TESTING NEURAL NETWORK INTELLIGENCE USING RAVEN'S PROGRESSIVE MATRICES

Nina Raganová

Supervisor: prof. Ing. Igor Farkaš, Dr.

Motivation

- AI mimics cognitive functions
- Compare with human performance

- "An intelligence test sometimes shows a man how smart he would have been not to have taken it."

  - *Laurence J. Peter*

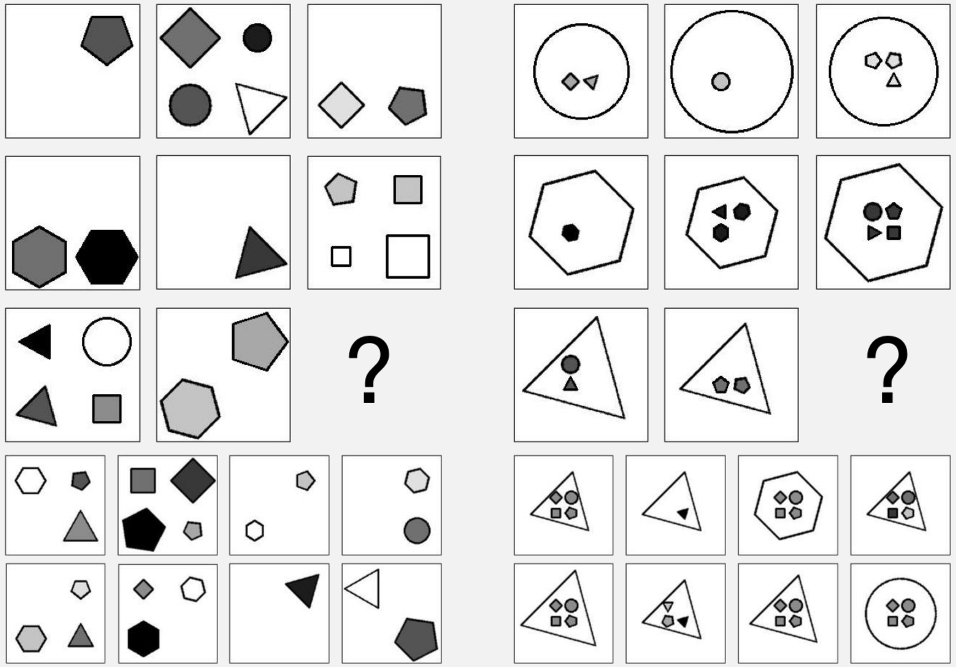
GOAL

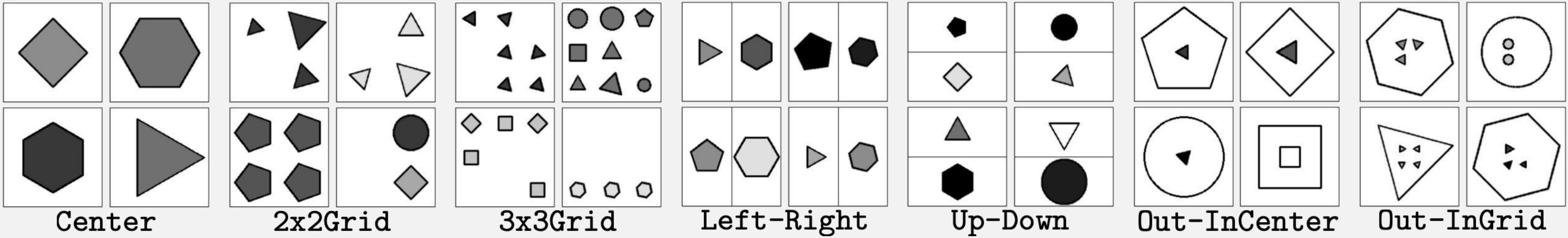
- Implement a chosen neural network and train it for solving the RPM problem.
- Test the model performance in various scenarios concerning the variability and the size of the training set.
- Interpret obtained results.

