universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Predicting Psychological Embeddings of Fear-Related Stimuli in a Multidimensional Space Using Deep Neural Networks

verfasst von / submitted by

Dominik Pegler, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Dr. Frank Scharnowski, MSc
Mitbetreut von / Co-Supervisor:	Dr. Filip Melinščak, BSc MSc

Acknowledgments

A big thank you goes to my co-supervisor Filip, who was there for me almost 24/7 with his witty mind and inspiring ideas. I am grateful to Mengfan Zhang (张 梦 凡) for always sitting down with me and doing all the data collection work. Then, hovering over it all, was my supervisor Frank, whom I would like to thank for inviting me into this lab he created, with its talented people who make it possible to generate all these ideas and put them into practice. 🕷

Contents

1 Abstract

Fear-related visual stimuli are often used in exposure therapy, a popular method of treating anxiety disorders. To better understand anxiety disorders, it is helpful to understand how fear-related stimuli are mentally represented. Similarity judgments of such stimuli are often used to infer mental representations as a multidimensional space. Collecting them is a time-consuming and often impossible task. The present study aims to investigate whether these representations can be generated using artificial neural networks, taking arachnophobia as a specific example. Previous research has only used this approach with images of simple objects, but not with more complex images. An online experiment was conducted to collect similarity judgments of spider images from 77 crowdsourced participants. Multidimensional scaling (MDS) was used to create a latent space of the images, in which their similarities were reflected by their distances along a predetermined number of dimensions. An ensemble of convolutional neural networks (CNNs) was then trained to reproduce these multidimensional embeddings. All four resulting MDS dimensions could be successfully predicted by the CNNs. Furthermore, when applied to an entirely new and unseen set of stimuli, the CNNs were able to create a latent space that resembled the dimensions of the original space. The results show that CNNs can predict the embeddings of complex, fear-related images. However, the study leaves open whether these stimuli are represented differently in clinical populations than in the general population. Other limitations, such as the small sample size and an unsystematically collected image set, could be overcome in future studies. In any case, the method could provide insight into the development of anxiety therapies by addressing the questions of where in this multidimensional space the images most associated with anxiety are located and how to select images for exposure therapy from this space to achieve optimal outcome for the individual patient.

Keywords: Arachnophobia, Anxiety Disorder, Mental Representations, Convolutional Neural Networks, Multidimensional Scaling, Exposure Therapy

2 Introduction

2.1 Anxiety Disorders

Anxiety disorders such as phobias are highly prevalent and a large burden on society. One in three people will experience them during their lifetime (Bandelow & Michaelis, 2015; Craske et al., 2017), so it is clearly important to get to know the main causes and symptoms of anxiety and how to treat them. Studying specific phobias is a good way to gain a better understanding, and spider phobia (arachnophobia) is the most prevalent example (Fredrikson et al., 1996), affecting about three percent of the population (Oosterink et al., 2009). Spiders evoke more fear and disgust in people than any other animal (Polák et al., 2020). This suggests that arachnophobia is a widespread problem that may affect a significant number of people, causing emotional distress and potentially limiting their quality of life due to avoidance behaviors. Although spider phobia does not typically come with such major societal costs such as other specific phobias (Fehm et al., 2005), we used it as the subject to study because of its high prevalence, and because settings for exposure therapy can be created very easily in order to better understand its mechanisms.

2.2 Exposure Therapy

The most common treatments for specific phobias are behavioral therapies, such as exposure therapy, which is considered the gold standard. In particular, in vivo exposure, which involves confrontation with the object or situation in real life (e.g., touching a real spider), consistently results in high rates of success (Wolitzky-Taylor et al., 2008). Despite the availability of effective interventions, many people with specific phobias are reluctant to seek treatment. One reason for this is that high levels of fear lead many phobic individuals to view themselves as simply untreatable when exposed to the feared object (Wolitzky-Taylor et al., 2008). To counteract this and to increase the rate of treatment completion, computer-aided exposure therapies were developed some time ago (Dewis

et al., 2001). The patient is exposed to images or videos of the feared object or situation, which are displayed on a screen (Piercey et al., 2012). We suggest that careful selection and sequencing of stimulus images for computer-aided treatments tailored to individual patients is critical to achieving the best possible outcome. This includes determining the sequence of stimuli that the patient will be exposed to. In particular, it may be important for the therapist or the selection algorithm to be aware of the properties of the available stimuli for optimal selection.

2.3 Psychological Embeddings of Fear-Related Stimuli

In order to understand the phobic response to a stimulus, it is helpful to try to understand how that stimulus is represented mentally. In the present study, we conceptualize these psychological embeddings or mental representations as representations of similarity (Shepard & Chipman, 1970). It refers to the mental distance or similarity between different objects or concepts. This can be represented mathematically as a multidimensional space, where each dimension represents a different feature or feature constellation. For example, in a psychological space for objects, the dimensions might represent the features of shape, texture, and color. In this space, each point represents a different object, and the distance between the points reflects their perceived similarity. Objects that are closer together are perceived as more similar, while objects that are further apart are perceived as more different. Using multidimensional embeddings as a model for mental representations is supported by recent work by Hebart et al. (2020), which has shown that multidimensional embeddings can accurately predict human categorization behavior of natural objects. Furthermore, it suggests that humans are able to rate objects along these dimensions with a high degree of accuracy. These results provide strong evidence for the validity and effectiveness of this approach in predicting and understanding human perception of natural objects.

2.3.1 Similarity Judgments

Similarity judgments are a common way to capture representations of similarity. Typically, these are numerical ratings between two stimuli from which a distance matrix can be computed. Unfortunately, pairwise ratings are time-consuming and can become intractable with large sets of stimuli. For this reason, more efficient methods have been developed to collect similarity judgments. The spatial arrangement method (SpAM; Hout et al., 2013) is such a method. In this task participants are shown multiple stimuli at once on a computer screen and are asked to place them on the canvas according to their similarity to one another. The distances between any two stimuli can then be used to compute a distance matrix.

2.3.2 Creating Psychological Embeddings Using Multidimensional Scaling

Multidimensional scaling (MDS) is a statistical technique for analyzing and representing complex data in a lower dimensional space (Torgerson, 1967). The objective is to provide a multidimensional representation of the data that preserves the (dis)similarities between objects in the form of their distances. The starting point for MDS is the distance matrix created from the pairwise similarity ratings. The resulting *n*-dimensional space can then be used to create simple two-dimensional plots to visualize the underlying properties that may have influenced the similarity ratings.

2.3.3 Prediction With Deep Convolutional Neural Networks

Collecting similarity judgments can be a time-consuming task, which often becomes unmanageable as the number of comparisons increases quadratically with the number of stimuli. With a quadratic increase in stimulus combinations, it would quickly become neither realistic nor reasonable to present them all to the participants. For example, 10 stimuli would require 45 pairwise ratings, 100 stimuli would require about 5,000, and 1,000 stimuli would require about 500,000 pairwise ratings. Formally, this can be represented as follows (where n is the number of stimuli and k the number of pairwise comparisons):

$$k = \frac{n(n-1)}{2}. \tag{1}$$

Advanced and adaptive data collection methods can increase the number of objects that can be embedded in this multidimensional space (Roads & Mozer, 2019), but will always be limited by the availability of data collection resources, while future research may require the ability to quickly embed a virtually unlimited number of objects. For this reason, we want to find out whether Convolutional Neural Networks (CNNs), which are increasingly used for all kinds of image categorization or rating problems (LeCun et al., 2015), can be used to generate such embeddings quickly and accurately. A recent work by Sanders and Nosofsky (2020) could show that CNNs were able to succesfully predict the coordinates of novel images of rocks in a multidimensional space derived from human similarity judgments. This approach treated CNNs solely as a machine learning (ML) method for making predictions, not as models for cognition. The present study takes inspiration from this work and aims to show that this approach can be successfully transferred to more complex imagery used in phobia research.

2.4 Expected Results

Our research question is whether CNNs are able to accurately predict these dimensions for novel spider images. For simple rock images, the correlations between observed and predicted MDS coordinates were $r > .80$ in three of the eight dimensions, and averaged $r = .68$ across all eight dimensions (Sanders & Nosofsky, 2020). This will serve as our benchmark. The level of accuracy required to work efficiently with these visual stimuli in real-world scenarios, such as exposure therapy, remains to be determined, but as we expect the results to be lower than in previous research, we set this level of sufficiency at a correlation coefficient of at least $r = .60$ ($R^2 = .36$) for each dimension. We also expect the MDS embeddings to have at least three dimensions and to be roughly related to visual aspects such as color, brightness, and the type of presence of a spider in the image (zoomed in, on a web, etc.). For this purpose, an advanced web-based online experiment will be used,

which is able to collect more than one pairwise similarity judgment per trial and can be combined with crowdsourcing platforms.

