Unlearnability of Symmetry-based Visual Relations by Deep Neural Networks

Guillermo Puebla[1] **Jorge Díaz-Ramírez**[2] **Domingo Araya**[3] **Gonzalo Fuentes**[3]

Jeffrey S. Bowers[4]

[1] Escuela de Administración y Negocios, Universidad de Tarapacá
[2] Departamento de Ingeniería y Tecnologías, Universidad de Tarapacá
[3] Instituto de Ingeniería Matemática y Computacional, Pontificia Universidad Católica de Chile
[4] School of Psychological Science, University of Bristol

Abstract

Relational reasoning, a fundamental ability of the human mind, remains an important challenge for visual deep neural network (DNN) systems. In this work, we present a new set of visual relational reasoning tasks with varying levels of relational complexity and show that, for a series of convolutional neural networks and vision transformers, a third-order relational task is not learnable—in the sense that extensive hyperparameter search and model training do not yield in-distribution above-chance performance on the test set. Our results highlight the limitations of current visual DNN architectures and suggest a path towards developing reliable methods to discriminate between human and visual DNN system responses.

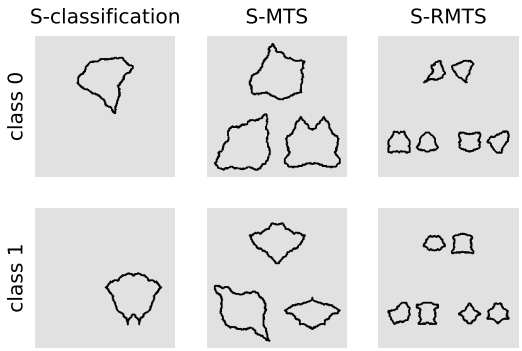

Figure 1: Symmetry-based relational tasks.

Introduction

With a single glance, the human visual system is capable of producing scene comprehension: a parsing of the current state of affairs into a set of objects comprising the scene and the *relations* among these entities (Biederman, 1981). Consequently, the ability of visual deep neural networks (DNNs)—which can match, and sometimes even surpass, human performance on tasks such as naturalistic image classification—to represent visual relations has been the focus of many investigations. In particular, many studies have examined whether visual DNNs can learn to represent the *sameness* relation (Adeli et al., 2023; Funke et al., 2021; Gupta et al., 2025; Kim et al., 2018; Messina et al., 2021, 2022; Puebla & Bowers, 2022, 2025; Stabinger et al., 2016; Tartaglini et al., 2025). These studies have relied heavily on the same-different task, which consist of classifying an image with two objects as “same” or “different”.

In this work, we extend the study of visual relations in DNNs by introducing a new set of tasks that leverage *bilateral symmetry*. This type of symmetry is central to biological vision systems, as most animals exhibit it (Pizlo et al., 2014). Importantly, bilateral symmetry is not a concrete feature of an object; rather, it is better defined as an inner *relation* between the two sides of an object. In contrast to previous research that implemented higher-order visual relational reasoning tasks based on the sameness relation (Puebla & Bowers, 2025), using bilateral symmetry as the basis for our tasks enabled us to instantiate third-order visual relational reasoning tasks with only six objects at a time.

Simulations

Tasks

Our symmetry-based tasks extend the sameness tasks used in studies of primate reasoning defined by Premack (1983) (see Figure 1). All our tasks are binary classification problems. We termed the first one *symmetry classification* (henceforth, S-classification). In this task, a given sample is labelled as 0 if the shape is asymmetric and as 1 if it is symmetric. S-classification is a first-order visual relational reasoning task because the reasoner only needs to make a judgment about an unnested relation.

We refer to our second task as *symmetry-based match-to-sample* (henceforth, S-MTS). In this task, there is a base shape at the top-center of the image, which is either symmetric or asymmetric, and two candidate shapes, one at the bottom-left and the other at the bottom-right. One of the candidates is symmetric, and the other asymmetric, and the goal is to choose the shape at the bottom that has the same symmetry inner relation as the base shape

Extended abstract presented at the *9th Conference on Cognitive Computational Neuroscience*, New York, NY, USA, 2026.

