

# Artificial Intelligence Systems for Solving IQ Tests

Mikołaj Małkiński

thesis supervisor: Prof. Jacek Mańdziuk, Ph.D., D.Sc.

Warsaw University of Technology, Faculty of Mathematics and Information Science

## Introduction

Intelligence can be decomposed into plenty of constituting factors such as pattern recognition, ability to perform arithmetic operations or **reasoning about abstract concepts**. The ability to solve abstract reasoning tasks is considered one of the hallmarks of human intelligence, which can be roughly estimated by measuring subjects' performance on **IQ tests**. We focus on a specific version of such problems—the **Raven's Progressive Matrices** (RPMs)—presented in Figure 1.

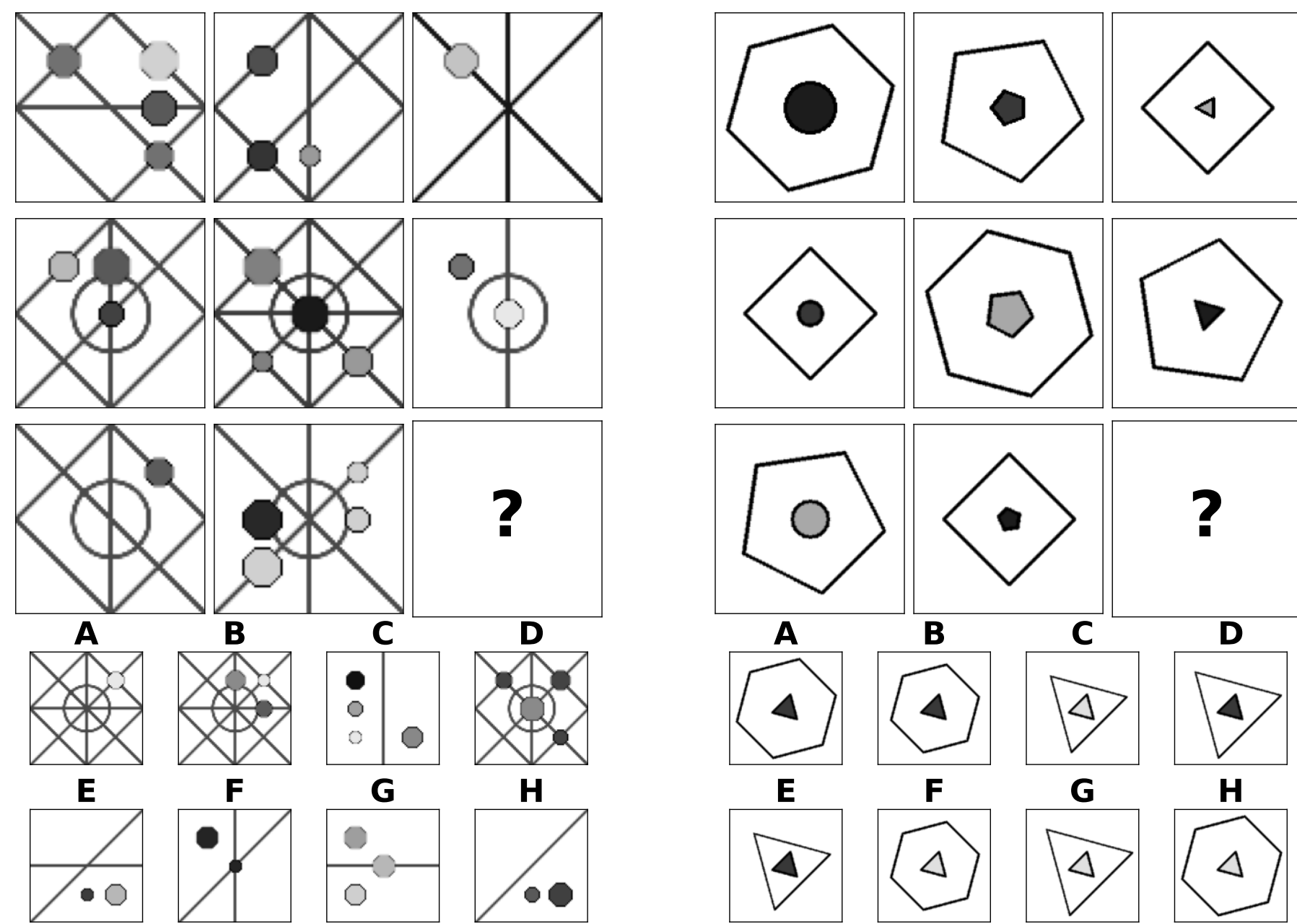


Figure 1. Solving RPMs requires to identify abstract relationships hidden behind random visual distractors and contrast possible answers to select the one which fits best. Left RPM, although has more visual details, is governed by only a single rule (row-wise AND applied to shape position), whereas perceptually simpler right RPM contains 8 distinct relations applied to both outer and inner structures. The examples come from the **PGM** [1] and **Balanced-RAVEN** [7] datasets respectively. In both cases the correct answer—which is to be placed in the bottom-right panel—is A.