RAVEN'S PROGRESSIVE MATRICES

- Measuring abstract reasoning
- Multiple choice questions
- Visual geometric design

- RAVEN dataset (Zhang, 2019)

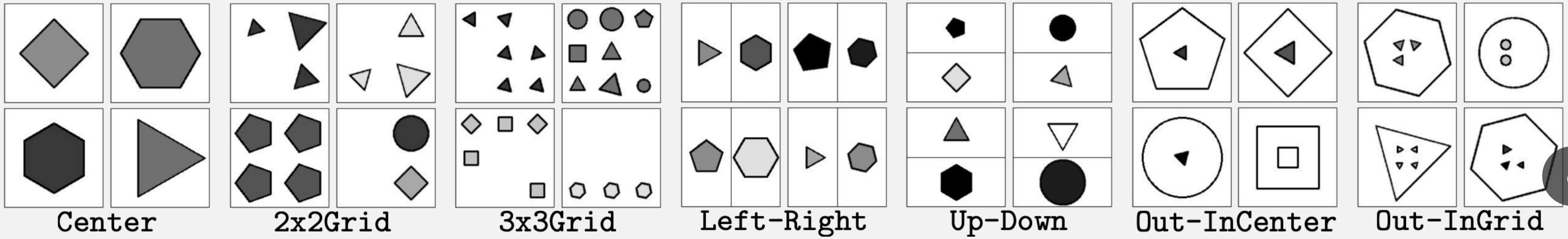



Center
2x2Grid
3x3Grid
Left-Right
Up-Down
Out-InCenter
Out-InGrid

Recent Work

- Solving Raven's Progressive Matrices with Neural Networks (Zhuo and Kankanhalli, 2020)
  - ResNets with ImageNet pre-training

Model	Average	Center	2x2Grid	3x3Grid	L-R	U-D	O-IC	O-IG
ResNet-18	77.18%	72.75%	57.00%	62.65%	91.00%	89.60%	88.40%	78.85%
Human	84.41%	95.45%	81.82%	79.55%	86.36%	81.81%	86.36%	81.81%



Experimental Setup

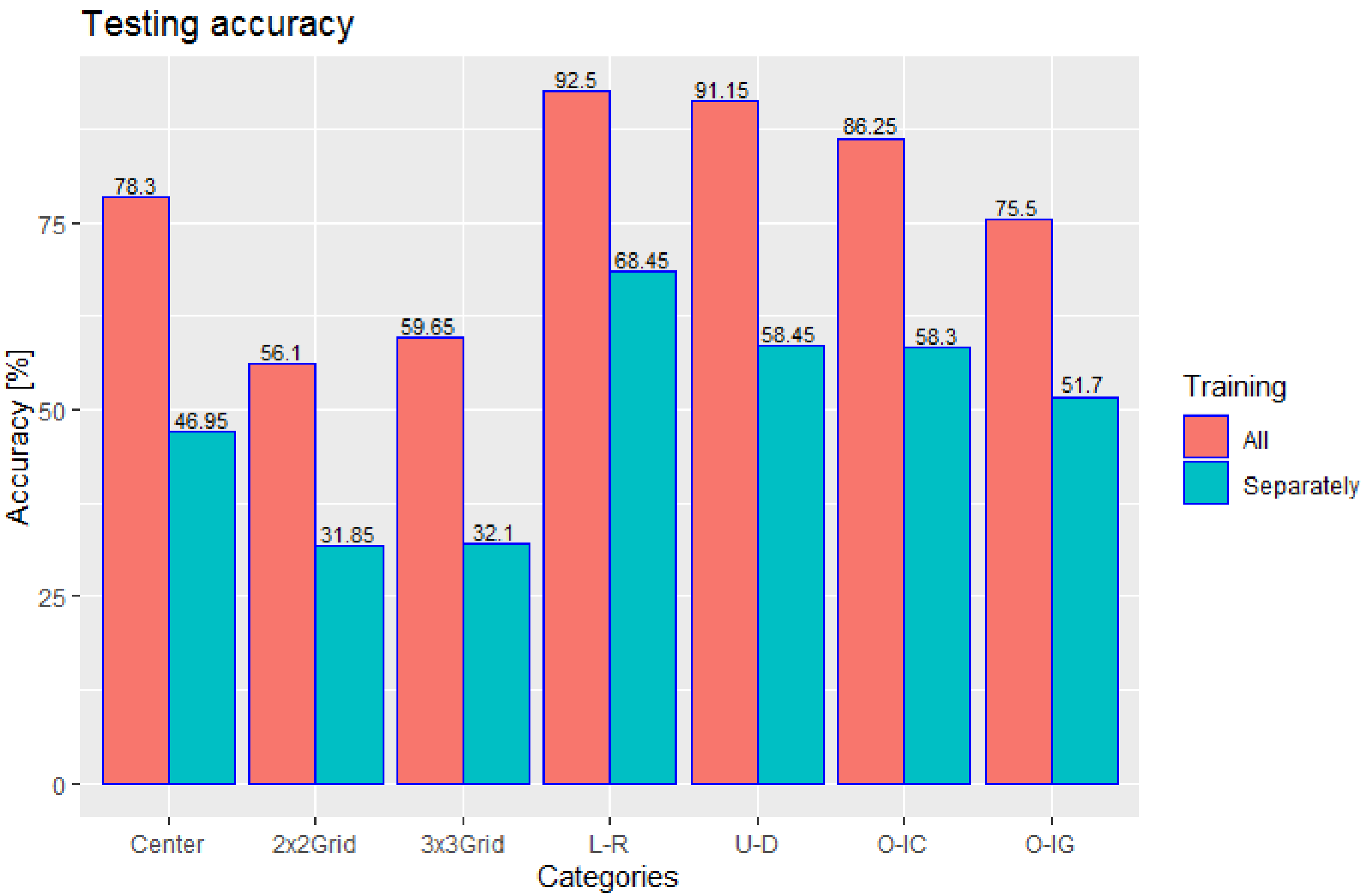
- PyTorch library
- ResNet-18 with extra dropout layer
- Adam optimizer[1]
- Dataset split 6000 : 2000 : 2000
  (training : validation : testing)
- 16-dimensional input (16 x 224 x 224)
- output layer with 8 neurons and softmax