2.5 Aims of This Study

In summary, there are two aspects that we want to highlight: the application of new research methods and a step towards a deeper study of phobias. Our objective is to show how these methods are a useful addition to the toolbox of psychological research and what new insights can be gained as a result. This research could potentially contribute to the development of exposure therapy for anxiety disorders. The use of multidimensional space analyses could improve the accuracy of identifying the most effective stimuli for individual patients, resulting in better outcomes for patients undergoing exposure therapy.

2.6 Planned Procedure

To briefly summarize what will be done: we will collect similarity judgments of fear-related images (spiders), for which we will then use MDS to embed these images in a multidimensional space. We will train an ensemble of CNNs to learn the association between the image content and the latent coordinates in the multidimensional space, and finally test the predictive performance of the CNNs on a hold-out dataset. We will also demonstrate an example use case by having the CNNs construct a multidimensional solution for a large set of 1,516 novel spider images. The interpretation of MDS dimensions is beyond the scope of the current study, but could be the subject of further research.

3 Methods

This study was approved by the Ethics Committee of the Faculty of Psychology at the University of Vienna, and all methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from the participants before enrollment in the study. Open science practices were applied throughout the study. All materials, experiment code, collected data and data analysis code are stored on

https://osf.io/e64x7 and will be made publicly available upon publication. Figure 1 provides an overview of the entire analysis pipeline.

3.1 Data

3.1.1 Experimental Design

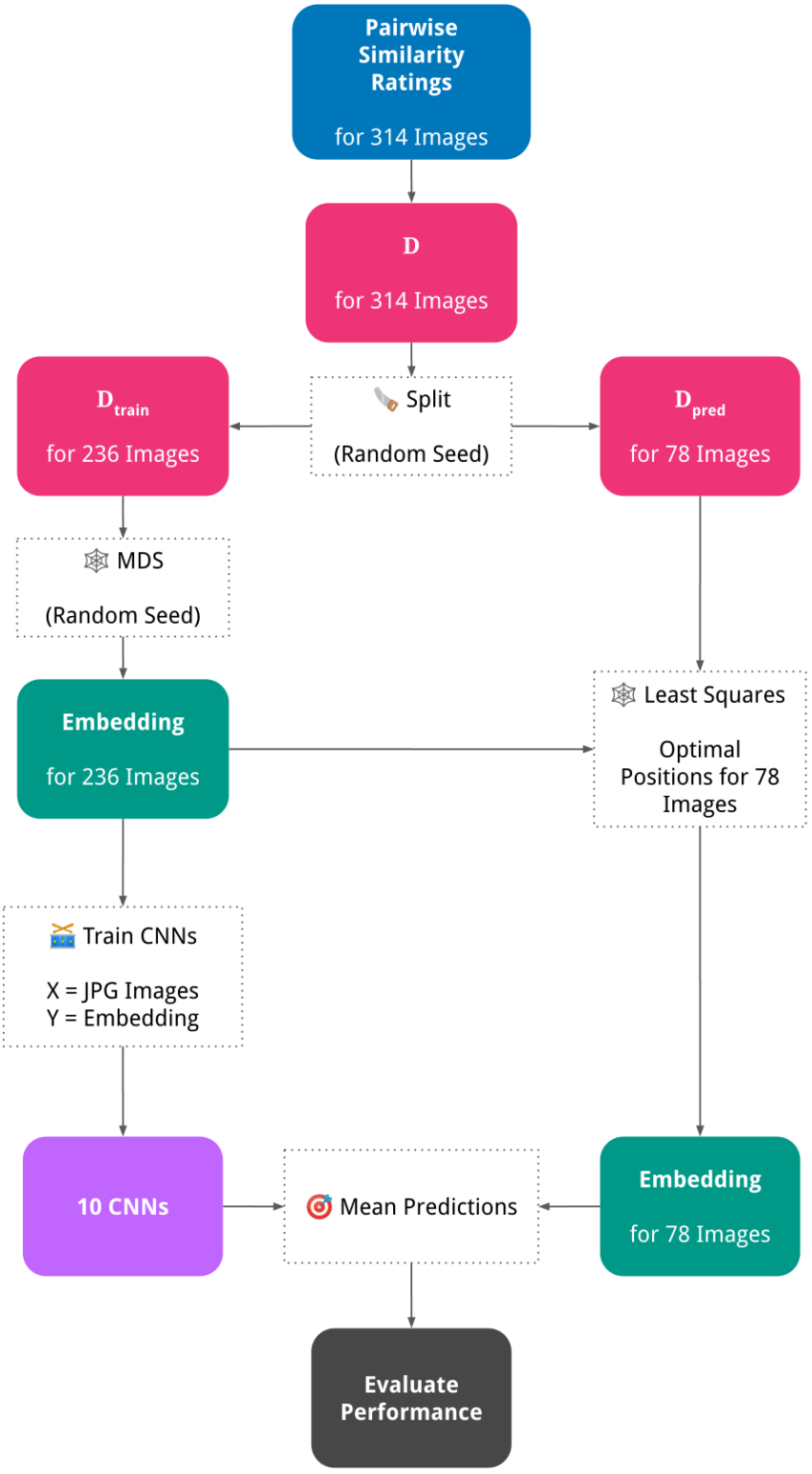
The Spatial Arrangement Method (SpAM; Hout et al., 2013) was used as the experimental approach to collect similarity judgments. This web-based technique presents a greater number of stimuli simultaneously on the screen in each trial, allowing for the collection of a greater number of pairwise similarity judgments per trial in very little time. This method is well suited for online implementation, which allowed us to recruit participants via crowdsourcing platforms. On each trial, participants were presented with 16 stimuli that were randomly selected and placed on the screen. They were then instructed to use drag-and-drop to arrange the stimuli so that the distances between each item reflected their perceived similarity. In this context, items that were placed closer together were considered to be more similar. Participants completed a total of 10 trials, eight of which were different and two of which repeated the same set of stimuli. The median time for each participant to complete the experiment was 25 minutes. In addition to the similarity judgments, the data set collected consisted of the Fear of Spiders Questionnaire (FSQ) and demographic data.

3.1.2 Participants

In this study, a total of 77 participants were recruited online through the crowdsourcing platform prolific.co and paid USD 9.60 per hour to participate. All participants were drawn from the general prolific.co pool and were not pre-screened for arachnophobia. The mean FSQ score of the participants was M = 50.06 (SD = 31.95). Of the 77 participants, 28 were female, 48 were male and one person did not declare their gender,

Figure 1

Overview of the procedure: From similarity judgments, to distance matrices, to multidimensional embeddings, to making predictions with the CNNs. Distance matrices are denoted by D.



giving a female percentage of 36.36%. The mean age of the participants was M = 34.1 years (SD = 12.2).

