Grounding Psychological Similarity Spaces in Deep Neural Networks

Lucas Bechberger, Institute of Cognitive Science, Osnabrück University, lbechberger@uos.de

Psychological similarity spaces are a useful tool for analyzing similarity judgments and categorization behavior made by humans. They are usually based on pairwise dissimilarity ratings (obtained in psychological experiments), which are processed with multidimensional scaling (MDS) [4] in order represent dissimilarity between stimuli as distances between points. A crucial shortcoming of this approach is its inability to generalize to novel stimuli [1]. Deep neural networks (DNNs) are a powerful machine learning tool and have become immensely popular in recent years due to their good generalization performance, especially in the area of computer vision [7]. The activation of the final layers of such networks are generally regarded as high-level representations of the raw input.

In recent years, there has been a growing interest in combining deep neural networks with cognitive models [1]: The high-level representations of DNNs are used as a substitute for human similarity spaces, allowing researchers to evaluate the predictions of similarity measures or categorization models on a large number of naturalistic inputs. Also a mapping from images to points in psychological similarity spaces has been investigated in this context [2, 8] (see Figure 1): First, a similarity space is obtained with MDS based on human similarity ratings, and then DNNs are used to learn a mapping from raw inputs to coordinates in this space. In order to prevent overfitting, the network is also trained on a secondary task (such as classification or reconstruction) using additional data.

We present a practical study on the domain of shapes, involoving a data set of 60 line drawings with pairwise visual dissimilarity ratings [3].[1] This focus on the cognitive domain of shapes distinguishes our work from the earlier study by Sanders and Nosofsky [8], who consider holistic similarity spaces involving a combination of multiple cognitive domains such as shape, texture, and color. Table 1 shows generalization performance of various machine learning approaches as measured by R^2 with respect to a four-dimensional target space. We observe that sketches (using the Sketch-a-Net architecture [11] on the TU Berlin [6] and Sketchy [9] data sets) are a more useful source domain than photographs (using the inception-v3 architecture [10] on ImageNet [5]). Moreover, multi-task learning (where coordinates are predicted as part of the network's penultimate layer and both tasks are optimized jointly) led to better performance than transfer learning (in the form of a lasso regression on top of a pre-trained network). Furthermore, sketch classification seems to be a more useful secondary task than sketch reconstruction. The overall performance level reached in our experiments is considerably below the level of $R^2 \approx 0.77$ reported by Sanders and Nosofsky [8] for a data set of 360 images, which indicates that a larger set of stimuli and more complex network architectures may be needed in future studies.

[1]Code available at https://github.com/lbechberger/LearningPsychologicalSpaces.

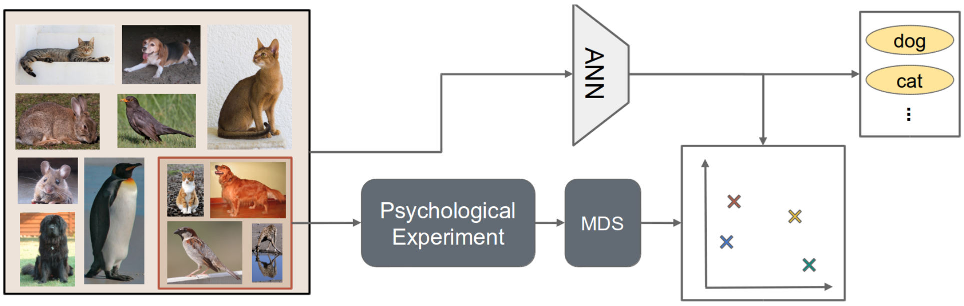

Figure 1: Illustration of the general hybrid approach from [2].

Source Domain	Secondary Task	Learning Regime	R^2
Photographs	Classification	Transfer	0.4924
Sketches	Classification	Transfer	0.5246
		Multi-Task	0.5775
	Reconstruction	Transfer	0.2605
		Multi-Task	0.4213

Table 1: Overview of our best regression results for various setups.

References

[1] Ruairidh M. Battleday, Joshua C. Peterson, and Thomas L. Griffiths. From Convolutional Neural Networks to Models of Higher-Level Cognition (and Back Again). *Annals of the New York Academy of Sciences*, 2021.

[2] Lucas Bechberger and Kai-Uwe Kühnberger. *Generalizing Psychological Similarity Spaces to Unseen Stimuli*, pages 11–36. Springer International Publishing, Cham, 2021.

[3] Lucas Bechberger and Margit Scheibel. Analyzing Psychological Similarity Spaces for Shapes. In *ICCS*, pages 204–207, Cham, 2020.

[4] I. Borg and J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer Series in Statistics. Springer-Verlag New York, 2nd edition, 2005.

[5] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, pages 248–255, 2009.

[6] Mathias Eitz, James Hays, and Marc Alexa. How Do Humans Sketch Objects? *ACM Trans. Graph.*, 31(4), July 2012.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *CVPR*, June 2016.

[8] Craig A Sanders and Robert M Nosofsky. Using Deep-Learning Representations of Complex Natural Stimuli as Input to Psychological Models of Classification. In *CogSci*, 2018.

[9] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies. *ACM Trans. Graph.*, 35(4), July 2016.

[10] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In *CVPR*, pages 2818–2826, 2016.

[11] Qian Yu, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy Hospedales. Sketch-a-Net that Beats Humans. In *BMVC*, pages 7.1–7.12. BMVA Press, 2015.