Using Convolutional Neural Networks to Map Line Drawings Into Psychological Shape Space

Lucas Bechberger[0000−0002−1962−1777] and Kai-Uwe Kühnberger[0000−0003−1626−0598]

Institute of Cognitive Science, Osnabrück University
{lucas.bechberger,kai-uwe.kuehnberger}@uni-osnabrueck.de

The interdisciplinary framework of conceptual spaces [6] proposes a geometric representation of conceptual knowledge through low-dimensional interpretable similarity spaces. The similarity spaces used in this framework are often based on psychological similarity ratings for a small set of stimuli, and are unfortunately often incapable of generalizing to novel stimuli.

Since recent studies [4,8] have found a link between the internal representations used by deep convolutional neural networks (CNNs) and human cognitive processes, we have proposed to use such neural networks for learning a mapping from raw images into the psychologically grounded similarity spaces [1].

In this study, we provide an example application of this idea, focusing on the cognitive domain of shapes, which is deemed very important for the recognition and classification of physical objects [7]. We compare the mapping performance of photograph-based and sketch-based CNNs on a data set of line drawings [2,3], exploring both transfer learning and multi-task learning settings.[1].

1 General Methods

We use the shapes data set by Bechberger and Scheibel [2,3], which contains 60 line drawings that are annotated with their coordinates in similarity spaces of varying dimensionality. As secondary data source, we employed the TU Berlin data set [5] and the Sketchy data set [10], which contain free-hand sketches annotated with their respective class. We used several data augmentation techniques (resizing, translation, rotation, horizontal flips, and shearing) to increase the variety of our training examples.

In all of our experiments, we applied 10% salt and pepper noise during training, but no noise on the test set. Due to the small number of line drawings, we make use of a five-fold cross validation scheme. We report the averaged results across all five trained networks. We use the coefficient of determination R^2 as an evaluation metric.

2 Experiments

For our transfer learning experiment, we considered the pre-trained photograph-based inception-v3 network [11] (using an internal representation with 2048 di-

[1] Code for reproducing our results is publicly available on GitHub: https://github.com/lbechberger/LearningPsychologicalSpaces.

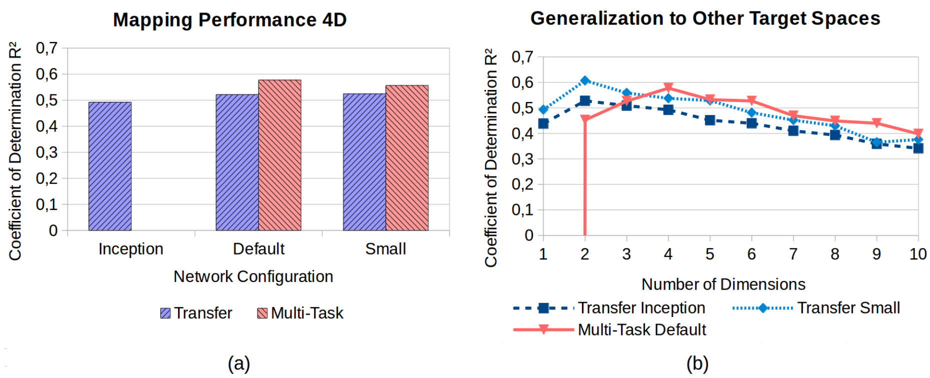

Fig. 1. Results of our mapping experiments for a four-dimensional target space (a) and target spaces of different dimensionality (b).

mensions) and a variant of Sketch-a-Net [12] (the first CNN to reach human level performance on the TU Berlin data set) which was trained from scratch to correctly classify the augmented sketches of both TU Berlin and Sketchy. We consider two variants of our sketch-based network: DEFAULT (with hyperparameters set to the values reported in the literature) and SMALL (with identical hyperparameters but a representation size of 256 instead of 512 dimensions). For each network configuration, we extracted the internal representation of all augmented line drawings and trained a lasso regression on the resulting feature space. Figure 1a shows that transfer learning from our sketch-based network was slightly more successful than transfer learning from the photograph-based network. The best results we were able to achieve with this approach are $R^2 \approx 0.52$.