3.1.3 Sample Size

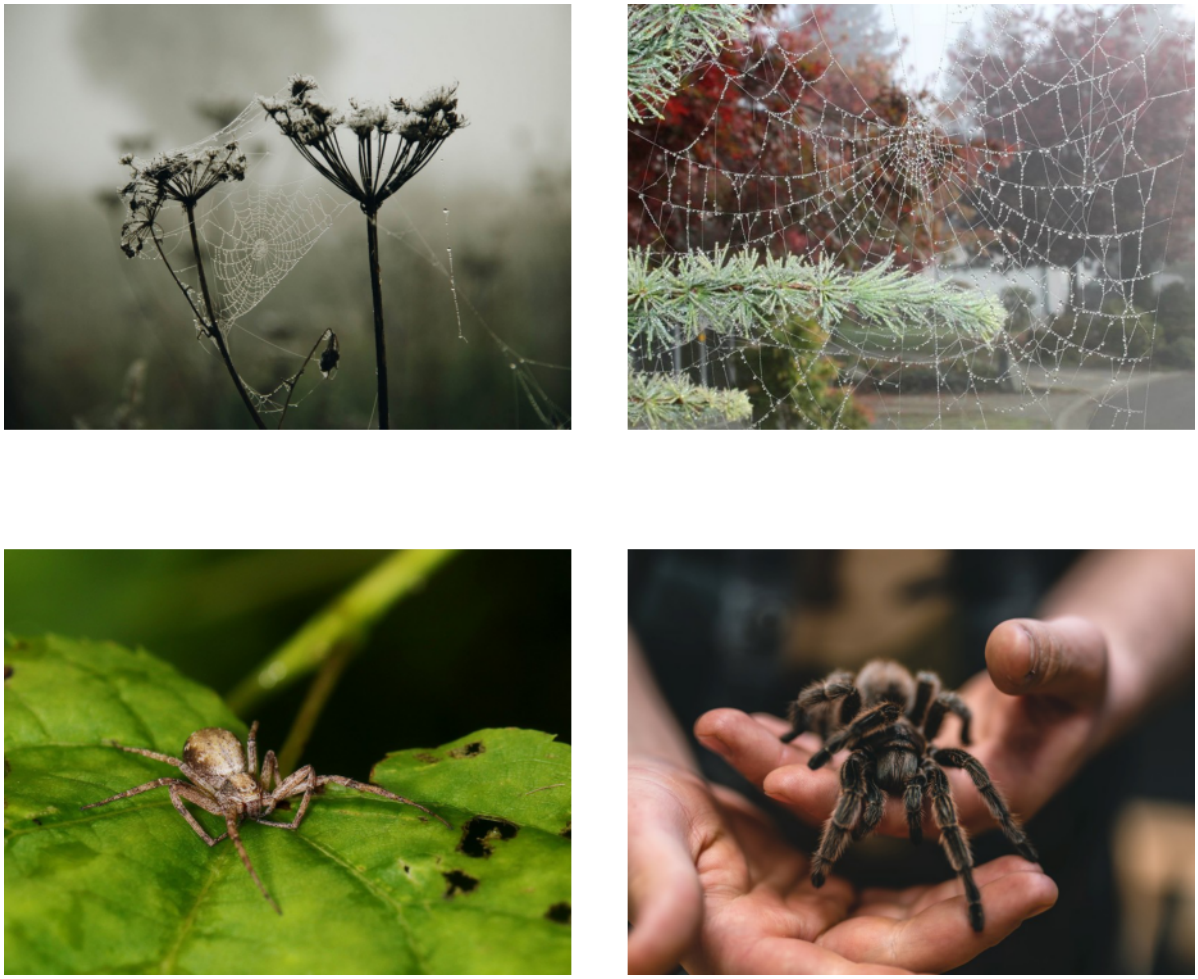
We did not use a traditional power calculation to determine the sample size for our study because there were no readily available calculations for MDS at the time. However, our goal was not to test a specific effect within a population, but to see if neural networks could replicate a human-generated MDS space. To ensure that the CNNs did not perform poorly due to a non-robust MDS space, we aimed to recruit enough participants to achieve a certain level of robustness. During data collection, we performed a convergence analysis by adding 11 participants at a time. We found that after 77 participants, the pairwise distances between MDS solutions (we computed eight separate MDS solutions, one for each number of dimensions) remained almost unchanged, indicating that a certain level of robustness had been achieved.

3.2 Materials

In our online experiment and subsequent construction of the MDS space and CNN training procedure, a diverse set of 314 spider-related images was used. These images depicted different types of spiders in different contexts, such as spiders on plants, humans, spider webs, and of different sizes. These images were obtained from a previous study conducted at the host research institution (Figure 2). Notably, this image dataset was not constructed in a representative or targeted manner, but rather as a convenience sample of online images. The second image set containing 1,516 novel spider-related images was obtained from https://images.cv.

Figure 2

A similar (top) and a dissimilar image pair (bottom) based on participant ratings



3.3 Models

3.3.1 Creating MDS Embeddings

Across all participants the average distance for each image pair was computed which built the basis for the distance matrix $\mathbf{D}$. This matrix was then randomly split into a training set with 236 (75%) and a prediction set with 78 (25%) images, the resulting matrices are denoted as follows:

$$\mathbf{D}_{\text{train}} = \begin{bmatrix} \mathbf{D_{1,1}}, \mathbf{D_{1,2}}, \ldots, \mathbf{D_{1,236}} \\ \mathbf{D_{2,1}}, \mathbf{D_{2,2}}, \ldots, \mathbf{D_{2,236}} \\ \vdots \\ \mathbf{D_{236,1}}, \mathbf{D_{236,2}}, \ldots, \mathbf{D_{236,236}} \end{bmatrix}, \tag{2}$$

$$\mathbf{D}_{\text{pred}} = \begin{bmatrix} \mathbf{D_{1,237}}, \mathbf{D_{1,238}}, \ldots, \mathbf{D_{1,314}} \\ \mathbf{D_{2,237}}, \mathbf{D_{2,238}}, \ldots, \mathbf{D_{2,314}} \\ \vdots \\ \mathbf{D_{236,237}}, \mathbf{D_{236,238}}, \ldots, \mathbf{D_{236,314}} \end{bmatrix}. \tag{3}$$

$\mathbf{D_{train}}$ was used to compute a metric MDS embedding (Lee, 2001; Shepard, 1987; Torgerson, 1967) using the Python library scikit-learn (Pedregosa et al., 2011). Maximum iterations was set to 5,000, the relative tolerance parameter (eps) to 1e-9. For each of the 78 images in the prediction set, based on the distances between these images and the 236 training images as described by $\mathbf{D_{pred}}$, we then used least-squares optimization to find the position in space that resulted in the minimum distance error. This allowed us to determine the positions of the images in the prediction set while leaving the training embedding untouched. Preventing this form of data leakage was essential, as falsely good predictions could have been made if the CNNs had been trained on a dataset containing information from the prediction set. The code for computing the MDS space and the code and description of the optimal embedding of new stimuli in an existing MDS space can be found in Appendix B.

Dimensions. The dimensionality of the image embeddings was determined using the normalized stress (Stress–1) and the Bayesian information criterion (BIC). The normalized stress, which is a measure of the discrepancy between the pairwise distances in the initial distance matrix and the corresponding distances in the resulting MDS space, was computed as part of the MDS analysis (see Appendix B for code examples). The BIC was calculated using the method proposed by Lee (2001), where m is the number of dimensions, n is the number of stimuli, K is the number of subjects (where each individual subject is indexed by k), and s is a sample estimate of the data precision population parameter σ:

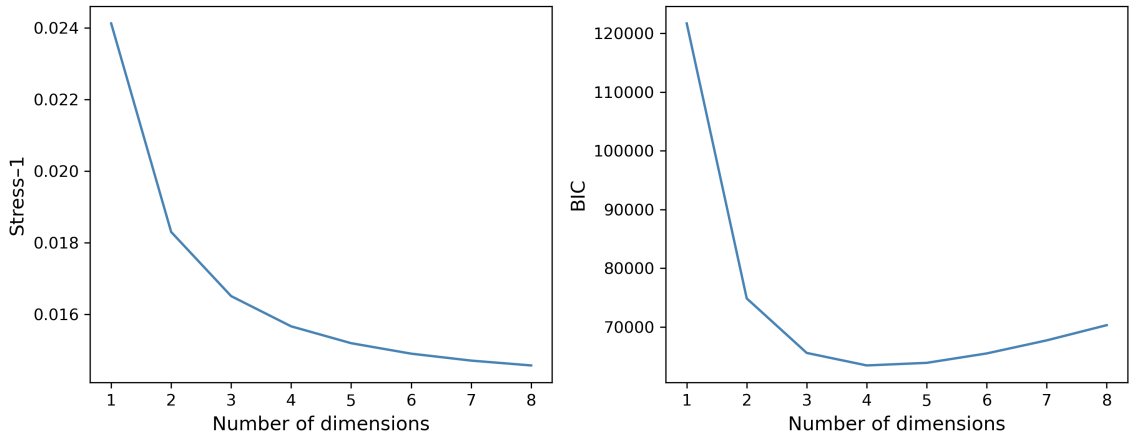
$$s = \frac{1}{n(n-1)/2} \sum_{i<j} \sqrt{\frac{\sum_k (d_{ij}^k - d_{ij})^2}{K-1}}, \tag{4}$$

$$\mathrm{BIC} = \frac{1}{s^2} \sum_{i<j} \left(d_{ij} - \hat{d}_{ij}\right)^2 + mn \log\left(\frac{n(n-1)}{2}\right). \tag{5}$$

Using the elbow criterion for the scree plot of the stress variable suggested three or four possible dimensions. The BIC curve, where the minimum was sought, suggested a four-dimensional solution (Fig. 3). The distances of all 314 available images were used to determine the number of dimensions.

