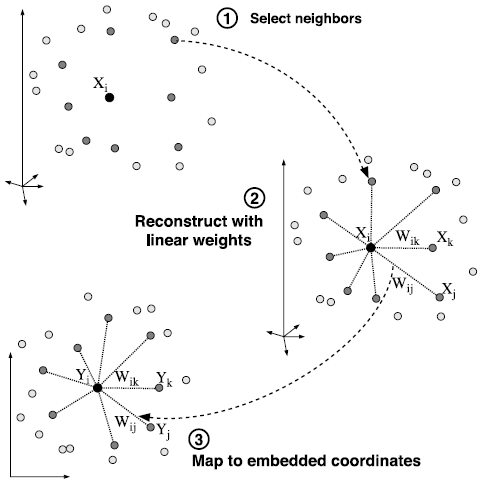


Figure 3: Schematic overview of the LLE algorithm [Roweis and Saul, 2000]

.

3.3 Stacked Denoising Autoencoders (SDAE)

Also deep artificial neural networks can be used for NLDR [DeMers et al., 1992] [Hinton and Salakhutdinov, 2006]. An example of a neural network that has this capability is the *autoencoder* (AE). Whereas a regular feed-forward neural network is trained by feeding it the data and showing it the corresponding targets, the AE is trained by reconstructing its input. This means that the AE consists of two parts: an encoder and a decoder part.

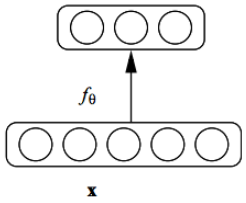


Figure 4: Schematic overview of an autoencoder [Vincent et al., 2010]

Figure 4 shows this model. The encoder encodes each vector $\boldsymbol{x}_i$ in the data from the input space, in this model $\mathbb{R}^5$, to a lower dimensional representation $\boldsymbol{y}$, $\boldsymbol{y} = s(\boldsymbol{W}\boldsymbol{x} + w_0)$, and the decoder decodes each vector back from their lower dimensional representation, here in $\mathbb{R}^3$, to their original input space, close enough to their original input, $\hat{\boldsymbol{x}} = s(\boldsymbol{W}'\boldsymbol{y} + w_0')$. The network is trained using gradient descent to change the weights in the weight matrix $\boldsymbol{W}$ such that the sum-of-squares error function is minimised, see equation 1. $\hat{\boldsymbol{x}}$ indicates the reconstruction of $\boldsymbol{x}$ after compression by the AE.

$$SSE = \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x} - \hat{\boldsymbol{x}})^2 \tag{1}$$

Often, a denoising criterion is added to the AE's that stochastically sets nodes from each input vector to zero or a random number. This corrupted version of the input vector is then compared to the uncorrupted input to calculate the loss, see figure 5. This criterion guides the learning of a good representation and often boosts the performance of an AE [Bengio et al., 2006] [Larochelle et al., 2009] [Vincent et al., 2010].

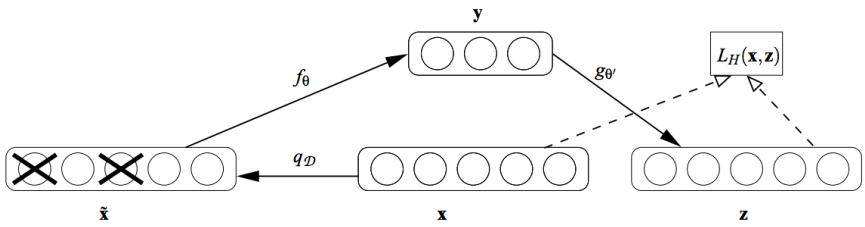


Figure 5: Schematic overview of a denoising AE. An example $\boldsymbol{x}$ is stochastically corrupted to $\tilde{\boldsymbol{x}}$ after which it is mapped to a lower dimensional representation $\boldsymbol{y}$. The decoder then attempts to reconstruct the vector $\boldsymbol{y}$ to its original input space $\mathbb{R}^5$ and produces reconstruction $\boldsymbol{z}$. With loss function $L(\boldsymbol{x}, \boldsymbol{z})$ the reconstruction is compared to the original input vector $\boldsymbol{x}$ and the loss is calculated [Vincent et al., 2010].

Multiple AE's can be stacked to form a deep architecture such as the one in figure 6. The AE's are trained separately before they are stacked together. The first AE is trained by feeding it vectors from the input space and the second AE is trained by feeding it the output generated by the first trained AE and so on. As each AE consists of an encoder and a decoder, we can train two encoders and two decoders by training two AE's. Stacked together, these AE's will form a 5 layered deep network, the SDAE. Sometimes, even deeper architectures are created as visible in figure 6. This final model is then fine-tuned using gradient descent. Training the whole model directly is often a bad idea as the model will converge only if the weights are already close to an optimum [Vincent et al., 2010].

It is possible to train the decoders similarly to how we train the encoders; by using gradient descent to minimise the SSE. It is however also possible to use tied weights as is visible in figure 6; in this case the transposes of the weights from the encoder are used to construct the decoder.

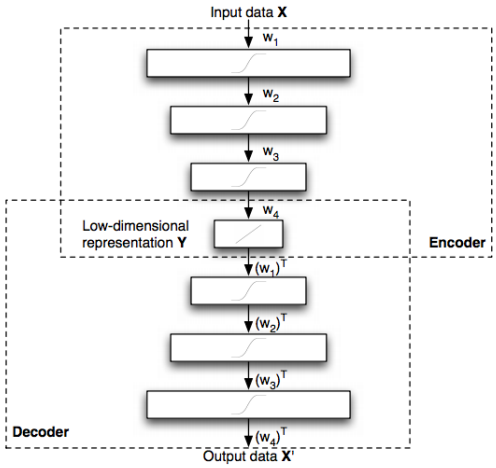


Figure 6: Schematic overview of stacked autoencoder [Maaten et al., 2009]

If the SDAE succeeds in finding a lower dimensional representation that has as many dimensions as the intrinsic dimensionality of the data, it means that the network of weights in the SDAE most likely created a good representation of the possible underlying manifold on which all data can be mapped with minimal information loss.

4 Method

To be able to analyse sketch maps using our proposed DR methods the sketch maps that were collected from participants had to meet certain requirements.

1. The sketch maps should all contain the same amount of information, meaning that all sketch maps should contain similar entities. If it is decided that landmarks can be drawn, *all* maps should contain the same landmarks.
2. The sketch maps should be externalisations of the same route. If one participant draws a route that follows a different path, we can not find commonalities between the routes as they are not referring to the same entity. Thus, although the routes can differ mutually, what they refer to should be equal.
3. The purpose in the environment of which the sketch maps were formed should be equal. As indicated before, intentionality influences perception and storage of environmental information.
4. Differences in sketch maps should not be an effect of the drawing skills of the participants. If drawings are very different as an effect of drawing skill alone, the algorithms will not be able to find good generalisations. To meet these requirements we have used one specific virtual environment, one route and one specific task to which we subjected all participants. There however were some options to do this and we have considered the following underneath.

Option 1: Participants receive a textual description of a three dimensional space which they are required to scan mentally, thus forming a mental representation of a fictional environment. This textual description could for instance be a dynamic textual map (such as the old text-based games, see figure 7), where participants receive textual cues to walk around in a fictional space.

Advantage: If a mental representation is formed, it is formed with knowledge structures from earlier encounters with events in the real world.

Disadvantages:

1. It is not certain that all participants will form a mental representation when they are merely subjected to textual descriptions in stead of a visual environment.
2. The variation in mental representations will be too large as they are constructed from earlier encounters with the real world which are personal.
3. Metric information is hard to transfer via textual information: the text has to be very explicit about this information as all routes should be similar, something that will affect and steer the storage of knowledge.

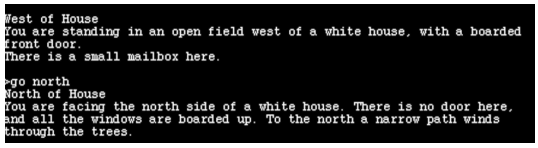


Figure 7: Example of a textual environment (a white house).

Option 2: Participants will be asked to interact with a visual virtual environment to form a mental representation, see figure 8.

Advantage: 'Walking' through a visual environment, although virtual, will enhance the construction of a mental representation of the environment. A research that compares spatial knowledge acquisition from navigation through real and virtual environments found that although participants were worse when learning a virtual environment compared to a real environment they could still show that their performance in a virtual environment was predictive for their performance in a real environment. They indicated that this suggests the working of similar mechanisms when learning a virtual environment and when learning a real environment [Richardson et al., 1999].

Disadvantage: As mentioned above, the contrast between cognitive mapping in the real world or a digital world affects learning of the digital environment.

Figure 8: Example of a visual (virtual) environment (a white house).

Both in option 1 and option 2, participants receive a task in that fictional world that requires spatial decision making, such as 'walk to the house at the other side of the city and retrieve a package'.

Option 3: Directly retrieving data from a sketch map database that is based on encounters with real environments. An example is *sketchmapia* [Institut für Geoinformatik, 2016], see figure 9.