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	7×7, 64, stride 2
conv2_x	$56 \times 56 \times 64$	3×3 max pool, stride 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	7×7 average pool
fully connected	1000	512×1000 fully connections
softmax	1000	

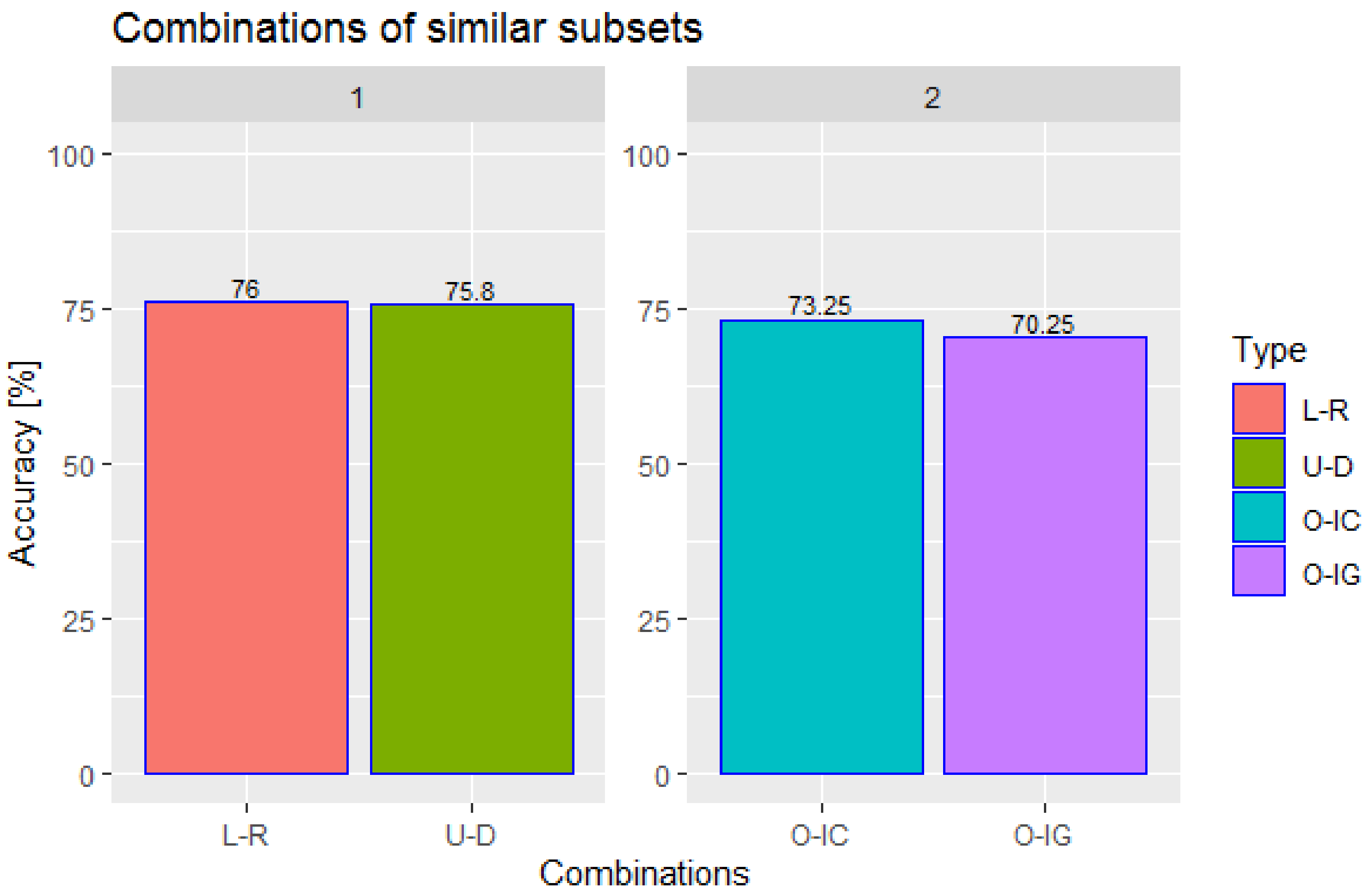
1. Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv–1412.