Figure 3

Determining the number of dimensions based on Stress–1 and Bayesian information criterion (BIC).



3.3.2 Transfer Learning With Convolutional Neural Networks

In scenarios where artificial neural networks are used, it is common to draw on architectures that have been proven in similar use cases. We used the same model as in the Sanders and Nosofsky (2020) study: ResNet50, a type of convolutional neural network (He et al., 2016). In addition, deep learning libraries like PyTorch (Paszke et al., 2019) allow the use of pre-trained models. These models were trained on a large image set of 10 million images to classify these images into 1,000 different image categories (Deng et al., 2009). In general, such pre-trained models are able to learn new image categories faster; this process is called transfer learning (Yosinski et al., 2014). Since in our investigation we were not dealing with categories but with coordinates to be predicted by the CNNs, we had to adapt the architecture of our original model by removing the final classification layers and replacing them with linear layers that can solve such a regression problem.

Twenty-five percent of the total dataset (78 out of 314 images) was used as hold-out data to test the predictions of the ensemble. From the remaining set, another 25% (59 out of 236 images) was used for validation, leaving 177 images for training. We then trained an ensemble of 10 CNNs on our training set of spider images (Fig. 2), where the labels were their coordinates in MDS space. Like Sanders and Nosofsky (2020), we did this in two steps: First, for each of the 10 CNNs, an intermediate model was trained in which all but the new final layers were frozen, so that only the weights and biases of these new layers could change. This first training run used the Adam optimizer (Kingma & Ba, 2017) with a starting learning rate of lr = 6e-2. In a second step, each of these intermediate models was fine-tuned using the AdaGrad optimizer (Duchi et al., 2011) with a learning rate starting at lr = 2e-3. Model performance during training was measured by the accuracy of predictions in the validation set (R^2 based on the correlations between predicted and observed coordinates). For both the intermediate and the fine-tuned model we used the rate scheduler `ReduceLROnPlateau` (PyTorch; Paszke et al., 2019) which reduced the learning rate by a factor of 0.8 after six epochs without improved accuracy. In this second run, the freeze was removed and the training now affected all layers. The maximum

number of epochs for each model was set to 500. To avoid excessive overfitting, training was stopped early if there was no improvement within 20 epochs. To artificially increase the training set and further account for overfitting, random image enhancements (e.g., flip, translate, rotate, shear, blur, etc.) were applied to all training images. For each of the 10 CNNs, the model with the best accuracy was selected.

The final predictions of the entire ensemble on the test set were calculated by averaging all predictions of the resulting 10 CNNs. The Pearson correlation coefficient r (or the corresponding R^2) was then used to evaluate the predictive performance of the CNNs.

The complete code for the entire transfer learning procedure can be found in the OSF repository.

4 Results

4.1 Multidimensional Scaling Results

Figure 4 shows the two embeddings generated through MDS: the one for training the CNNs, the main embedding, and the one for testing the predictions of the CNNs. Since it was not the objective of this study to further elaborate on the dimensions and interpret them, we only briefly touched upon their properties. These four dimensions seem to represent simple visual characteristics – dimension one is somewhat related to the clarity of the spider representation (from right to left, the spider stands out more clearly and distinctly from the background), dimension two could be interpreted as the color spectrum from green (left) to orange (right). Alternatively, this dimension could be explained as the location of the spider (in green nature vs. on human skin). Weak intercorrelations developed in the training embedding, whereas in the prediction embedding dimensions one and three correlated relatively high (r = .38; see Table 1). Appendix C contains further illustrations of the individual dimension (for high resolution images we refer to the OSF repository).

Figure 4

Main MDS embedding for training the CNNs (top), and the smaller prediction set (bottom) against which the predictive performance of the CNNs is measured. The latter set was not part of the main MDS space, its coordinates were determined in an additional step based on the distances to the images in the main set, without changing the main space.

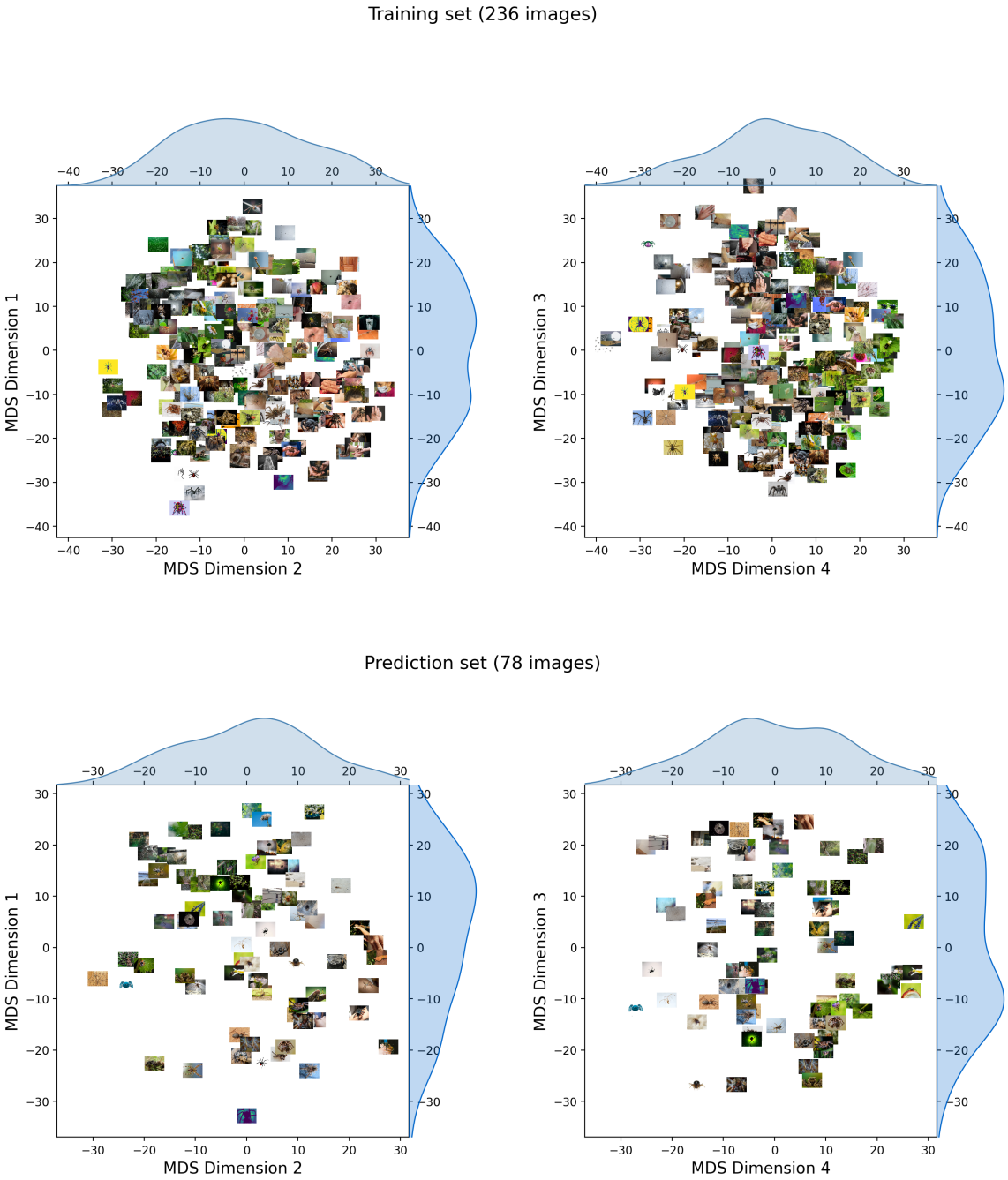


Table 1

Intercorrelations of dimensions in the original (MDS) and the predicted (CNN) embeddings with 95% CIs.