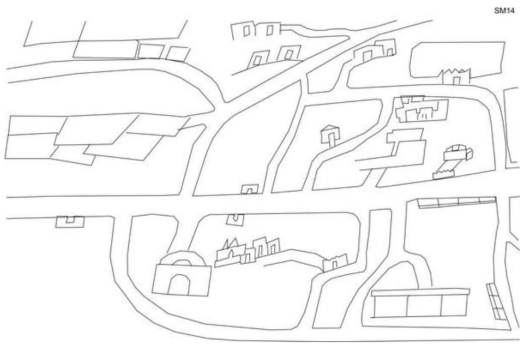


Figure 9: A sketch maps from a sketch map database sketchmapia [Institut für Geoinformatik, 2016].

Advantage: With sketch maps drawn of real environments, we are sure to capture some knowledge on environmental knowledge processing.

Disadvantages:

1. This database does not contain enough sketches per dataset, ±30 sketches per database. Further, no other sketch map databases could be found online. As the variation in the drawings is too big, a lot more data is required to find underlying patterns.

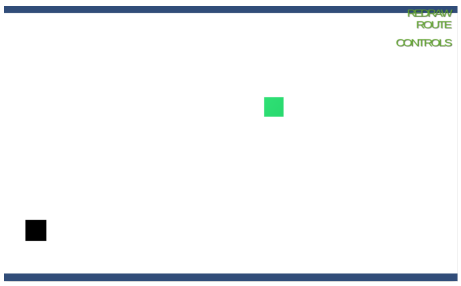
2. The data is too variant, also mentioned in number 1, and intentionality is not clear enough. In short, this data does not comply to the requirements mentioned at the beginning of this section.

4.1 Material

Eventually, a visual digital environment was constructed for participants to learn, see figure 10a. Participants were asked to walk through a fictional virtual city from first person view to retrieve a package that was located at the other end of the city. The road that lead to the package was predetermined, meaning that although participants could walk through the environment by themselves using keyboard arrows to navigate and view the scene around them, the road was blocked in several places such that participants could only walk one specific route from one end of the city to the other where the package was located.

The route was of a length that is hard to envision in the mind as a whole. In this way we hoped to avoid that participants were able to retrieve the entire route from short-term memory rather than having to reconstruct the route from spatial knowledge that they had stored in long-term memory.

(a) Full environment first person view



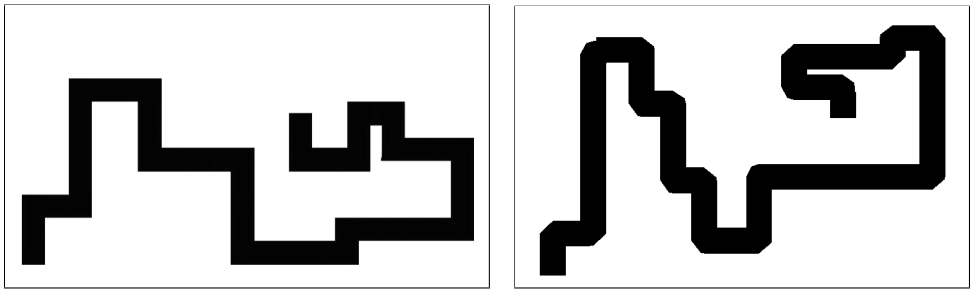
(b) Full environment top view

Figure 10: The fictional virtual environment

After being immersed in this virtual environment, participants were asked to externalise their cognitive maps by creating a sketch map of the fictional space from memory. To give participants drawing freedom but at the same time adhering to the requirements of the data, participants were subjected to a restricted drawing task. They were faced with a to top view of the environment but without any of the streets or landmarks, merely a white rectangle filling the screen, see figure 10b. Here they were asked to walk the exact same route they walked earlier from first person view, but this time they were expected to reconstruct the environment from memory, transforming the gained spatial knowledge to a two dimensional top view representation. The black dot in figure 10b was the indicator for the participants' starting positions and the green dot from figure 10b referred to the endpoint where the package was initially located in the 3 dimensional environment. While walking over this top view of the environment their pawns left a trace, forming a sketch that represented the route they constructed from memory, see figure 11b. Figure 11a represents the real route that participants walked from first person view. It is important to realise that the route commenced in the bottom left corner and ended in the middle to the right of the map.

You can see that the participant that drew figure 11b performed quite well as the reconstruction already is a good approximation of the real route. There are however errors; we can see that he

overestimated the length of the longest most left street, and that he approached the endpoint from the north in top view while from first person he actually approached the endpoint from the south.



(a) The real route (b) A sketch map

Figure 11: The sketch mapping exercise

4.2 Participants

The experiment was spread online such that as much participants as possible could be gathered. It is possible that due to the online gathering of data, causing us to have less control over the data that was gathered, some images polluted the dataset by reasons such as misinterpretation of the task at hand or other. In this study this is however not necessarily a bad thing as the methods could prove able to filter out outliers from the data. It is even interesting to see if the proposed methods can find such outliers easily when they do not adhere to the way the other data is structured. Before commencing with the experiment, participants had to fill in their gender, age and country of origin. Table 1 shows the information of all people that participated in the experiment. Most participants were between the age of 20 and 30, and a few were around the age of 50.

Country		**Gender**	
Argentina	1	Male	49
Aruba	1	Female	31
Australia	1		
Azerbaijan	1		
Brasil	1		
Britain	2		
Chile	1		
China	2		
Germany	2		
Finland	1		
Greece	1		
Luxembourg	1		
the Netherlands	61		
New Zealand	1		
Spain	1		
Unknown	2		

Table 1: Participants data

4.3 Sketch map data

Finally, a total of 80 sketch maps were collected from all participants of which in total 49 male and 31 female. The images that were collected from the sketch mapping tool were of size 46×80. For analysis the pixels of each image were concatenated from a matrix of size 46×80 to a vector of size

$46 \times 80 = 3680$ and thus each image was a vector in $\mathbb{R}^{3680}$. The dimensions of each image vector hold a value in the range [0, 255] which indicates the intensity of a pixel: 255 for a white pixel and 0 for a black pixel.

5 Experimental validation

Due to the sparsity of our sketch map dataset, some problems will likely arise when we use our DR methods to approximate the possible underlying manifold. The dimensionality d of the data is 3680 whereas n, our sample size, is merely 80. Usually, n is a multiple of d [Raudys and Jain, 1991]. Due to the small sample size, the used algorithms might not be able to find a good generalisation of the underlying pattern in our data. Thus, before experimenting with the gathered sketch map data, the methods were tested on a large and more familiar dataset to test the effectiveness of our approach and the interpretability of its results. And, by testing the methods on both a sparse and a dense subset of this large dataset, we can generalise about the effects of sparsity on the capabilities of the methods to find generalisable patterns.

The DR methods were tested using the *MNIST* dataset containing images of handwritten digits from 0 to 9 [LeCun et al., 1998]. The MNIST dataset contains many observations, 70.000 in total, and is thought to reside on an embedded manifold in multidimensional space. These qualities make it a good dataset to test and evaluate our methods and to compare the outcomes to the outcomes of the methods when we use them on our own sketch map dataset.

Only a subset of the MNIST dataset was used for this study as we are not interested in classification, but rather in finding common structures and variations in images that are of the same class and thus refer to the same concept. The subset that we selected existed of all the images that represent the handwritten digit 7. This gave us a dense training set consisting of 6000 images of size 28×28, with a dimensionality of $28 \times 28 = 784$, and a test set of 1000 images with the same characteristics as the training set. This test set was used to test whether the deep neural network could create a good generalisable model of the MNIST data by monitoring whether the model did not overfit to the training data. We will call these 7000 samples the *dense MNIST data*. Also a subset was selected from the training as well as the test set with an equal sample size as our own sketch map dataset, $n = 80$, to serve as sparse dataset. Accordingly, we will name this dataset the *sparse MNIST data*. In the MNIST dataset, as with our sketch map dataset, each dimension of each image vector holds a value in the range [0, 255] that indicates the intensity of a pixel. These values were normalised, as is common practice with dimensionality reduction on images, to fall in the range [0,1].

In the coming subsections we will first estimate the dimensionality of both the dense and the sparse MNIST data with use of the MLE algorithm [Levina and Bickel, 2004], an algorithm that estimates the intrinsic dimensionality of a dataset by treating all observations as a homogeneous Poisson process.

Subsequently we will experiment with PCA on the dense MNIST data. We will discuss what can be interpreted from the variation in this dataset and explore how and if we can visually interpret the eigenvectors that are extracted from this dataset. For PCA it is not necessary to test if a good model is constructed as this method guarantees a perfect linear mapping without loss of information to at least $n - 1$ dimensions even when the data is sparse.

Following, the LLE algorithm is used to create a mapping for both the dense and sparse MNIST data such that we can compare the mappings and perhaps venture to speculate about the effects of sparsity on the quality of the mapping.

Finally, the SDAE is trained on the training set of the dense MNIST data and the model it produces is validated on the test set of the dense MNIST data. If a model is constructed that is able to encode each input vector from the input space to a useful lower dimensional mapping as indicated by the MLE, we will try to interpret the meaning of each remaining dimension. After, the SDAE is trained on the sparse MNIST data of $n = 80$ and the mapping is compared to the mapping that is created by

training the SDAE on the dense MNIST data. If the SDAE is able to find comparable models when mapping highly dimensional dense data to lower dimensions and when mapping highly dimensional sparse data to lower dimensions, we can possibly generalise about the underlying structure of the sketch map data using merely a sparse dataset.