## Augmentation of RPMs

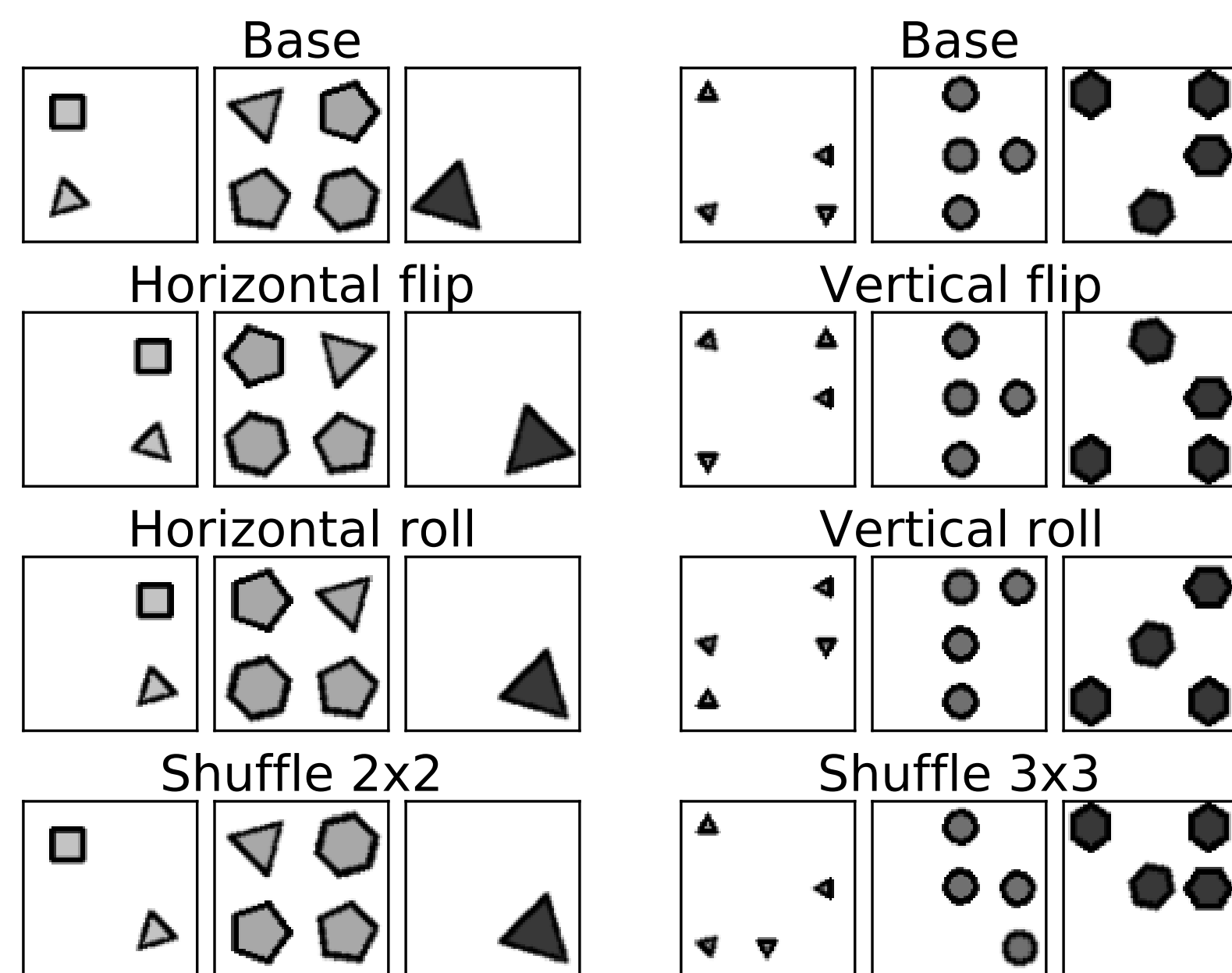


Figure 2. Augmentation of RPMs from the Balanced-RAVEN dataset. Selected transformation is applied in the same way to all images in a given RPM. The data augmentation module applies a randomly selected combination of presented methods with additional random rotation and transposition. For clarity, we only depict different views of a single row for two RPMs belonging to configurations **2x2Grid** (left part) and **3x3Grid** (right part), respectively.

## Neural modules

In order to better understand which neural network components are useful for solving abstract relational reasoning problems, we propose and evaluate **11 new architectures** based on:

- **HriNet** [2]: 1) we replace the original perceptual ResNet backbone with a simpler CNN; 2) incorporate Global Reasoning units (GloRe); 3) introduce a novel rotation-based *Vortex Convolution* module; combine with a variety of reasoning modules including 4) Neural Arithmetic Logic Units, 5) Relation Networks, 6) Predinet, 7) Multi-Head Attention and 8) Area Attention,
- **WReN** [1]: 9) we extend the relational component to handle object triplets,
- **SCL** [6]: 10) we propose a generalized non-linear scattering transformation which considers all available group assignments and 11) a multi-head scattering operation.

## Learning to solve RPMs

We have considered the following methodologies of learning to solve RPMs:

- **Supervised** (CE) - trains the network to correctly predict index of the correct answer, by estimating the true probability distribution  $p = \text{onehot}(k)$  (see *Scoring* branch in Fig. 3),
- **Auxiliary** (AUX) - in this setup, the model has to predict an encoded representation of the abstract rules (see *Rule prediction* branch in Fig. 3). For this purpose we propose a novel **sparse encoding** [4], which is superior in all experiments on Balanced-RAVEN to previously used *dense encoding* [1, 7].
- **Multi-Label Contrastive Learning** (MLCL) - a novel method of learning to solve RPMs which casts the problem into a **multi-label classification framework**, employs the proposed sparse encoding and introduces a **contrasting mechanism** in the objective function (see Fig. 3).