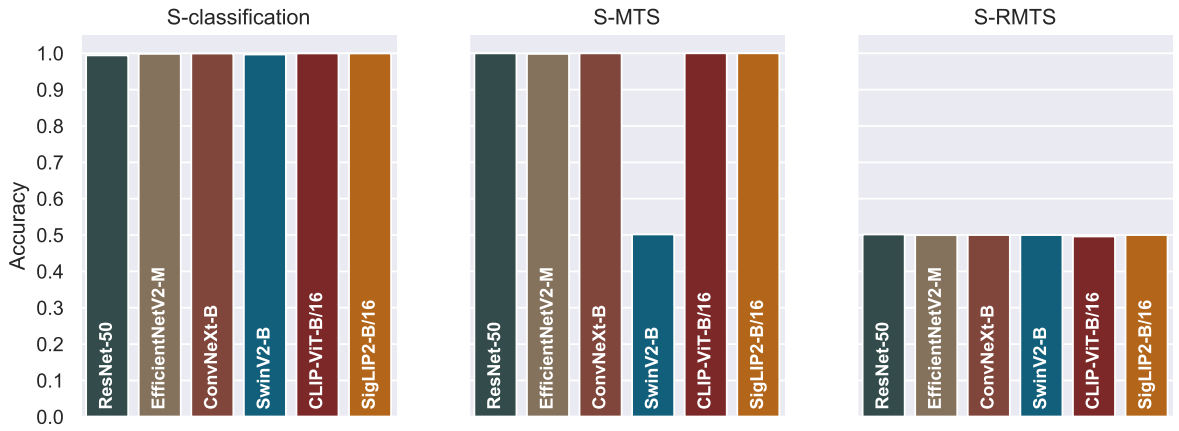

Figure 2: In-distribution test accuracy after hyperparameter optimization and training.

at the top; the label is 0 if the matching shape is at the bottom-left and 1 if it is at the bottom-right. S-MTS is a second-order visual relational reasoning task because the reasoner needs to make a nested judgment about the individual symmetry inner relations of the objects.

We termed our final task *symmetry-based relational match-to-same* (henceforth, S-RMTS). In this task there is a base pair of shapes at the top-center of the image instantiating the (second order) relation *same-symmetry* (when both shapes are symmetric or both are asymmetric) or *different-symmetry* (when one shape is symmetric and the other asymmetric) and two candidate pairs of shapes at the bottom-left and bottom-right. The goal is to identify which candidate pair matches the base pair's symmetry relation; the label is 0 if the matching pair is at the bottom-left and 1 if it is at the bottom-right. S-RMTS is a third-order visual relational reasoning task because the reasoner needs to make a nested judgment about the second-order symmetry relations between pairs of objects.

Models

We included six models in our simulations. Three convolutional neural networks: ResNet-50 (He et al., 2016), EfficientNetV2-M (Tan & Le, 2021), ConvNeXt-B (Liu et al., 2022b); and three vision transformers: SwinV2-B (Liu et al., 2022a), CLIP-ViT-B/16 (Radford et al., 2021), SigLIP2-B/16 (Tschannen et al., 2025).

Datasets

For the S-classification and S-MTS tasks, the datasets consisted of disjoint sets of 28000, 5600, and 11200 images for training, validation, and testing, respectively. For the S-RMTS task, we increased the training set size to 280000 after observing that no model could learn the task with the original training size. For each model, we generated datasets matching its preferred image size.

Trainig and Test

Before training each model, we optimized hyperparameters using the Optuna framework (Akiba et al., 2019). For all models, we used the AdamW optimizer with an exponential learning rate scheduler and a batch size of 128. With these settings, we optimized the learning rate and the scheduler's multiplicative decay factor over 100 trials. We then selected the best-performing training configuration based on in-distribution validation accuracy and trained a single model instance with those hyperparameters for up to 100 epochs. We report test accuracy for those particular seeds.