5.1 Dimensionality estimation

The MLE algorithm indicates that the intrinsic dimensionality of the dense MNIST data is 12. Thus, the MLE estimates that there are 12 degrees of freedom in our supposed manifold. When we estimate the intrinsic dimensionality of the sparse MNIST data, $n = 80$, the dimensionality estimator indicates that the intrinsic dimensionality is 8. This sample size dependency is something to consider when analysing the intrinsic dimensionality of the sparse sketch map data. It is however also logical as data will most likely contain less variation with less data points and more variability with a larger sample size.

5.2 Principal Component Analysis

PCA was conducted only on the dense MNIST data, as mentioned priorly. Looking at figure 12, we can see that the cumulative contribution of the eigenvectors $\boldsymbol{e_i}$ to the variance in the data increases rapidly, meaning that most of the data points approximate a linear combination of only part of the eigenvectors. The first 6 principal components already explain ±50% of the variation in the dataset, and ±90% of the variance is explained by the first 65 components.

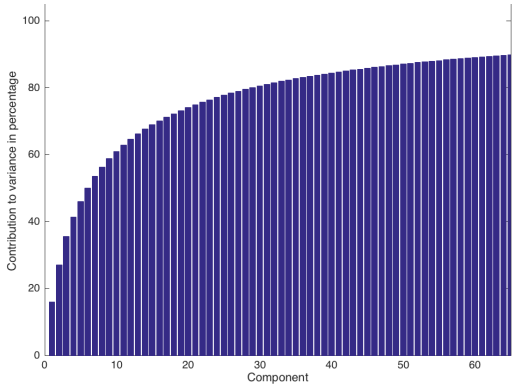


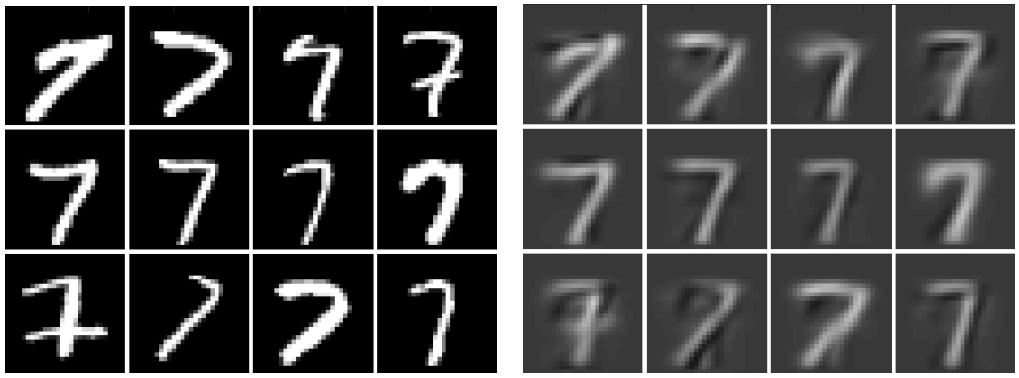
Figure 12: Scree plot of dense MNIST data: cumulative percentage the principal components contribute to the variance.

Figure 13 shows a reconstruction of a few samples from the dataset after compression of the data by use of the eigenvectors. To the left we can see a subset of the original dense MNIST data and to the right the same images after reconstruction from ±63% of the variance; a reconstruction from the first 12 eigenvectors. We choose a compression to 12 dimensions as this is the estimated intrinsic dimensionality of the dense MNIST data. The images are compressed with equation 2, and reconstructed with equation 3. Here, $\boldsymbol{A}$ indicates a matrix with the eigenvectors as row vectors, the amount of eigenvectors equal to the amount of dimensions to which the image is compressed, $\boldsymbol{y}$ indicates an image vector that is mapped to a lower dimensional representation and $\hat{\boldsymbol{x}}$ the image vector after reconstruction.

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} - \overline{\boldsymbol{x}}) \tag{2}$$

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{y}) + \overline{\boldsymbol{x}} \tag{3}$$

As we can see from figure 13, angle and width are mostly maintained while characteristics of different writing styles and other features are removed. The reconstructions are however surprisingly similar to the original images.

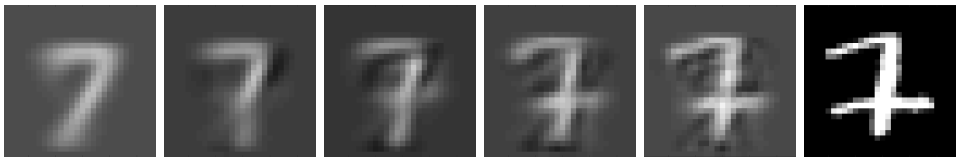


(a) Original images

(b) Reconstructed images after PCA compression to 12 components

Figure 13: Subset dense MNIST data and reconstruction after PCA compression

As the number of eigenvector with which each image is compressed and reconstructed is incremented, the accuracy of the reconstruction increases, see figure 14, until the number of eigenvectors k are equal to n-1 or d-1 which would generate a lossless reconstruction.



(a) $\boldsymbol{e}_1$ (b) $\boldsymbol{e}_{1:5}$ (c) $\boldsymbol{e}_{1:10}$ (d) $\boldsymbol{e}_{1:30}$ (e) $\boldsymbol{e}_{1:60}$ (f) $\boldsymbol{e}_{1:d-1}$

Figure 14: Reconstruction of the original bottom-left image of figure 13a after compression to 1, 5, 10, 30, 60 & d-1 eigenvectors.

To obtain a better understanding of the meaning of the principal components we can look at a selection of the eigenvectors, for instance the first six eigenvectors, see figure 15, with which the images in figure 13 are reconstructed. As these eigenvectors are of the same dimensionality as the original images, we can visualise them as images of the same size as the original images. Every image of a 7 in figure 13b is a linear combination of the first 12 eigenvectors each multiplied by a scalar and the output is then added to the mean vector, see equation 3. The eigenvectors themselves show the biggest variations in the dataset; they indicate a kind of general structure that all these digits share. As we know that black pixels indicate low values and white pixels indicate high values, it is possible by inspection of the eigenvector images to interpret the meaning of each principal component. Dimension 1 changes gradually from a broad 7 to a narrow seven with a sharper angle, while dimension 2 slowly changes between a narrow 7 with an 80°angle to a broader 7 with a 30°angle. If we would like to create a 7 that has a narrow top and an 80°angle we can create it using only the second eigenvector. So for example the equation $(\boldsymbol{e}_2^T x) + \overline{\boldsymbol{x}}$ will create this image. x in this equation indicates a lower dimensional representation of the image, in this case in $\mathbb{R}^1$.

(a) Mean (b) Eig 1 (c) Eig 2 (d) Eig 3 (e) Eig 4 (f) Eig 5 (g) Eig 6

Figure 15: First 6 eigenvectors and mean image visualised

This method nicely shows the most important linear variations in the dataset without any difficult calculations or without preprocessing the data too much by selecting specific features manually. We could use such a method to hypothesise about the way in which the concept of a number 7 is encoded in the minds of participants; what commonalities and variations we can extract from the handwritten digits. In this study we are however not interested in the concept of a number 7 but we will later use this method to hypothesise about environmental information processing in the minds of participants from commonalities and variations hidden in the sketch map dataset. It however remains difficult to draw conclusions from the mean vector and the eigenvectors as these are not robust under rotations and translations in the data. For this we require non-linear methods. Still, PCA can provide us some visual insights about some of the characteristics of our data.

5.3 Locally Linear Embeddings

Underneath we can see non-linear mappings of the MNIST data representing the digit 7 from $\mathbb{R}^{784}$ to $\mathbb{R}^2$ found by the LLE algorithm both for the dense MNIST data, visible in figure 16, as for the sparse MNIST data visible in figure 17.

Figure 16: Plot of the dense MNIST data mapped to 2 dimensions by LLE. $n = 6000$.

We have created these mappings in an effort to visualise the effects of sparsity on the LLE algorithm to find good generalisations. The mappings do appear comparable; the shapes of both data clouds are in accordance with one another and the gradual changes of the 7's over the manifold appear to capture the same structural information.

Apart from these comparisons we can conclude that the mapping which the LLE algorithm has found is a useful one as similar images appear to be close together. If it is more useful than the mapping found by PCA can not be concluded from these plots alone.

Figure 17: Plot of the sparse MNIST data mapped to 2 dimensions by LLE.

Later, when we will plot the sketch map data in $\mathbb{R}^2$ by use of the LLE algorithm, we can compare the result to the two dimensional plot created by use of PCA, to see whether either one provides us with useful information about the structure of the data cloud and the meanings of the each dimension.

5.4 Stacked denoising autoencoders

A five layered network was used to train the SDAE on the MNIST dataset. Two encoders were set up to encode the data to lower dimensions and two decoders to decode the data back to their original dimensionality. When stacked together, the bottleneck layer overlaps as is visible in figure 6 and a 5-layered network is formed. No tied weights were used to reconstruct the decoders; each encoder and decoder was trained using gradient descent.