	MDS-236		MDS-78		CNN-78		CNN-1516	
Dim	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI
1–2	−.12	[−.24, .01]	−.15	[−.36, .07]	−.48	[−.63,−.29]	−.19	[−.24,−.14]
1–3	.11	[−.01, .24]	.38	[.17, .55]	.37	[.16, .55]	.56	[.53, .60]
1–4	.03	[−.10, .16]	.01	[−.21, .24]	.00	[−.22, .23]	.02	[−.03, .07]
2–3	.04	[−.09, .17]	−.03	[−.25, .19]	.03	[−.19, .25]	−.08	[−.13,−.03]
2–4	−.01	[−.14, .12]	−.14	[−.35, .09]	−.15	[−.36, .08]	−.13	[−.18,−.08]
3–4	−.09	[−.21, .04]	−.10	[−.31, .13]	−.38	[−.56,−.18]	−.11	[−.16,−.06]

4.2 Convolutional Neural Network Results

4.2.1 Training and Prediction Time

Using a standard PC workstation equipped with an NVIDIA Quadro K2200 GPU with CUDA (NVIDIA et al., 2020) enabled, it took approximately 104 minutes to train the ensemble of 10 CNNs, including intermediate and fine-tuned models. Specifically, each intermediate model took about 6 minutes to train, while each fine-tuned model took about 4 minutes to train. Additionally, it took the CNN ensemble 40 seconds to make predictions for the set of 78 images.

4.2.2 Predictions

A consistent picture emerged when we looked at the performance of the CNNs (Fig. 5). All four dimensions were predicted almost equally well, the correlations between the observed and predicted coordinates were r = .74 (R^2 = .55), r = .70 (R^2 = .49), .78 (R^2 = .61) and .73 (R^2 = .53), which was in line with expectations. The average predictive performance of our CNN ensembe was r = .74 (R^2 = .54). In addition, these results were framed by narrow 95% confidence intervals. It is also noteworthy that dimensions one and two and dimensions three and four were more highly correlated with each other than in the original MDS dimensions, suggesting that the CNNs partly determine these two dimensions from

the same image features (Table 1). The other four of the six intercorrelations were almost identical to those in the MDS embedding. The results also showed that the ensemble of CNNs consistently outperformed each individual model in terms of prediction accuracy (Table 2).

Table 2

Predictive performance of the CNN ensemble (E) and the individual models on the 78 image set.

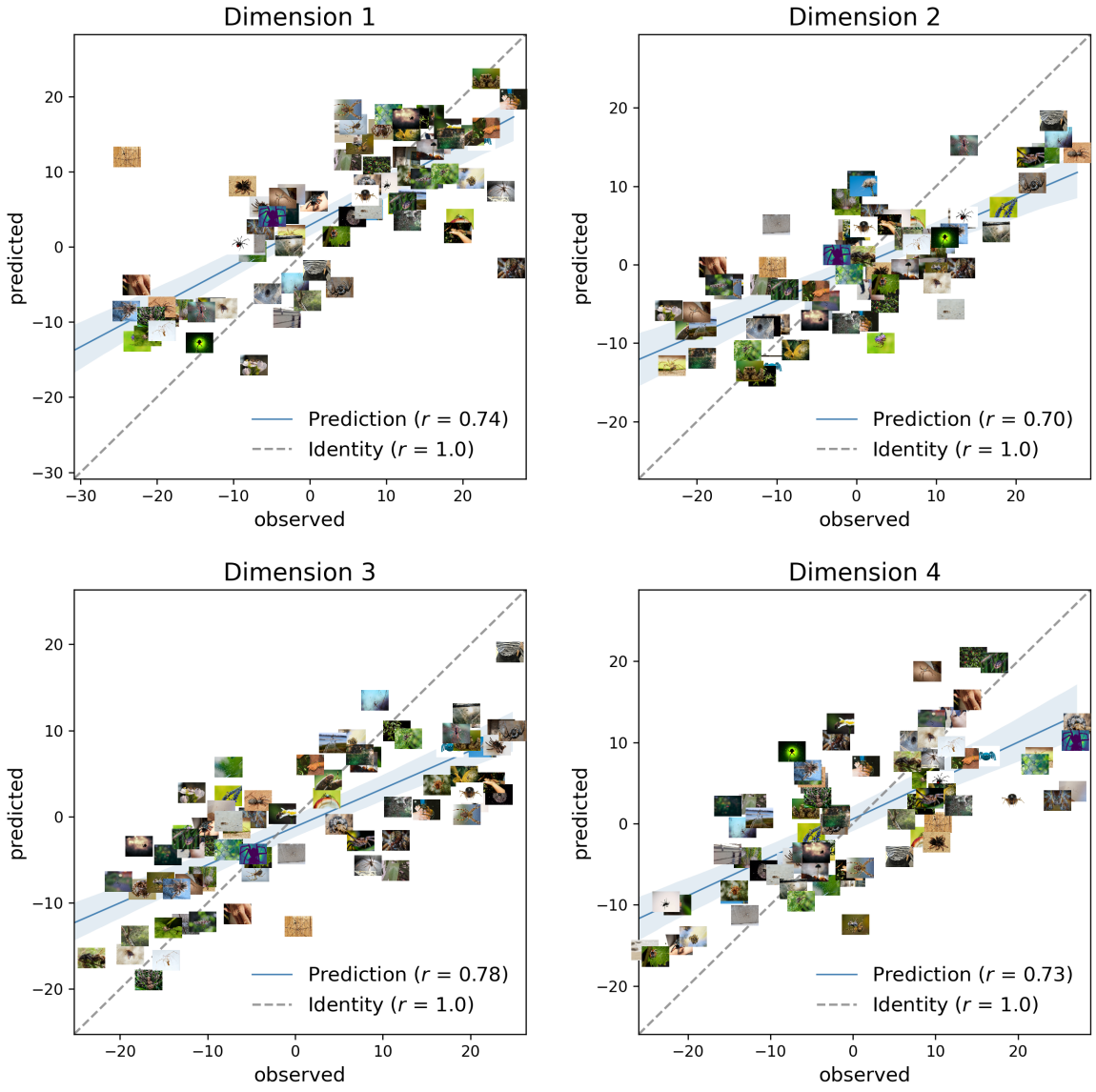
Model Dim		1	2	3	4	5	6	7	8	9	10	E
1	*r*	.71	.71	.68	.73	.66	.70	.72	.71	.65	.72	.74
	R^2	.50	.51	.46	.53	.44	.50	.51	.50	.43	.52	.55
2	.	.63	.61	.68	.69	.69	.63	.65	.65	.67	.65	.70
	.	.40	.38	.47	.47	.48	.40	.42	.43	.44	.42	.49
3	.	.72	.76	.76	.63	.74	.79	.73	.72	.73	.72	.78
	.	.52	.57	.57	.40	.55	.62	.53	.52	.53	.52	.61
4	.	.68	.70	.69	.66	.70	.66	.68	.71	.71	.71	.73
	.	.46	.49	.48	.44	.48	.43	.47	.50	.50	.50	.53
Mean	.	.69	.70	.70	.68	.70	.69	.69	.70	.69	.70	.74
	.	.47	.48	.49	.46	.49	.48	.48	.49	.47	.49	.54

4.3 Example Use Case

As already announced in the introduction, the study also included the demonstration of an example use case, namely the prediction of coordinates of a large set of 1,516 images. In real life, researchers usually do not have real coordinates at hand, which is why they would use CNNs in the first place. In the absence of such ground truth, the results for the large, novel image set could not be evaluated in terms of a numerical accuracy measure. Instead, we relied on visual inspection to demonstrate the predicted multidimensional space. Figure 6 shows the resulting dimensions, and upon examination, we found that the dimensions of the new image set were similar to the smaller original MDS, which contained 236 images. What is also worth noting is that the intercorrelations between dimensions showed a similar trend as in the original embedding for the 78 images