## Multi-Label Contrastive Learning

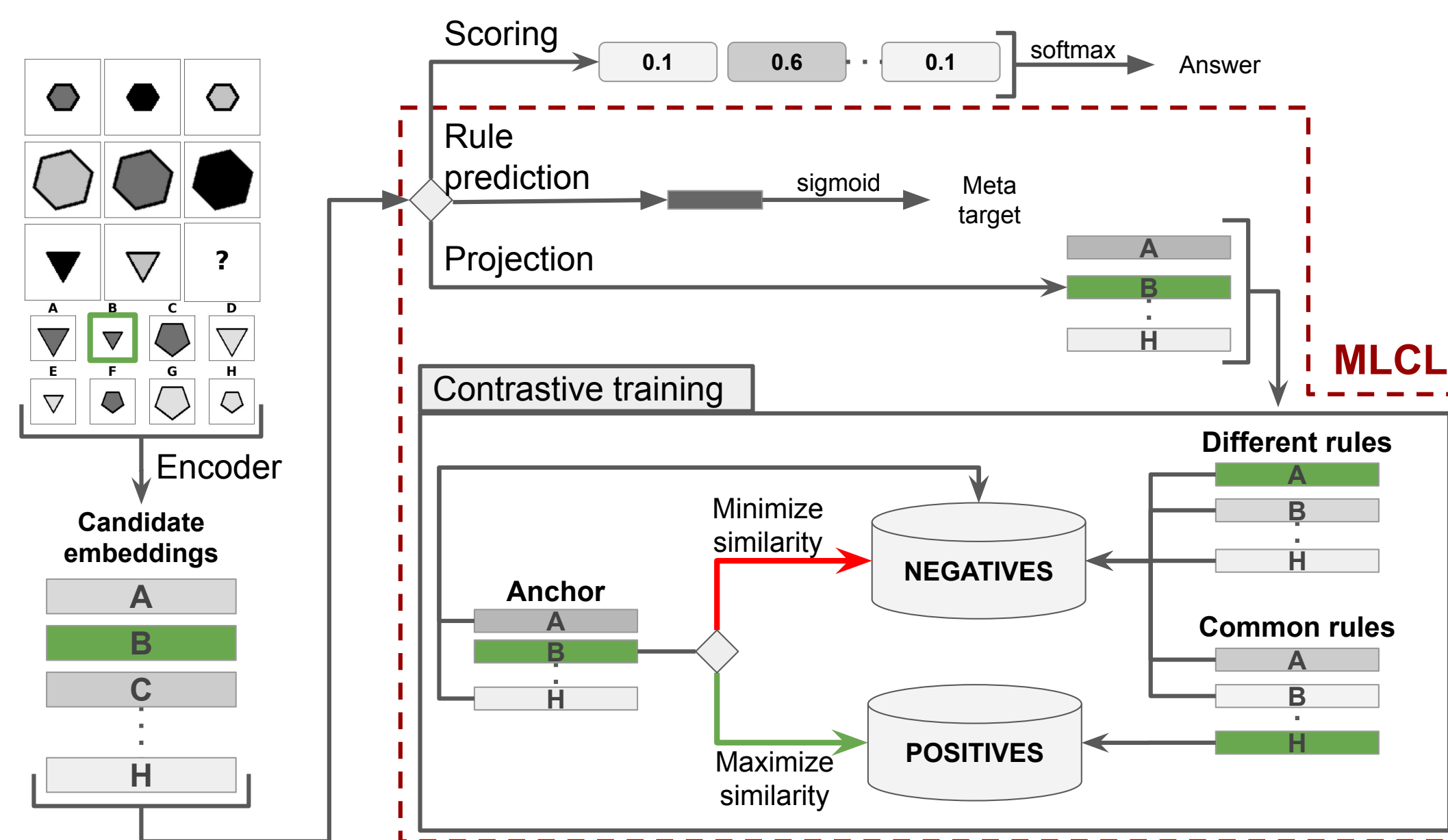


Figure 3. In each training setup we complete the context matrix with each of the choice panels and generate *candidate embeddings*  $h$  using an encoder network  $\mathcal{N}$ . The candidate embedding corresponding to the answer which correctly completes the matrix is depicted in green. The embeddings are processed via three different paths: 1) a scoring head  $\psi$  predicts the probability distribution  $\hat{p}$  of answers, 2) a rule prediction network  $\rho$  estimates the underlying rules  $\hat{\pi}$  as encoded meta-targets and 3) a contrastive projection network  $g$  transforms them into lower-dimensional vectors  $z$ , which are later used for the purpose of computing the contrastive loss. Both rule prediction and contrastive projection branches constitute the proposed MLCL framework.