As priorly mentioned in section 3.3, a denoising criterion called *masked noise* was used to stochastically set the dimensions from each input vector and each layer to 0 such that the network was better able to find generalisable patterns. As the hidden layers of each AE are under-complete, meaning that the dimensionality of the input vector is reduced instead of increased, the hidden layer of each AE is not capable of learning each exact input vector and thus it becomes harder for the network to overfit to the data. It is however still beneficial to use a denoising criterion to find useful features in the image data [Vincent et al., 2010].

5.4.1 Experimental setup - subset 1

By trial and error we have finally used a network with the following architecture: 784-100-12-100-784. Here, 784-100 indicates the AE that reduces the dimensionality from the input space to 100, hereby searching for features in the data. The second AE reduces the dimensionality from 100 features to 12, indicating the intrinsic dimensionality as estimated by the MLE. The size of the hidden layer of the first AE was inspired by a research by Palm [Palm, 2012] and Hinton [Hinton and Salakhutdinov, 2006].

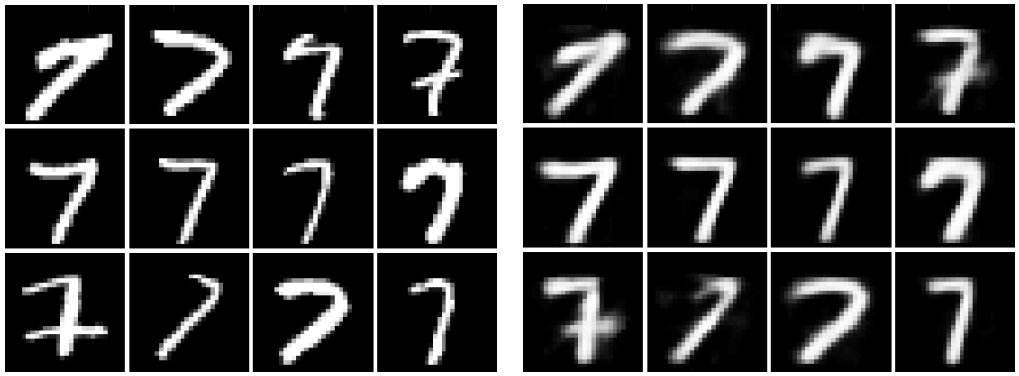
We have experimented with different values for the parameters to train each AE and the final model. Of all these values, all combinations were tried to find the parameter combination that performed the best on the dataset. Table 2 shows the final parameters that were used for training each DAE and the final model (the SDAE).

Parameters model	DAE 1	DAE 2	Final model
Learning rate	0.1	0.1	0.01
Momentum	0.5	0.5	0.5
L2 Regularisation	0	0	0
Dropout	0	0	0
Masked noise	0.5	0.5	0.5
Batchsize	10	10	1000
Epochs	50	50	2000

Table 2: SDAE different training parameters for training on the dense MNIST data

5.4.2 Results - subset 1

Viewing some of the test set images after construction from the compressed 12 dimensional data in figure 18 we can see that although the reconstruction is not perfect, the correct structure of each digit can be reconstructed. The model thus captures most of the intrinsic structure of the data. Here we can conclude that the model that the autoencoder has constructed is an approximation of the '7' manifold possibly underlying the data.



(a) Test set images

(b) Reconstructed test set images after SDAE compression to 12 dimensions

Figure 18: Subset dense MNIST data and reconstruction after SDAE compression

The model that is constructed can be interpreted as capturing the correlations in the data. The lower dimensional data, mapped by the autoencoder to lower dimensions, can be interpreted as decorrelated data. Each dimension represents different variations, translations and deformations. The decoder that is created by the autoencoder maps each vector in $\mathbb{R}^{12}$ to "7 space" as each vector in $\mathbb{R}^{12}$ that we feed into the decoder will turn into a vector in $\mathbb{R}^{784}$ that, if we would transform the output vector to an image, would resemble a handwritten digit 7. This means that our trained network can be interpreted as a structure that approaches the possibly embedded 7 manifold. We can inspect the 7 space found by the AE by inserting different vectors in $\mathbb{R}^{12}$ into the decoder. The behaviour of the sigmoid functions that are used to transform the activations of each neuron ensures that the activations of the middle bottleneck layer always fall into the range [0,1]. This means that each dimension in 7 space can only take on values between 0 and 1. The vector $\boldsymbol{x}$ underneath relative to our mapping M is thus in the centre of our 7 space.

$$\boldsymbol{x} = \begin{bmatrix} 0.5_1 \\ 0.5_2 \\ 0.5_3 \\ \vdots \\ 0.5_{12} \end{bmatrix} \qquad \boldsymbol{u} = \begin{bmatrix} 1_1 \\ 0.5_2 \\ 0.5_3 \\ \vdots \\ 0.5_{12} \end{bmatrix}$$

The centre vector $\boldsymbol{x}$ fed into the decoder produces a vector in $\mathbb{R}^{784}$ that is reconstructed to an image in figure 19a. All different vectors in $\mathbb{R}^{12}$ that are fed into the decoder will come out as different 7's with different features. If we would insert the vector $\boldsymbol{u}$ into the decoder, a vector which only differs from the centre vector by the first dimension, we would get an image that looks like the image produced by the centre vector but which varies on one dimension or with one feature.

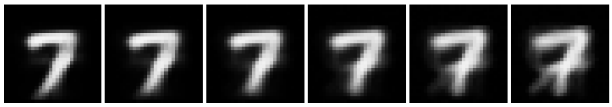
By entering different values for the first dimension in the range [0,1] and by comparing the output of the decoder to the output of the middle vector image we can somewhat visualise the meaning of the first dimension. See figure 19 for a selection of the dimensions visualised. Dimensions that appeared most meaningful were selected for visualisation. The values that were inserted in each dimension were 0, 0.2, 0.4, 0.6, 0.8 and 1, resulting in 6 images per dimension. Every time, all remaining dimensions were kept constant with the value 0.5.

(a) Centre

(b) Dim 2

(c) Dim 3

(d) Dim 7



(e) Dim 11

Figure 19: The SDAE model trained on the dense data: Visualisation of the centre vector relative to our mapping M and the dimensions that appear most meaningful: 2, 3, 7 & 11.

By visualising the degrees of freedom of the constructed model, we can gain a lot of information about the variations in our dataset. The information that appears encoded in each dimension varies between the thickness of a line, rotation of a digit, addition of a stripe, and similar 7's that have shorter or longer pedicels.

5.4.3 Experimental setup - subset 2

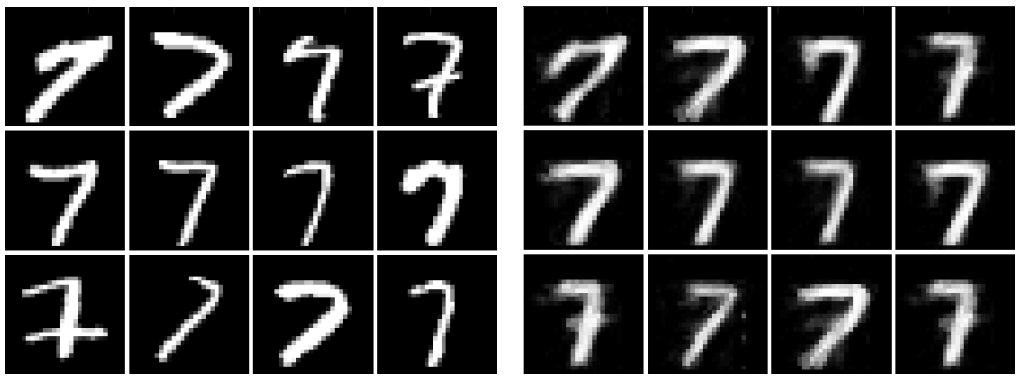
The sparse MNIST data was trained with a similar structure as the dense dataset, but this time the network architecture reduced the dimensionality of the data to 8, the intrinsic dimensionality as estimated by the MLE algorithm, obtaining the following architecture: 784-100-8-100-784. Although some parameters were not changed, other parameters had to be adjusted to retrieve reasonable reconstructions. The final parameters can be found underneath in table 3.

Final Parameters	DAE 1	DAE 2	Final model
Learning rate	0.005	0.01	0.1
Momentum	0.5	0.5	0.5
L2 Regularisation	0	0	0
Dropout	0	0	0
Masked noise	0.5	0.5	0.5
Batchsize	10	10	10
Epochs	500	200	400

Table 3: SDAE different training parameters for training on the sparse MNIST data

5.4.4 Results - subset 2

With a sparse subset of the possible embedded manifold we were not able, as suspected earlier, to found a good mapping to 8 dimensions. See figure 20 for the original data from the test set and a reconstruction from the SDAE after compression to 8 dimensions.



(a) Test set images

(b) Reconstructed test set images after SDAE compression to 8 dimensions

Figure 20: Subset sparse MNIST data and reconstruction after SDAE compression.