RESULTS

Joint and Separate Training

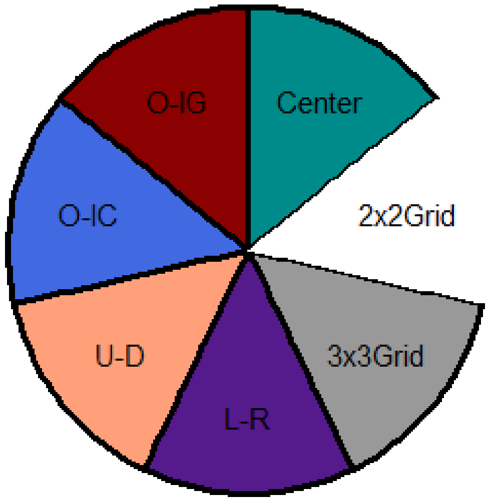


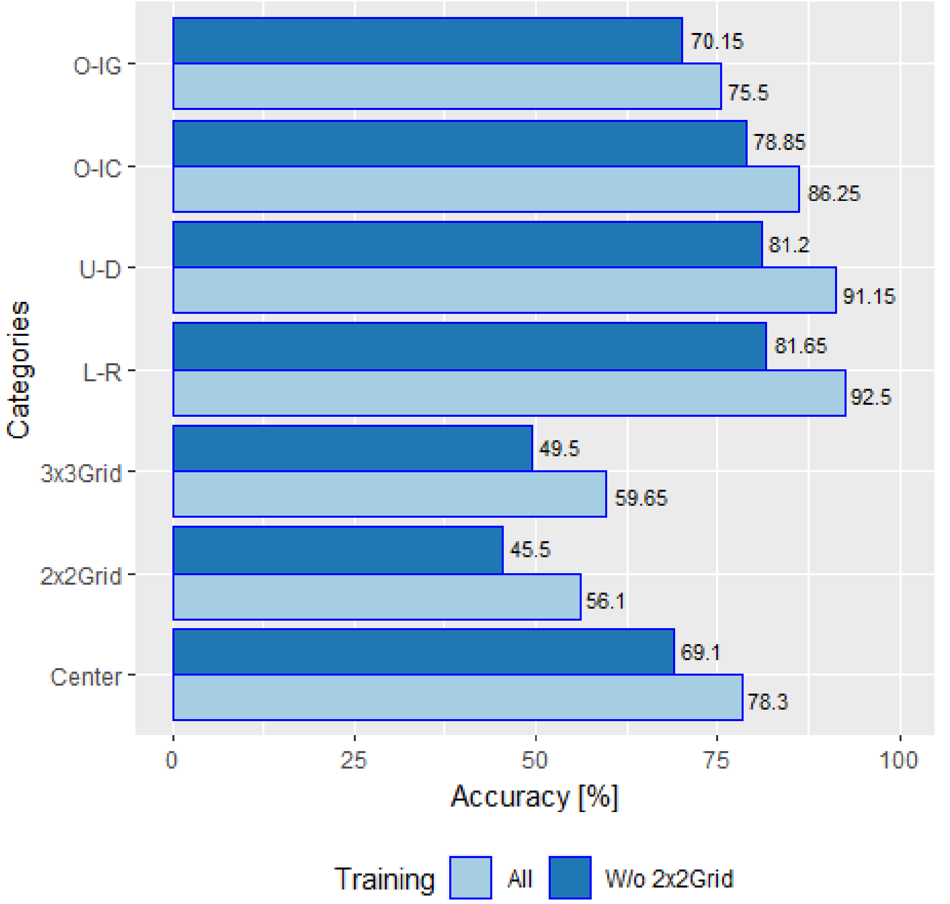
Combination of 2 Similar Subsets