Figure 5

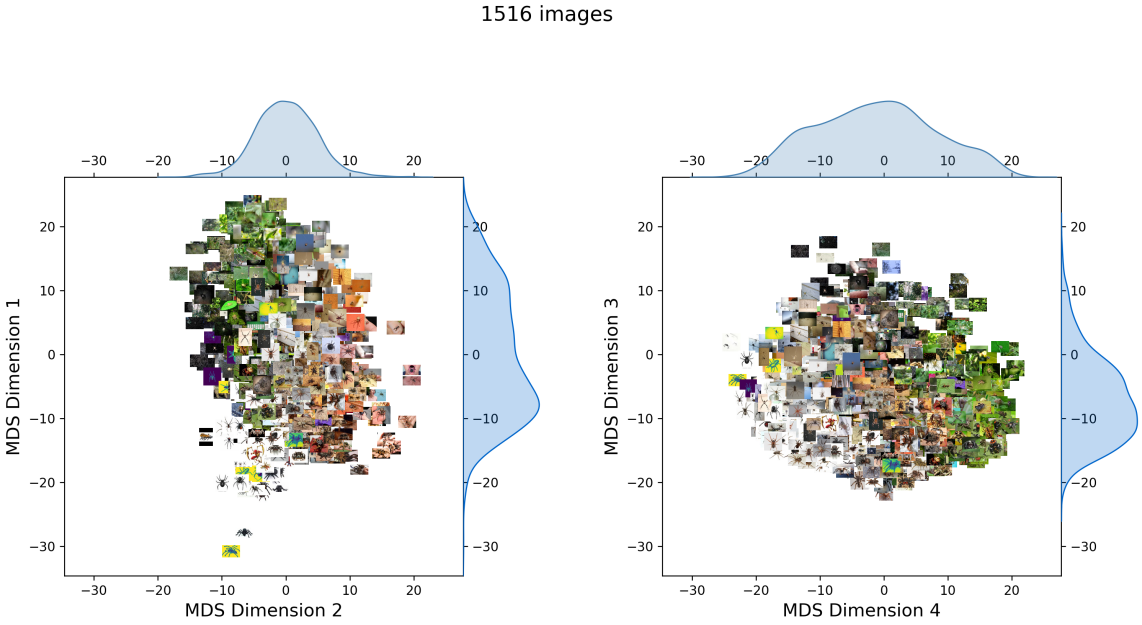
Predictive performance of CNNs on the 78 image set based on the correlations between observed and predicted dimensions with 95% confidence intervals.



set (Table 1). Embedding all 1,516 images took the CNN ensemble 13 minutes.

Figure 6

Example use case: CNN predictions for a large set of novel spider images.



5 Discussion

This study demonstrated the use of deep learning techniques to detect and classify the underlying structure of similarity judgments of fear-related images. To date, no research has approached mental representations of fear-related stimuli in this way. The dataset created as part of the study was crowdsourced from volunteers.

As expected, the results of the study were consistent with earlier work by Sanders and Nosofsky (2020) using simpler images of rocks as an example. The results showed that CNNs were successful at embedding even more complex images in multidimensional psychological spaces. Also the fact that the intercorrelations between the dimensions were almost identical in four of the six dimension pairs between the MDS embedding and the CNN-predicted embedding is another indicator of the good predictive performance of the CNNs. Notably, the performance was equally impressive across all four dimensions. Furthermore, it is worth noting that the training process can be accomplished within a

single day using commercially available hardware, and the embedding of new stimuli can be performed almost instantaneously. These results suggest that the method outlined here represents a valuable tool in psychological research.

5.1 Limitations

There are several limitations to note. First, the set of images used in the study was not systematically selected for this kind of research. Although the images were collected in a previous study, their use in the present study was not known at the time. The images represent more of a convenience sample from online sources that were readily available, but may not be representative of the broader population of images in the domain. In addition, without a more targeted selection process, the image set may be biased toward certain features or properties, further limiting its usefulness for generalization to other contexts. Furthermore, the size of the image set is smaller than that used in the previous study by Sanders and Nosofsky (2020), despite the relatively more complex images. Second, the sample size of 77 participants is rather small compared to the number of images, as each image pair has only one to two distance ratings. In addition, participants were not screened based on arachnophobia scores, but were recruited through convenience sampling. This may limit the generalizability of the results to populations with higher levels of arachnophobia. Another important point is that it is highly likely that for certain images individual CNNs make conflicting predictions, but this form of uncertainty has not been accounted for. As with the predictions for all other images, the average was formed from such conflicting predictions, which can negatively affect the result for otherwise very good predictions. Such images could, for example, be withheld and presented again to participants to check their position in multidimensional space or to further investigate the reasons for poor prediction performance. Finally, we used only one random split to create the training and prediction sets, and only one random seed for the initial MDS configuration. More precise accuracy measures could be obtained by cross-validation, for example, using different splits in the training and prediction data, using different seeds for the initial MDS configurations, and thus training the networks on different MDS solutions.

However, this could become intractable as the number of different data splits and MDS configurations increases, since an entire ensemble of CNNs would have to be computed for each of these combinations. Because this study is only a proof of concept of whether or not a given MDS configuration can be predicted from a set of novel images (and in real-world scenarios we would be confronted with an existing MDS space with specific dimensions), at least the lack of cross-validation of the different seeds for MDS initializations is less of an issue.

5.2 Added Value of This Study

The main advantage of this study is the generation of complete MDS spaces with a pre-trained ensemble of CNNs for complex, fear-related images. Furthermore, new stimuli can be easily embedded into existing MDS spaces without requiring time-consuming similarity judgments from participants. In principle, any number of new stimuli can be embedded into existing MDS spaces with little effort. Previous research has only demonstrated this for simple, non-valenced images. This could be advantageous in future exposure therapy settings for assessing where in this multidimensional space the most fear-related stimuli are for a given patient. Further steps, such as determining how similar patients are, can then be used to determine which stimuli are most effective for a particular patient and ensure the best success rate for anxiety therapy.

5.3 Future Directions

In addition to addressing some of the limitations of this study, there are now several future directions that could be explored in further research. A major motivation for this study was to aid in exposure therapy for anxiety. To further this goal, an important future direction could be to investigate how the positions of stimuli in the multidimensional space map onto measures of fear, disgust, and willingness to approach, potentially in an arachnophobic population. In addition, it would be valuable to explore how the stimuli can be sequentially selected from this space in a personalized manner to optimize

exposure therapy outcomes. To further study spider images one approach that may be worth exploring is to remove the surrounding context and present the spider on a plain background, similar to how rocks were studied by Sanders and Nosofsky (2020). This might help participants focus more on the spider's intrinsic features when making similarity judgments, as they would be less distracted by the background. In addition, such an approach could help identify the influence of context on participants' judgments. To validate the predictive performance of the CNNs, a straightforward task would be to collect similarity judgments for a subset of the generalization image set used in our example use case. To improve the predictive performance of CNNs, it could help to determine the certainty with which CNNs can predict the position of objects, in order to avoid accepting inaccurate predictions. To ensure that the results of the present study are not solely due to the current experimental conditions, it would be important to replicate the current experiment using a different setup, such as the choose-m-of-n-stimuli task used by Roads and Mozer (2019). To speed up data collection in the future, an active learning approach could be used by prioritizing trials that contain the most information for the initial MDS space. This approach can make data collection more economical and significantly reduce the total number of trials required to create MDS spaces (Roads & Mozer, 2019). Finally, our study does not answer the question about the nature of the similarities between our fear-related images and whether they differ from similarities between non-fear-related images. In other words, it is unclear whether the similarities are based on the fear-related nature of the images or just on similarities of images in general. Future studies should investigate this question further.

5.4 Conclusion

The present study successfully demonstrated the ability to reconstruct MDS spaces of complex, fear-related images using CNNs, consistent with previous work on simple images. This finding suggests that CNNs are a promising method for studying fear-related stimuli and their internal representations. However, certain caveats, such as the experimental design and the lack of uncertainty measures in the predictions, need to

be addressed to fully exploit this approach and improve the practicality of working with fear-related images. Overall, further exploration of CNNs in understanding fear-related stimuli is warranted and may lead to significant advances in the field.

6 References

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335. https://doi.org/10.31887/DCNS.2015.17.3/bbandelow

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H.-U. (2017). Anxiety disorders. *Nature Reviews. Disease Primers*, *3*, 17024. https://doi.org/10.1038/nrdp.2017.24

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dewis, L. M., Kirkby, K. C., Martin, F., Daniels, B. A., Gilroy, L. J., & Menzies, R. G. (2001). Computer-aided vicarious exposure versus live graded exposure for spider phobia in children. *Journal of Behavior Therapy and Experimental Psychiatry*, *32*(1), 17–27. https://doi.org/10.1016/S0005-7916(01)00019-2

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(61), 2121–2159. http://jmlr.org/papers/v12/duchi11a.html

Fehm, L., Pelissolo, A., Furmark, T., & Wittchen, H.-U. (2005). Size and burden of social phobia in Europe. *European Neuropsychopharmacology*, *15*(4), 453–462. https://doi.org/10.1016/j.euroneuro.2005.04.002

Fredrikson, M., Annas, P., Fischer, H., & Wik, G. (1996). Gender and age differences in the prevalence of specific fears and phobias. *Behaviour Research and Therapy*, *34*(1), 33–39. https://doi.org/10.1016/0005-7967(95)00048-3

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, *4*(11), 1173–1185. https://doi.org/10.1038/s41562-020-00951-3