The model that is created for this subset appears to capture somewhat of the intrinsic structure of the data, but is much less accurate than the model that was created for subset 1, illustrated in figure 18. The SDAE appeared to train very well on the training set, but actually started to overfit to the train set very quickly as is visible from the error on the test set in figure 21. As for this, the training was stopped earlier to prevent the model overfitting to the training set.

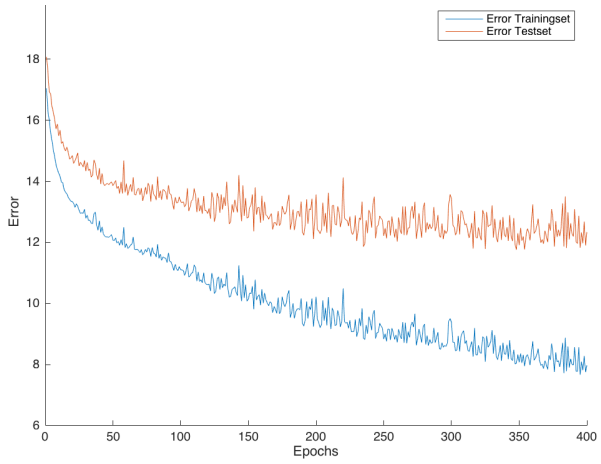
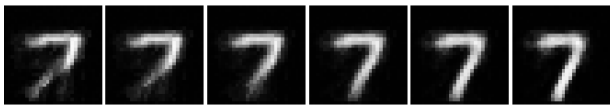


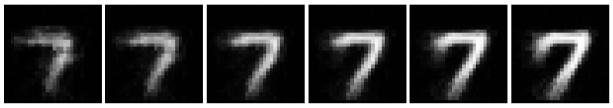
Figure 21: SDAE overfitting to the training data of the dense MNIST data when fine-tuning the SDAE.

Some of the general structure of the 7 data is captured in the model as for instance the first and second image from the first row of figure 20a and b are reasonably similar. However, other images such as the first image in the third row are poorly reconstructed; similar features show but in general the images are very different.

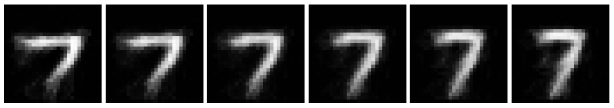
(a) Centre



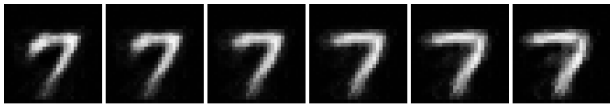
(b) Dim 1



(c) Dim 4



(d) Dim 5



(e) Dim 6

Figure 22: The SDAE model trained on the sparse data: Visualisation of the centre vector relative to our mapping M and the dimensions that appear most meaningful: 1, 4, 5 & 6.

As with subset 1, each dimension could be visualised by inserting our own vectors in $\mathbb{R}^8$ into the decoder of the trained autoencoder, see figure 22 for a selection of the dimensions visualised. Although the model is worse than the model that was created on the dense dataset, some similarities arise when we compare the visualised dimensions and the centre vector.

6 Experiments

The 80 images of the sketch map dataset were normalised similarly to the MNIST dataset; all pixel values were divided by 255 to scale all values to be in the range [0,1].

The same structure of testing is used as in section 5: first, the dimensionality of the sketch map dataset is estimated with use of the MLE algorithm. Subsequently we experiment with PCA on the sketch map dataset; we discuss what can be interpreted from the variation in this dataset and explore how and if we can visually interpret the eigenvectors that are extracted from this dataset. Following, the LLE algorithm is used to create a mapping to two dimensional space to retrieve insight in the topology of the data. Lastly the SDAE is trained on sketch map dataset. If a model is constructed that is able to encode each input vector from the input space to a useful lower dimensional mapping as indicated by the MLE, we will try to interpret the meaning of each remaining dimension. Additionally, we will compare some of the routes from "sketch map space" to the real route that the participants walked through the city to see if we are able to draw a comparison.

6.1 Dimensionality estimation

The maximum likelihood estimator estimated the intrinsic dimensionality of the sketch map dataset to be 4. This result could be an effect of the size of the dataset, which is small compared to its dimensionality, causing sparsity in a high-dimensional space and a decrease in accuracy when estimating the structure of the possible embedded manifold. If we compare this apparent intrinsic dimensionality from the sketch map data to that of the MNIST data, we can remark that the sparse MNIST data were estimated to have an intrinsic dimensionality of 8, whereas the dense MNIST data were estimated to have an intrinsic dimensionality of 12. It is thus good to notice that the intrinsic dimensionality of our sketch map data as estimated by the MLE possibly also shows this sample size dependency.

However, it could also mean that indeed much of the structure of the sketch maps that were drawn is shared, and a 4 dimensional representation could be found that correctly explains nearly all variations in the dataset.

The results from the experimental validation section do show us that it is probably still possible to generalise about the structure of the possible manifold from merely a sparse dataset, as reconstructions from the generated model for the sparse MNIST data were worse than those from the dense MNIST data, but still capable of showing some of the general structure of the data.

6.2 Principal component analysis

Figure 23 underneath shows the percentage each component contributes to the overall variance cumulatively. Eyeballing this bar plot, we can see that the cumulative variance each component explains seems to rise rapidly as the first 15 components already explain 50% of the variance in the data and 90% of the variance is explained by the first 50 components. This can indicate that the data has a lower intrinsic dimensionality.

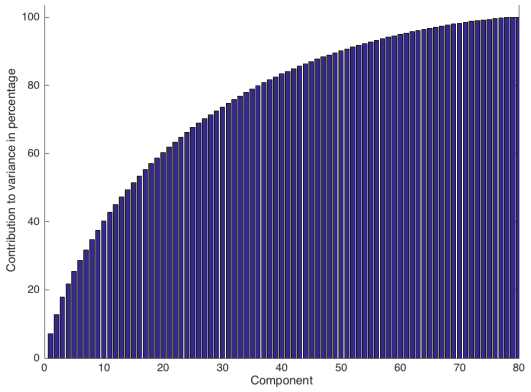
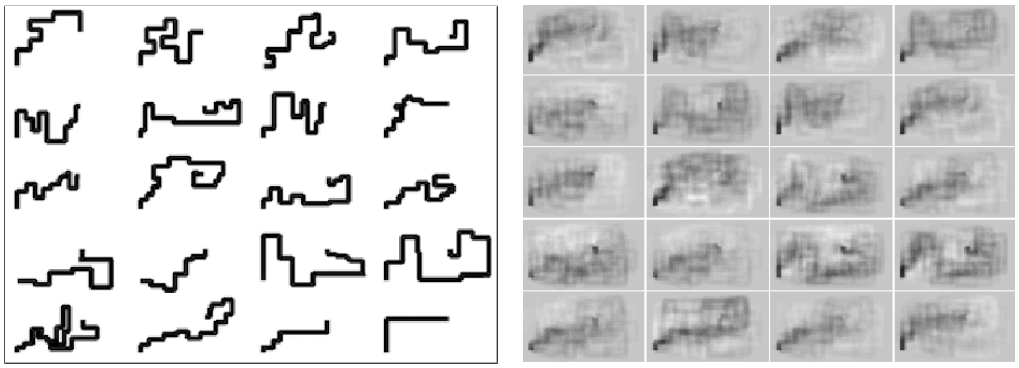


Figure 23: Scree plot of sketch map data: cumulative percentage the principal components contribute to the variance.

Figure 24 shows the reconstruction of images from the sketch map dataset after compression by the first four principal components, ±22% of the variance in our data. Also here, we choose 4 dimensions as this was the intrinsic dimensionality as estimated by the MLE. You can see that some of the characteristics such as length are somewhat preserved but that the reconstructions after compression to the amount of dimensions as estimated by the MLE are much worse than when we were dealing with the MNIST dataset. This could mean that there is little structure in the data but, as the MLE estimates the dimensionality to be 4, one could also hypothesise that the underlying topology is likely to be more non-linear than the MNIST dataset.



(a) Subset of the original data

(b) Reconstructed images after PCA compression to 4 components

Figure 24: Original sketch data and reconstruction after PCA compression

When visualising the first four eigenvectors of the images, we will call them *eigenmaps*, we seem to be able to deduce where the biggest variations in the sketch maps come from, see figure 25.

(a) mean map (b) eigenmap 1 (c) eigenmap 2 (d) eigenmap 3 (e) eigenmap 4

Figure 25: Mean and first 4 eigenmaps of the sketch map data visualised.

The most darkest or lightest pixels indicate areas with a larger variation to the mean route, visible in figure 25a. We know that the original route started in the southwest (bottom-left corner of the image) and ended in the northeast. It looks as if the first eigenvector and thus the most variation in the data separates the shorter from the longer routes. The second eigenvector seems to separate routes that mostly cover the upper part of the map from routes that mostly cover the bottom part of the map. The bottom two eigenmaps, component 3 and 4, are not so easy to interpret. They seem to involve rotation (eigenmap 3) and width of the loop of the second half of the route.

When we plot the first two dimensions, see figure 26, we can see that the images that are alike visually are close together in euclidean distance and that indeed the x-axis seems to indicate the length of the route. The y-axis seems less well interpretable. From figure 26 alone it is not clear if there are gender differences in performance on the sketch mapping task.