Results

Figure 2 presents the in-distribution test accuracy after hyperparameter optimization and training. As can be seen, all models achieved ceiling test performance on the S-classification task. On the S-MTS task, all models achieved ceiling performance, except SwinV2-B, which exhibited random-level performance. Finally, all models exhibited random-level test performance in the S-RMTS task.

Discussion

These results provide strong evidence that contemporary visual deep neural networks, including both convolutional architectures and vision transformers, fail to acquire higher-order relational representations when these are defined over abstract properties such as bilateral symmetry. While all models readily achieved ceiling performance on first- and second-order tasks, their consistent collapse to chance-level accuracy on the third-order S-RMTS task—even after extensive hyperparameter optimization and substantial increases in training data—raises the possibility of a fundamental limitation rather than a contingent failure of training.

Acknowledgments

This work was supported by the Agencia Nacional de Investigación y Desarrollo (ANID) through FONDECYT de Iniciación Grant No. 11241282.

References

Adeli, H., Ahn, S., & Zelinsky, G. J. (2023). A brain-inspired object-based attention network for multiobject recognition and visual reasoning. *Journal of Vision*, *23*(5), 16–16.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Lawrence Erlbaum Associates.

Funke, C. M., Borowski, J., Stosio, K., Brendel, W., Wallis, T. S., & Bethge, M. (2021). Five points to check when comparing visual perception in humans and machines. *Journal of Vision*, *21*(3), 16–16.

Gupta, M., Rane, S., McCoy, R. T., & Griffiths, T. L. (2025). Convolutional neural networks can (meta-) learn the same-different relation. *arXiv preprint arXiv:2503.23212*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Kim, J., Ricci, M., & Serre, T. (2018). Not-so-clevr: Learning same–different relations strains feedforward neural networks. *Interface focus*, *8*(4), 20180011.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., et al. (2022a). Swin transformer v2: Scaling up capacity and resolution. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12009–12019.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022b). A convnet for the 2020s. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11976–11986.

Messina, N., Amato, G., Carrara, F., Gennaro, C., & Falchi, F. (2021). Solving the same-different task with convolutional neural networks. *Pattern Recognition Letters*, *143*, 75–80.

Messina, N., Amato, G., Carrara, F., Gennaro, C., & Falchi, F. (2022). Recurrent vision transformer for solving visual reasoning problems. *Image Analysis and Processing–ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part III*, 50–61.

Pizlo, Z., Li, Y., Sawada, T., & Steinman, R. M. (2014). *Making a machine that sees like us*. Oxford University Press.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences*, *6*(1), 125–136.

Puebla, G., & Bowers, J. S. (2022). Can deep convolutional neural networks support relational reasoning in the same-different task? *Journal of Vision*, *22*(10), 11–11.

Puebla, G., & Bowers, J. S. (2025). Visual reasoning in object-centric deep neural networks: A comparative cognition approach. *Neural Networks*, *189*, 107582.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. *International conference on machine learning*, 8748–8763.

Stabinger, S., Rodríguez-Sánchez, A., & Piater, J. (2016). 25 years of CNNs: Can we compare to human abstraction capabilities? *International conference on artificial neural networks*, 380–387.

Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. *International conference on machine learning*, 10096–10106.

Tartaglini, A. R., Feucht, S., Lepori, M. A., Vong, W. K., Lovering, C., Lake, B., & Pavlick, E. (2025). Deep neural networks can learn generalizable same-different visual relations. *8th Annual Conference on Cognitive Computational Neuroscience*.

Tschannen, M., Gritsenko, A., Wang, X., Naeem, M. F., Alabdulmohsin, I., Parthasarathy, N., Evans, T., Beyer, L., Xia, Y., Mustafa, B., et al. (2025). Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *arXiv preprint arXiv:2502.14786*.