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256–281. https://doi.org/10.1037/a0028860

Kingma, D. P., & Ba, J. (2017, January 29). *Adam: A method for stochastic optimization*. arXiv: 1412.6980 [cs]. Retrieved April 1, 2023, from http://arxiv.org/abs/1412.6980

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, *45*(1), 149–166. https://doi.org/10.1006/jmps.1999.1300

NVIDIA, Vingelmann, P., & Fitzek, F. H. (2020). CUDA, release: 10.2.89. https://developer.nvidia.com/cuda-toolkit

Oosterink, F. M. D., De Jongh, A., & Hoogstraten, J. (2009). Prevalence of dental fear and phobia relative to other fear and phobia subtypes. *European Journal of Oral Sciences*, *117*(2), 135–143. https://doi.org/10.1111/j.1600-0722.2008.00602.x

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019, December 3). *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. arXiv: 1912.01703 [cs, stat]. Retrieved November 19, 2022, from http://arxiv.org/abs/1912.01703

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

Piercey, C. D., Charlton, K., & Callewaert, C. (2012). Reducing anxiety using self-help virtual reality cognitive behavioral therapy. *Games for Health Journal*, *1*(2), 124–128. https://doi.org/10.1089/g4h.2012.0008

Polák, J., Rádlová, S., Janovcová, M., Flegr, J., Landová, E., & Frynta, D. (2020). Scary and nasty beasts: Self [U+2010] reported fear and disgust of common phobic animals. *British Journal of Psychology*, *111*(2), 297–321. https://doi.org/10.1111/bjop.12409

Roads, B. D., & Mozer, M. C. (2019). Obtaining psychological embeddings through joint kernel and metric learning. *Behavior Research Methods*, *51*(5), 2180–2193. https://doi.org/10.3758/s13428-019-01285-3

Sanders, C. A., & Nosofsky, R. M. (2020). Training deep networks to construct a psychological feature space for a natural-object category domain. *Computational Brain & Behavior*, *3*(3), 229–251. https://doi.org/10.1007/s42113-020-00073-z

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, *1*(1), 1–17. https://doi.org/10.1016/0010-0285(70)90002-2

Torgerson, W. S. (1967). *Theory and methods of scaling* (7. print). Wiley.

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review*, *28*(6), 1021–1037. https://doi.org/10.1016/j.cpr.2008.02.007

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in neural information processing systems*, 3320–3328. https://doi.org/10.48550/arXiv.1411.1792

A Deutsche Zusammenfassung

Angstbezogene visuelle Stimuli werden häufig in der Konfrontationstherapie eingesetzt, einer gängigen Methode zur Behandlung von Angststörungen. Um Angststörungen besser zu verstehen, kann es hilfreich sein zu wissen, wie Personen die dazugehörigen Reize mental repräsentieren. Häufig werden Ähnlichkeitsbewertungen von solchen Reizen verwendet, um mentale Repräsentationen als multidimensionalen Raum mathematisch darzustellen. Diese zu sammeln, ist eine zeitaufwändige, oft unlösbare Aufgabe. Hier wurde am konkreten Beispiel der Spinnenphobie untersucht, ob solche multidimensionalen Repräsentationen von künstlichen neuronalen Netzen erzeugt werden können. Bisherige Forschung hat diesen Ansatz nur für Bilder von simplen Objekten verwendet. In der vorliegenden Studie haben 77 Teilnehmer Ähnlichkeitsbewertungen von Spinnenbildern abgegeben. Mittels multidimensionaler Skalierung (MDS) wurde ein mehrdimensionaler Raum der Bilder erstellt, in welchem Unähnlichkeiten durch Distanzen dargestellt waren. Anschließend wurde ein Ensemble künstlicher neuronaler Netze (Convolutional Neural Networks; CNNs) trainiert, um diese mehrdimensionalen Repräsentationen zu reproduzieren. Alle vier resultierenden MDS-Dimensionen konnten von den CNNs erfolgreich vorhergesagt werden. Die Ergebnisse zeigen, dass CNNs mehrdimensionale Repräsentationen komplexer, angstbezogener Bilder generieren können. Offen bleibt, ob diese Reize in klinischen Populationen anders repräsentiert werden als in der Allgemeinbevölkerung. Weitere Einschränkungen wie geringe Stichprobengröße und ein unsystematisch zusammengestelltes Bilderset könnten in zukünftigen Studien überwunden werden. Die Methode könnte jedenfalls richtungsweisende Erkenntnisse für die Entwicklung von Angsttherapien bringen, wenn man sich zukünftig den Fragen widmet, wo in diesem multidimensionalen Raum jene Bilder liegen, die am meisten mit Angst assoziiert sind, und wie man aus diesem Raum Bilder für Konfrontationstherapie auswählen kann, um für die jeweilige Person den optimalen Therapieerfolg zu erzielen.

Stichworte: Spinnenphobie, Angststörung, Mentale Repräsentationen, Künstliche neuronale Netze, Multidimensionale Skalierung, Konfrontationstherapie

B Code Examples

Computing MDS With Python Package scikit-learn

Input

```python
from sklearn.manifold import MDS
import numpy as np


# load distance matrix
D = np.loadtxt("data/MDS/D.txt")

# number of dimensions
n_dims = 4

# create MDS instance
mds = MDS(
    n_components=n_dims,
    dissimilarity="precomputed",
    max_iter=5000,
    eps=1e-9,
    normalized_stress=False,
    random_state=1,
)

# create embedding
emb_train = mds.fit_transform(D)

# compute normalized stress
normalized_stress = np.sqrt(mds.stress_ / (n_dims * np.sum(D**2)))

# print coordinates for first 5 images
print(emb_train[:5], "\n")

# print normalized stress
print("Stress-1:", normalized_stress)
```

```
[[-13.70181093 -16.91064281   2.71926094  10.97214829]
 [ -7.61645155  -2.92077816  18.72345587 -16.85966161]
 [ -9.48634174  -9.72497882 -35.7147028   -7.0947897 ]
 [  5.97309733   2.37160332   6.44072848 -22.4689668 ]
 [-22.14430258  -1.63495331  10.22555141  13.92709506]]


Stress-1: 0.13878798881393278
```

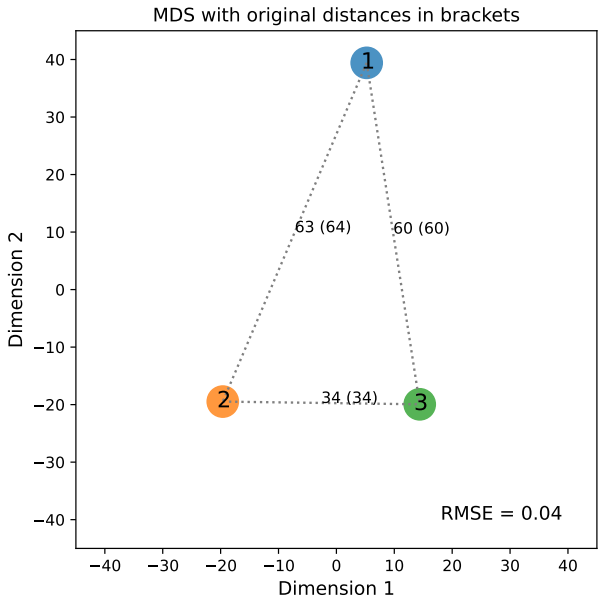
MDS Prediction for Novel Stimuli

We did not find a dedicated method in the literature or in established Python libraries to embed new points in existing MDS spaces without changing the latter. The method we found suitable, finding the optimal position via least squares of the distance error, is briefly described here with a code example.