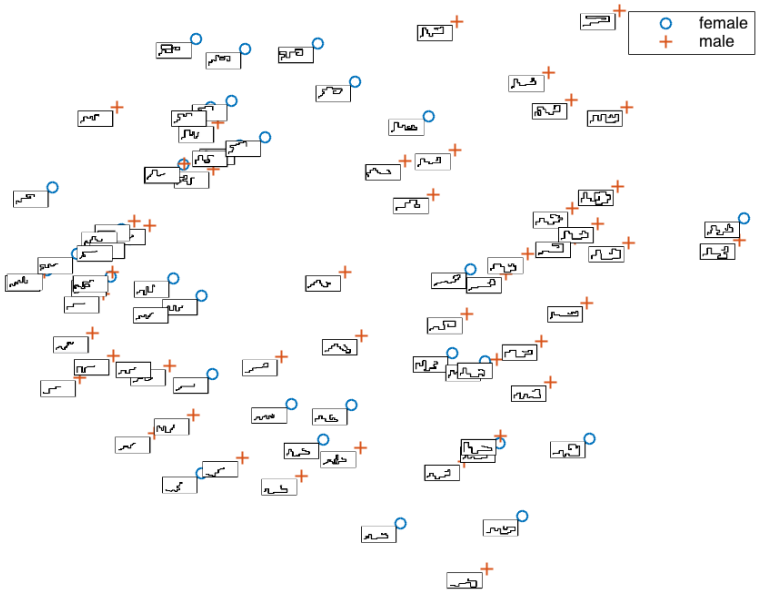


Figure 26: Plot of the sketch data mapped to 2 dimensions by PCA

Although these features can be interesting to analyse, we are not only interested in linear commonalities. To analyse more complex structures of our data, we must analyse images that are closer together in *geodesic* distance, the distance that follows the topology of the possible embedded manifold. To study these geodesic distances, we must find a non-linear mapping for our data.

6.3 Locally Linear Embeddings

Figure 27 shows a plot of the sketch maps mapped to two dimensions by the LLE algorithm, which aims to preserve locally linear distances and in this way the geodesic distances between the data, following the structure of the possible manifold. What is nice to see from this plot is that two samples that pollute the data are correctly seen as outliers. These were drawings from participants that very likely did not understand the task or were not able to reproduce the sketch map from memory: they took the shortest route from the begin point to the end point. Apparently, these fall outside of the structure of the possible manifold and thus do not possess the same intrinsic structure, found by the LLE algorithm, as the other sketch maps.

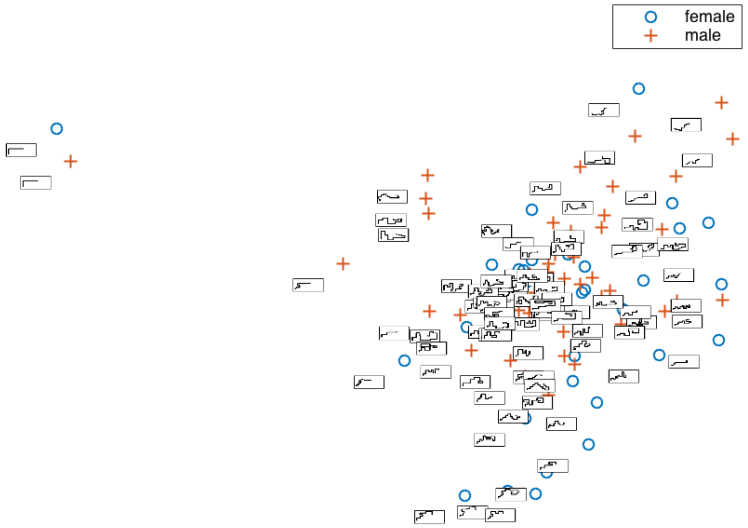


Figure 27: Plot of the sketch data mapped to 2 dimensions by LLE

Although gender does not perfectly separate the data, there definitely seem to be differences between the two classes: male sketch maps seem to reside higher on the plot than the sketch maps drawn by females. This result seems to correspond to the experiences that were reported by women during the experiment. Women reported to notice other things than men when walking through the fictional environment. They mentioned having troubles reproducing the route whereas men did not seem to report this issue. An interesting observation is that some women in this study explicitly reported to have had trouble remembering the route as they mostly noticed objects and colours in the environment. This observation conforms to the results of studies on innate gender differences in wayfinding; males share an innate preference to process metric information whereas females share the preference to process landmark information [Gavrielidou and Lamers, 2009] [Dabbs et al., 1998] [Eals and Silverman, 1994].

Because of differences between male and female wayfinding we have decided to leave out the sketch maps that were drawn by females to train the SDAE. In our research we are mainly interested in shared systematic distortions in metric information and as we are not sure if the females in our study processed this information, we will continue our study with the sketch map data acquired from males.

6.4 Stacked denoising autoencoders

A five layered network was used to encode the data of the sketch map dataset to a 4 dimensional encoding, and back to its original input space. Also here, as with the AE's that were constructed for the MNIST data, no tied weights were used: the weights from the encoder as well as the decoder were trained by use of gradient descent.

6.4.1 Experimental setup

The following architecture was used to set up the SDAE: 3680-36-4-36-3680. Here, 3680-36 indicates the AE that reduces the dimensionality from the image input space to 36, to search for features in the data. The second AE then reduces the dimensionality from 36 to 4, the size of the bottleneck, to ensure compression of the image data to 4 dimensions.

The separate DAE's and the final SDAE were trained with different combinations of values for the different parameters and the error plots of the combinations were eyeballed for the best learning curve. Table 4 shows the parameters that worked best and were finally used to train each DAE and the final SDAE model.

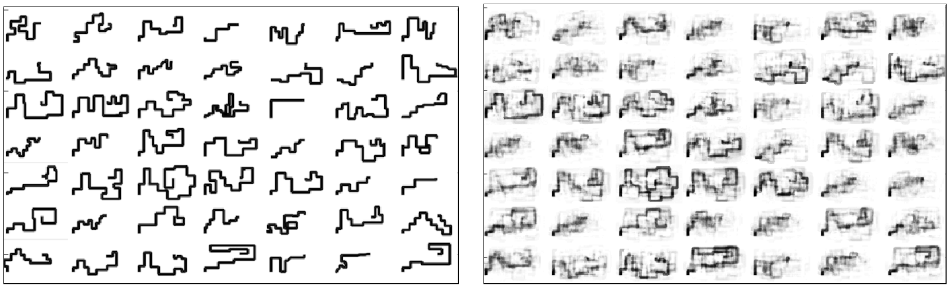
Final Parameters	DAE 1	DAE 2	Final model
Learning rate	0.01	0.01	0.002
Momentum	0	0	0
L2 Regularisation	0	0	0
Dropout	0	0	0
Masked noise	0.5	0.5	0.5
Batchsize	7	7	49
Epochs	15000	500	200000

Table 4: SDAE different training parameters sketch map data.

6.4.2 Results

After training the network on the subset of the sketch map dataset where $n = 49$, the SDAE was able to find a model that is able to reduce the sketch map images to 4 dimensional data, see figure 28.

We can see from image 28 that it is possible to reconstruct the images after the SDAE reduced the dimensionality of the images from 3680 to 4. The experimental validation section shows us that the quality of this representation is of course dependent on the sample size, but that it can still approximate the embedded structure. We can inspect this structure by feeding our own vectors in $\mathbb{R}^4$ into the decoder part of the SDAE.



(a) Subset of the original data

(b) Reconstructed images after SDAE compression to 4 dimensions

Figure 28: Subset of the sketch map dataset and reconstruction after AE compression

Also here, just as with the model that was created for the MNIST dataset, the mapped 4 dimensional data can possibly be viewed as the decorrelated data and the dimensions themselves represent variations, translations and deformations. Here, the decoder that is created by the autoencoder maps each vector to sketch map space, as each vector we feed into the decoder will turn out to be a different sketch map. We can inspect this sketch map space by inserting our own vectors in $\mathbb{R}^4$ into the mapping. The vector $\boldsymbol{v}$ relative to our mapping M is in the centre of our new sketch map space, see figure 30b.

$$\boldsymbol{v} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$$

All vectors in $\mathbb{R}^4$ that are fed into the decoder part of the SDAE produce different sketch maps with different features. By entering different values for a dimension in the range [0,1] we can attempt to visualise the meaning of this dimension, see figure 22. The values that were inserted in each dimension were again 0, 0.2, 0.4, 0.6, 0.8 and 1, resulting in 6 images per dimension. Also here, other values were kept constant with the value 0.5.

(a) Dim 1

(b) Dim 2

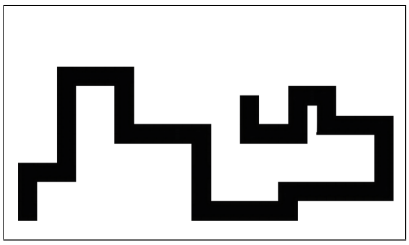
(c) Dim 3

(d) Dim 4

Figure 29: Visualisation of all dimensions

Analysing these images, we notice that they are noisy and that the transitions they represent are quite difficult to interpret, much more difficult than the smooth transitions that are visible when visualising the MNIST dimensions in 7 space.