## Multi-Label Supervised Contrastive Loss

We aim to encourage abstract reasoning models to recognise the underlying RPM rules, which are necessary to identify the correct answer. Our extension of the Supervised Contrastive Loss [3] for multi-label data points  $\mathcal{L}^{\text{mlc}}$  minimizes the similarity of the anchor to negative samples (incorrectly completed RPMs and correctly completed RPMs with different rules) and maximizes the similarity to positive pairs (correctly completed RPMs with common rules) (see Fig. 3):

$$\mathcal{L}^{\text{mlc}} = \sum_{i=1}^{2N} \frac{1}{2N_{\tilde{Y}_i} - 1} \sum_{j=1}^{2N} \mathbb{1}_{i \neq j} \cdot \mathbb{1}_{\tilde{Y}_i \cap \tilde{Y}_j \neq \emptyset} \cdot \mathcal{L}_{i,j}^{\text{mlc}}$$
$$\mathcal{L}_{i,j}^{\text{mlc}} = -\log \frac{\exp(z_i \cdot z_j / \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp(z_i \cdot z_k / \tau) + \mathcal{L}_{i,j,k}^{\text{mlc}}}$$
$$\mathcal{L}_{i,j,k}^{\text{mlc}} = \sum_{l=1}^7 \exp(z_i \cdot z'_{k,l} / \tau)$$

## Main results

Table 1. Test accuracy on Balanced-RAVEN averaged across 4 random seeds. Each model achieves best results when trained with the proposed MLCL approach with data augmentation (DA). Most notably, SCL trained with MLCL and DA sets a **new state-of-the-art** result with an absolute improvement of 1.8 p.p. over best result from the literature [6].

Method	DA	Mean test accuracy $\pm$ Std (%)		
		SCL	HriNet	CoPINet
CE	×	82.8 $\pm$ 0.7	50.6 $\pm$ 1.3	44.8 $\pm$ 0.8
	✓	89.7 $\pm$ 1.0	56.7 $\pm$ 1.6	47.2 $\pm$ 3.4
AUX dense	×	86.9 $\pm$ 3.5	51.6 $\pm$ 1.5	32.2 $\pm$ 1.9
	✓	91.4 $\pm$ 2.9	56.7 $\pm$ 1.2	33.6 $\pm$ 2.4
AUX sparse	×	95.9 $\pm$ 0.3	57.0 $\pm$ 1.6	46.8 $\pm$ 2.3
	✓	96.0 $\pm$ 0.9	64.2 $\pm$ 1.2	46.3 $\pm$ 1.4
MLCL	×	95.7 $\pm$ 0.5	56.2 $\pm$ 1.5	57.1 $\pm$ 0.4
	✓	<b>96.8 <math>\pm</math> 0.4</b>	<b>65.7 <math>\pm</math> 2.4</b>	<b>58.0 <math>\pm</math> 0.4</b>

## References

- [1] David G. T. Barrett, Felix Hill, Adam Santoro, Ari S. Morcos, and Timothy P. Lillicrap. Measuring abstract reasoning in neural networks. *International Conference on Machine Learning (ICML)*, 2018.
- [2] Sheng Hu, Yuqing Ma, Xianglong Liu, Yanlu Wei, and Shihao Bai. Hierarchical Rule Induction Network for Abstract Visual Reasoning. *arXiv preprint arXiv:2002.06838*, 2020.
- [3] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. Supervised Contrastive Learning. *arXiv preprint arXiv:2004.11362*, 2020.
- [4] Mikołaj Małkiński and Jacek Mańdziuk. Multi-Label Contrastive Learning for Abstract Visual Reasoning. *arXiv preprint arXiv:2012.01944*, 2020.
- [5] Jacek Mańdziuk and Adam Żychowski. DeepIQ: A Human-Inspired AI System for Solving IQ Test Problems. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.
- [6] Yuhui Wu, Honghua Dong, Roger Grosse, and Jimmy Ba. The Scattering Compositional Learner: Discovering Objects, Attributes, Relationships in Analogical Reasoning. *arXiv preprint arXiv:2007.04212*, 2020.
- [7] Chi Zhang, Feng Gao, Baoxiong Jia, Yixin Zhu, and Song-Chun Zhu. Raven: A dataset for relational and analogical visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5317–5327, 2019.
- [8] Chi Zhang, Baoxiong Jia, Feng Gao, Yixin Zhu, Hongjing Lu, and Song-Chun Zhu. Learning perceptual inference by contrasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1075–1087, 2019.