Creating the Initial Embedding

Input

```python
import numpy as np
import matplotlib.pyplot as plt
from sklearn.manifold import MDS
from scipy.spatial.distance import pdist, squareform
from itertools import combinations

# Distance matrix
D = np.array([
    [0.0, 64.0, 60.0, 29.0],
    [64.0, 0.0, 34.0, 23.0],
    [60.0, 34.0, 0.0, 84.0],
    [29.0, 23.0, 84.0, 0.0],
])

# Create MDS
mds = MDS(n_components=2,
          dissimilarity="precomputed",
          normalized_stress=False,
          random_state=1)
# Fit MDS
D_part = D[:-1, :-1]
emb = mds.fit_transform(D_part)
D_emb = squareform(pdist(emb))


# Plot MDS
def plot_mds(emb, D):
    plt.figure(figsize=(6, 6))
    D_emb = squareform(pdist(emb))
    # dots
    for i, (x, y) in enumerate(emb):
        plt.scatter(x, y, s=400, alpha=0.8)
        plt.text(x - 1, y - 1, i + 1, fontsize="x-large")

    # lines
    combs = list(combinations(range(len(emb)), 2))
    for i, j in combs:
        arr = np.vstack([emb[i], emb[j]]).T
        plt.plot(arr[0], arr[1], linestyle=":", color="gray")
        plt.text(
            arr.mean(axis=1)[0],
            arr.mean(axis=1)[1],
            f"{int(D_emb[i, j])} ({int(D[i, j])})",
            fontsize="medium",
        )

    plt.xlim((-45, 45))
    plt.ylim((-45, 45))
    plt.xlabel("Dimension 1", fontsize="large")
    plt.ylabel("Dimension 2", fontsize="large")
    plt.title("MDS with original distances in brackets")
    rmse = np.sqrt(np.mean((np.tril(D_emb) - np.tril(D))**2))
    plt.text(
        18,
        -40,
        f"RMSE = {round(rmse,2)}",
        fontsize="large",
    )


plot_mds(emb, D_part)

fname = "mds_1.pdf"
plt.savefig(fname)
fname
```

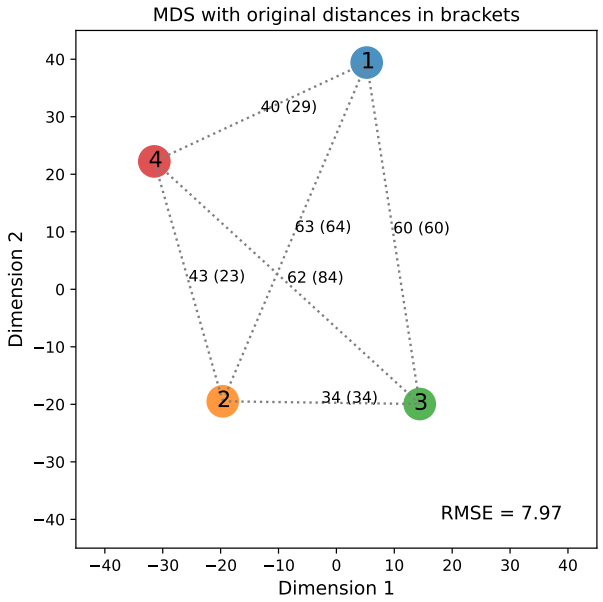


Embedding a New Point Using Least Squares

Input

```python
import numpy as np
from scipy.optimize import least_squares


def predict_points(distances, points):
    """Predict new points to a set of existing points while minimizing distance errors.

    Parameters
    ----------
    distances : array-like of shape (m, n_points)
        The distances from the new points to each of the existing points.
    points : array-like of shape (n_points, n_dims)
        The coordinates of the existing points.

    Returns
    -------
    p_new : ndarray of shape (n_dims,)
        The coordinates of the new point that minimize the sum of squared distance errors.

    Raises
    ------
    AssertionError
        If the number of distances does not match the number of points.

    Notes
    -----
    This function uses the least squares method to find the coordinates of a new point that minimize
    the distance errors between the new point and a set of existing points. The number of dimensions
    is inferred from the input data.

    Examples
    --------
    >>> distances = [[1.5, 1.0, 1.2, 1.1]]
    >>> points = [[0, 0, 0], [1, 0, 0], [0, 1, 0], [0, 0, 1]]
    >>> predict_points(distances, points)
    array([[0.5, 0.5, 0.5]])
    """

```

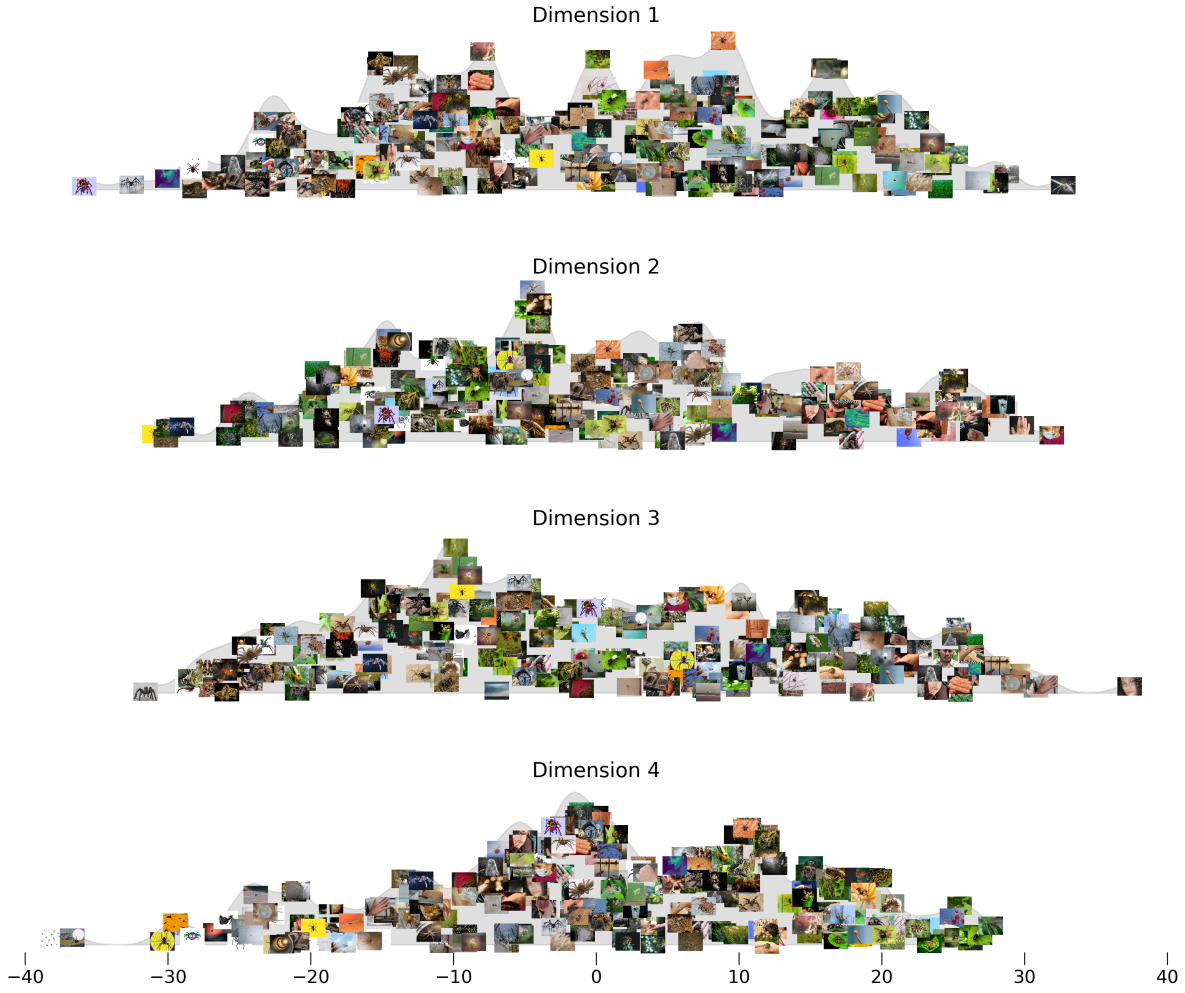
```
    dims = points.shape[1]
    n_points = points.shape[0]
    m = distances.shape[0]

    assert (
        distances.shape[1] == n_points
    ), "Number of distances must match the number of points"

    def objective(x):
        p_new = np.reshape(x, (dims, 1))
        errors = []
        for i in range(len(d)):
            p_existing = np.reshape(points[i], (dims, 1))
            error = np.linalg.norm(p_new - p_existing) - d[i]
            errors.append(error)
        return errors

    points_new = np.zeros(shape=(m, dims))
    for i, d in enumerate(distances):
        x0 = np.zeros(dims)
        res = least_squares(objective, x0)
        points_new[i] = res.x

    return points_new


lstsqr_coords = predict_points(np.expand_dims(D[-1, :-1], axis=0), emb)


plot_mds(np.vstack([emb, lstsqr_coords[0]]), D)

fname = 'mds_2.pdf'
plt.savefig(fname)
fname
```



C Dimension Inspection

Figure C.1

Density plots of the spider images along the four dimensions.



D List of Figures and Tables

List of Figures

List of Tables