Figure 30 shows the real route that participants had to walk through the fictional virtual city and puts it next to the route that is in the centre of the mapping as created by the SDAE. This image is constructed by feeding the vector $\boldsymbol{v}$ shown above into the decoder part of the SDAE and visualising its output.



(a) Real route

(b) Centre route

Figure 30: The real route and the route in the centre of map space created by the SDAE

Although both maps are resemblant of one another, it is difficult to draw any conclusions regarding the model that is created by the SDAE. Likely, the model learnt most of the features of this sparse dataset and merely reconstructs these throughout the model. If more data is gathered, more generalisable features will possibly be learnt such as with the large MNIST dataset.

7 Conclusion

Inspired by recent advances in machine learning, this paper explores the use of pattern recognition algorithms to search for meaningful structures in products of the human mind, specifically sketches. We were motivated by the idea that sketches, although idiosyncratic, could show us what form concepts take in mental processes. If we would be able to find patterns in sketches that all refer to the same concept or entity, we could capture some of the shared characteristics of a mental representation and implicitly some of the mechanisms of the mind.

In this study we narrowed our scope of interest down to the field of environmental knowledge processing; we searched for patterns in sketch maps that could teach us more about the storage and retrieval of spatial knowledge. Specifically, we started this study with an interest in finding commonalities and variations in sketch maps that could lead us to generalise about common errors that people make or tendencies that people have when storing and retrieving spatial knowledge. Examples are over- or underestimating distances, the rotation of locations relative to each other and the simplification of certain shapes of routes.

Although it proves difficult to analyse the patterns found in our experiments, we will mention some of the interesting discoveries, possible interpretations of results and possible improvements that could ensure better results in the future.

First, it is interesting to notice that the MLE algorithm estimates the intrinsic dimensionality of the sketch map data to be 4, meaning that the data possibly lie on a much lower embedded manifold. These results coincide with the results from the SDAE experiments, which show good reconstructions after a compression of the data to 4 dimensions. The problem remains that the SDAE likely overfits to the sketch map data, whereas when we would have a much larger dataset and a test set to validate the training, we could likely find a more generic model to fit the data. Also, if we keep in mind sample size dependency, we could hypothesise that the intrinsic dimensionality is most likely somewhat higher than 4. From a psychological viewpoint it is exciting to research what each feature from our obtained model would capture. If there is only little variation between the sketches, variation across 4 dimensions, there must be many shared features. When visualising the dimensions of a well trained model, we could capture transitions in the data that could explain to us some of the workings of spatial cognition. For instance, a dimension from our sketch map data cloud could show a transition from small rotations to large rotations circumnavigating the endpoint, which could indicate that a percentage of the participants shows a tendency to experience the route to the endpoint as a circle rather than two 90°angles.

Even though PCA is only capable of finding linear structures in the data, it is interesting to inspect the eigenvectors. If the topology of sketches is somehow fixed or sketches are rotated and translated prior to PCA to remove some of the non-linearity, PCA is able to find interesting patterns. One of the nicest properties of PCA is that the eigenvector corresponding to the largest eigenvalue points to the biggest variation in the data and thus this eigenvector can be interpreted as an important linear dimension. In this study, as the routes in each of our sketches had the exact same begin and endpoint, it is interesting to interpret these biggest linear variations. These appeared to be the length of the route and the usage of the lower or upper part of the map. What we can hypothesise about this finding is that some participants approached the endpoint from the left as their routes were much shorter than the other routes, while the endpoint was in reality approached from the right. Also we can hypothesise that some participants experienced the route as being much shorter than it actually was.

Further we discovered that, by use of the LLE algorithm, we were able to reveal outliers in the data; samples that did not conform to the structure of the possible embedded manifold as found by the algorithm. Also participants' gender differences became visible when reducing the dimensionality non-linearly. This leads us to believe that other class differences can also be detected using this method. If for instance, after reducing the dimensionality non-linearly and plotting our data in two or three dimensions, classes become clearly separable, we could hypothesise that measures of spatial

cognition are dependent of a specific class. What these differences possibly are can be discovered by analysing the dimensions that appear from the model constructed by the SDAE, see figure 29.

Using a SDAE to construct a model of our data proved interesting as it allowed us to view certain images that were not in the original dataset but which were created by the decoder from the SDAE. The decoder, as mentioned extensively in prior sections, maps each vector, in this case in $\mathbb{R}^4$, to map space. The images that were created along each dimension gave us insights into the meaning of these dimensions. Unfortunately the images in figure 29 are noisy and difficult to interpret. When visualising the centre vector in map space, see figure 30, we could see that the image was very much comparable to the real route, except for some of the streets that appeared longer and some that were estimated shorter. It remains uncertain how we can interpret this centre vector and whether its variations from the real route are in any way interpretable as shared systematic errors in the cognitive maps of participants. To gain access to these correlations that exist in the data we probably require a method to visualise the features *learned* by the nodes in the neural network.

One thing that we noticed that could be improved in further research was the difficulty of the sketch map exercise. Many participants indicated that they believed the mapping exercise was very difficult and that they thought they had not produced an accurate sketch map that resembled the real route. We noticed that the short time participants spent in the fictional virtual environment affected the learning of that environment negatively. In further research it would be better to increase this period of time, allowing participants to better construct a mental representation and to, as a result, produce better externalisations of the route.

Further, for future research, it would be a good idea to try out a different representation of our data. In this study we used the pixel data of our sketch maps to represent each sketch in $\mathbb{R}^{3680}$. This approach allows for a visual interpretation of the outcomes, but also adds unnecessary dimensions and noise to the data (we do not need all white pixels and black pixels to capture the information in our data). Another option could be to analyse a distance matrix extracted from each sketch map. In this way, noise is reduced and possibly better representations can be found with less data. Additionally, we would likely require less data to construct a good model of the underlying structure and possible embedded manifold as we decrease the amount of features per sketch map.

Finally, as this thesis merely researches the use of three DR methods on sketches, we could of course also test the effectiveness of different methods within this approach. The t-SNE algorithm for instance is able to very accurately visualise high-dimensional data in low dimensional space [Maaten and Hinton, 2008], but other methods such as Isometric Mapping (ISOMAP) could also be used to map the data to lower dimensions. Additionally it would be nice to experiment with class differences: more meta data could be gathered about participants. Plotting the sketches in $\mathbb{R}^2$ or $\mathbb{R}^3$ and adding colours for different classes could possibly reveal interesting clusters in the data.

Regarding the explorative nature of this study, we are happy with its results. These show that pattern recognition methods might provide us with objective knowledge about spatial cognition and with a tool to easily visualise class differences from sketches. We believe that this study will further encourage others to research how pattern recognition can aid psychology research.

Acknowledgements

I want to thank Maarten Lamers for guiding me through this research project and providing me with useful insights and expertise on the topic of dimensionality reduction, Thomas Barkowsky for his expertise on both the field of spatial cognition and informatics, Rasmus Berg Palm for providing access to his implementations of different neural networks, Wojtek Kowalczyk for his expertise on deep learning, Laurens van den Maaten for being able to use his online dimensionality reduction toolkit and Purple Jump, professional artists building virtual reality from whom we used the town constructor pack for constructing the digital environment.

References

[Abdi and Williams, 2010] Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

[Barkowsky, 2002] Barkowsky, T. (2002). *Mental representation and processing of geographic knowledge: a computational approach*. Springer.

[Bengio et al., 2006] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al. (2006). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153–160.

[Billinghurst and Weghorst, 1995] Billinghurst, M. and Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings of the Virtual Reality Annual International Symposium*, pages 40–47. IEEE Computer Society Washington.

[Blades, 1990] Blades, M. (1990). The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, 10(4):327–339.

[Dabbs et al., 1998] Dabbs, J. M., Chang, E.-L., Strong, R. A., and Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior*, 19(2):89–98.

[DeMers et al., 1992] DeMers, D., Cottrell, G., et al. (1992). no. *Advances in neural information processing systems*, 5:580–587.

[Eals and Silverman, 1994] Eals, M. and Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, 15(2):95–105.

[Gavrielidou and Lamers, 2009] Gavrielidou, E. and Lamers, M. H. (2009). Landmarks and time-pressure in virtual navigation: Towards designing gender-neutral virtual environments. In *Facets of Virtual Environments*, pages 60–67. Springer.

[Golledge et al., 1985] Golledge, R. G., Smith, T. R., Pellegrino, J. W., Doherty, S., and Marshall, S. P. (1985). A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *Journal of Environmental Psychology*, 5(2):125–152.

[Hinton and Salakhutdinov, 2006] Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

[Institut für Geoinformatik, 2016] Institut für Geoinformatik (2016). Sketchmapia. [Online; accessed 15-April-2016], *http://www.uni-muenster.de/Geoinformatics/en/sketchmapia/sketch-map-database.php*.

[Kaplan, 1973] Kaplan, S. (1973). *Cognitive maps in perception and thought*. Aldine.

[Kitchin, 1994] Kitchin, R. M. (1994). Cognitive maps: What are they and why study them? *Journal of environmental psychology*, 14(1):1–19.