For our multi-task learning experiment, we only considered the two variants of our sketch-based network. We supplemented the classification loss with an additional incentive to use a dedicated part of the internal representation to predict the coordinates of the image in the similarity space for all line drawings. As we can see in Figure 1a, performance is considerably better than for the transfer learning setup with results of up to $R^2 \approx 0.58$

In a final experimental step, we applied the different approaches to target spaces of varying dimensionality, without further optimizing their hyperparameter settings. Figure 1b shows that the sketch-based network can maintain its advantage over the photograph-based network in the transfer task across all target spaces. Moreover, for the transfer task, a two-dimensional similarity space yields the best performance, indicating a good trade-off between representational capacity and compactness. The multi-task learning approach outperforms the transfer learning setup on all spaces with at least four dimensions, but collapses for lower-dimensional spaces. One may however speculate that we could improve performance on the low-dimensional target spaces through additional hyperparameter tuning.

3 Conclusion

Our experiments show that mapping images into psychological shape space with CNNs is possible. We found advantages of sketch-based over photograph-based networks and of multi-task learning over transfer learning. Our overall performance level is still too low for practical applications and falls behind other related work, where levels of $R^2 \approx 0.77$ have been reported [9] (using however a larger data set and a more complex network structure). Thus, further research is needed to arrive at a more stable mapping function.

References

1. Bechberger, L., Kühnberger, K.U.: Generalizing Psychological Similarity Spaces to Unseen Stimuli (2020)
2. Bechberger, L., Scheibel, M.: Analyzing Psychological Similarity Spaces for Shapes. In: Alam, M., Braun, T., Yun, B. (eds.) Ontologies and Concepts in Mind and Machine. pp. 204–207. Springer International Publishing, Cham (2020)
3. Bechberger, L., Scheibel, M.: Representing Complex Shapes with Conceptual Spaces. In: Second International Workshop 'Concepts in Action: Representation, Learning, and Application' (CARLA 2020) (2020), `https://openreview.net/forum?id=OhFQNQicgXy`
4. Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A.: Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific reports **6**, 27755 (2016)
5. Eitz, M., Hays, J., Alexa, M.: How Do Humans Sketch Objects? ACM Trans. Graph. **31**(4) (Jul 2012), `https://doi.org/10.1145/2185520.2185540`
6. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press (2000)
7. Landau, B., Smith, L., Jones, S.: Object Perception and Object Naming in Early Development. Trends in Cognitive Sciences **2**(1), 19 – 24 (1998), `https://doi.org/10.1016/S1364-6613(97)01111-X`
8. Peterson, J.C., Abbott, J.T., Griffiths, T.L.: Evaluating (and Improving) the Correspondence Between Deep Neural Networks and Human Representations. Cognitive Science **42**(8), 2648–2669 (2018)
9. Sanders, C.A., Nosofsky, R.M.: Using Deep-Learning Representations of Complex Natural Stimuli as Input to Psychological Models of Classification. In: Proceedings of the 2018 Conference of the Cognitive Science Society, Madison. (2018)
10. Sangkloy, P., Burnell, N., Ham, C., Hays, J.: The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies. ACM Trans. Graph. **35**(4) (Jul 2016), `https://doi.org/10.1145/2897824.2925954`
11. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2818–2826 (2016)
12. Yu, Q., Yang, Y., Song, Y.Z., Xiang, T., Hospedales, T.: Sketch-a-Net that Beats Humans. In: Xianghua Xie, M.W.J., Tam, G.K.L. (eds.) Proceedings of the British Machine Vision Conference (BMVC). pp. 7.1–7.12. BMVA Press (September 2015), `https://dx.doi.org/10.5244/C.29.7`