[Kosslyn, 1973] Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, 14(1):90–94.

[Kosslyn, 1994] Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

[Larochelle et al., 2009] Larochelle, H., Bengio, Y., Louradour, J., and Lamblin, P. (2009). Exploring strategies for training deep neural networks. *Journal of Machine Learning Research*, 10(Jan):1–40.

[LeCun et al., 1998] LeCun, Y., Cortes, C., and Burges, C. J. (1998). The mnist database of handwritten digits. [Online; accessed 20-July-2016], *http://yann.lecun.com/exdb/mnist/*.

[Lee and Verleysen, 2007] Lee, J. A. and Verleysen, M. (2007). *Nonlinear dimensionality reduction*. Springer.

[Levina and Bickel, 2004] Levina, E. and Bickel, P. J. (2004). Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17:777–784.

[Lynch, 1960] Lynch, K. (1960). *The image of the city*. MIT press.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

[Maaten et al., 2009] Maaten, L. v. d., Postma, E., and Herik, J. v. d. (2009). Dimensionality reduction: A comparative review. *Technical Report, TiCC TR 2009-005*, Tilburg University.

[Meneghetti et al., 2015] Meneghetti, C., Pazzaglia, F., and De Beni, R. (2015). Mental representations derived from spatial descriptions: the influence of orientation specificity and visuospatial abilities. *Psychological research*, 79(2):289–307.

[Neisser and Kerr, 1973] Neisser, U. and Kerr, N. (1973). Spatial and mnemonic properties of visual images. *Cognitive Psychology*, 5(2):138–150.

[Palm, 2012] Palm, R. B. (2012). Prediction as a candidate for learning deep hierarchical models of data. *MSc. Thesis Informatics*, Technical University of Denmark.

[Pylyshyn, 2002] Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and brain sciences*, 25(02):157–182.

[Raudys and Jain, 1991] Raudys, S. J. and Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on pattern analysis and machine intelligence*, 13(3):252–264.

[Richardson et al., 1999] Richardson, A. E., Montello, D. R., and Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & cognition*, 27(4):741–750.

[Rosnow, 2002] Rosnow, R. L. (2002). The nature and role of demand characteristics in scientific inquiry. *Prevention & Treatment*, 5(1):article 37.

[Roweis and Saul, 2000] Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

[Sanders and Porter, 1974] Sanders, R. A. and Porter, P. W. (1974). Shape in revealed mental maps. *Annals of the association of american geographers*, 64(2):258–267.

[Stevens and Coupe, 1978] Stevens, A. and Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive psychology*, 10(4):422–437.

[Tobler, 1976] Tobler, W. R. (1976). The geometry of mental maps. In *Spatial choice and spatial behavior*, pages 69–81. Ohio State University Press.

[Tversky, 1992] Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23(2):131–138.

[Tversky, 1993] Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In *Spatial Information Theory: A Theoretical Basis for GIS*, pages 14–24. Springer.

[Tversky, 2002a] Tversky, B. (2002a). Navigating by mind and by body. In *International Conference on Spatial Cognition*, pages 1–10. Springer.

[Tversky, 2002b] Tversky, B. (2002b). What do sketches say about thinking. In *AAAI Spring Symposium, Technical Report SS-02-08*, pages 148–151, Stanford University.

[Tversky and Lee, 1999] Tversky, B. and Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In *International Conference on Spatial Information Theory*, pages 51–64. Springer.

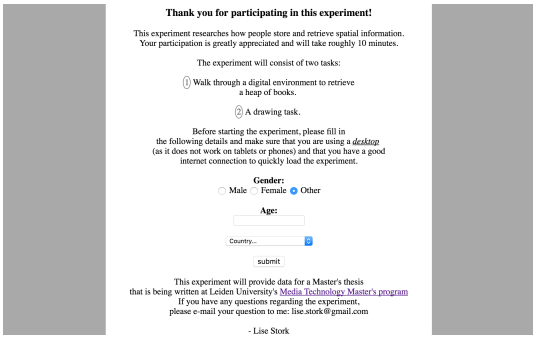
[Vincent et al., 2010] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408.

[Wang and Schwering, 2015] Wang, J. and Schwering, A. (2015). Invariant spatial information in sketch maps: a study of survey sketch maps of urban areas. *Journal of Spatial Information Science*, 11(1):31–52.

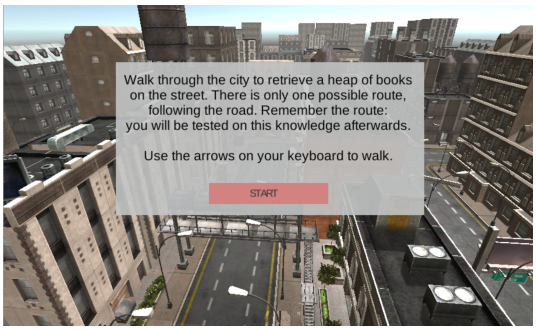
[Waterman and Gordon, 1984] Waterman, S. and Gordon, D. (1984). A quantitative-comparative approach to analysis of distortion in mental maps. *The Professional Geographer*, 36(3):326–337.

Appendix 1: Experiment

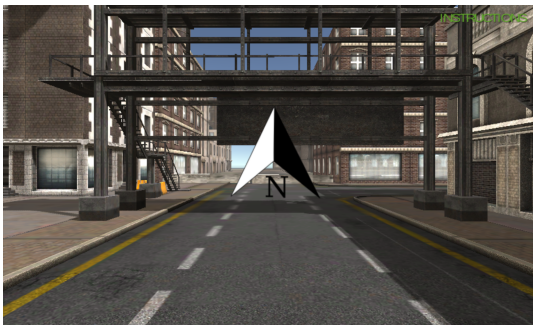
Flow of the online experiment from (a) to (r).



(a)



(b)

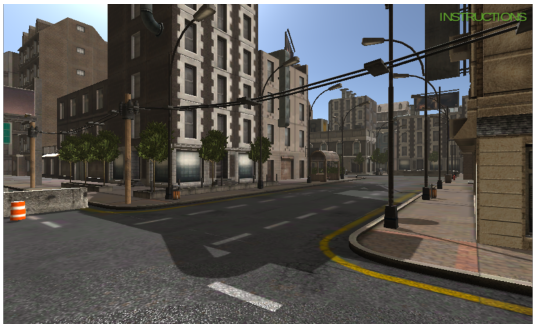


(c)

(d)

(e)

(f)



(g)

(h)


INSTRUCTIONS

(i)


INSTRUCTIONS

(j)


INSTRUCTIONS

(k)


INSTRUCTIONS

(l)


INSTRUCTIONS

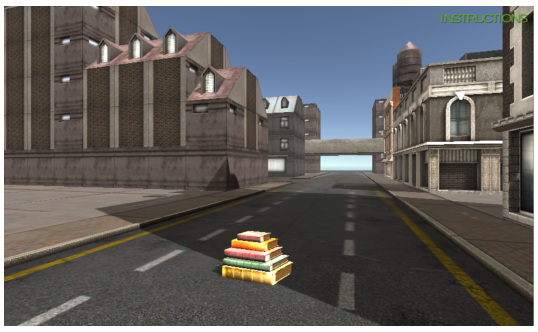
(m)


INSTRUCTIONS

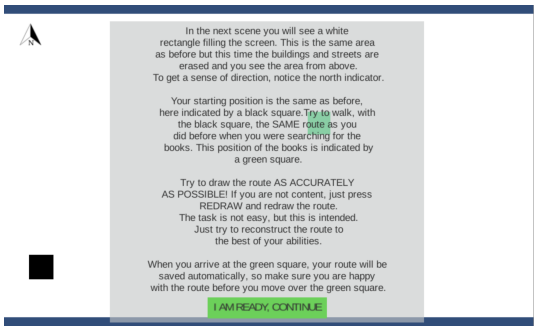
(n)


INSTRUCTIONS

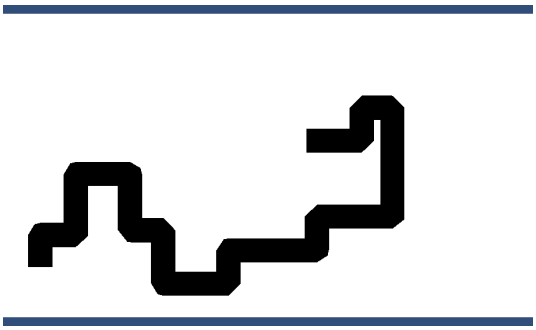
(o)


INSTRUCTIONS

(p)


In the next scene you will see a white
rectangle filling the screen. This is the same area
as before but this time the buildings and streets are
erased and you see the area from above.
To get a sense of direction, notice the north indicator.
Your starting position is the same as before,
here indicated by a black square.Try to walk, with
the black square, the SAME route as you
did before when you were searching for the
books. This position of the books is indicated by
a green square.
Try to draw the route AS ACCURATELY
AS POSSIBLE! If you are not content, just press
REDRAW and redraw the route.
The task is not easy, but this is intended.
Just try to reconstruct the route to
the best of your abilities.
When you arrive at the green square, your route will be
saved automatically, so make sure you are happy
with the route before you move over the green square.
I AM READY, CONTINUE

(q)



(r)