Trained without 2x2Grid

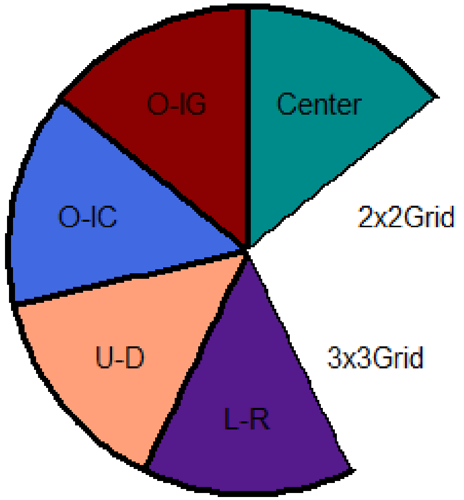
Training w/o 2x2Grid

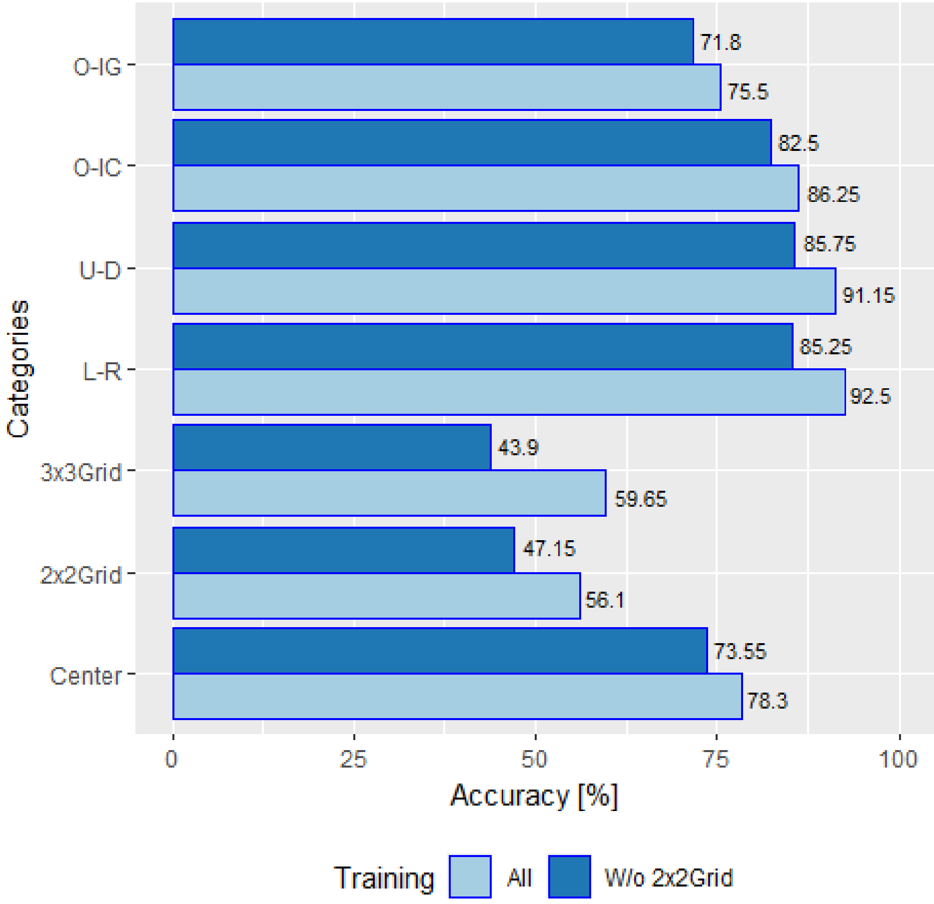




Trained without 2x2Grid and 3x3Grid

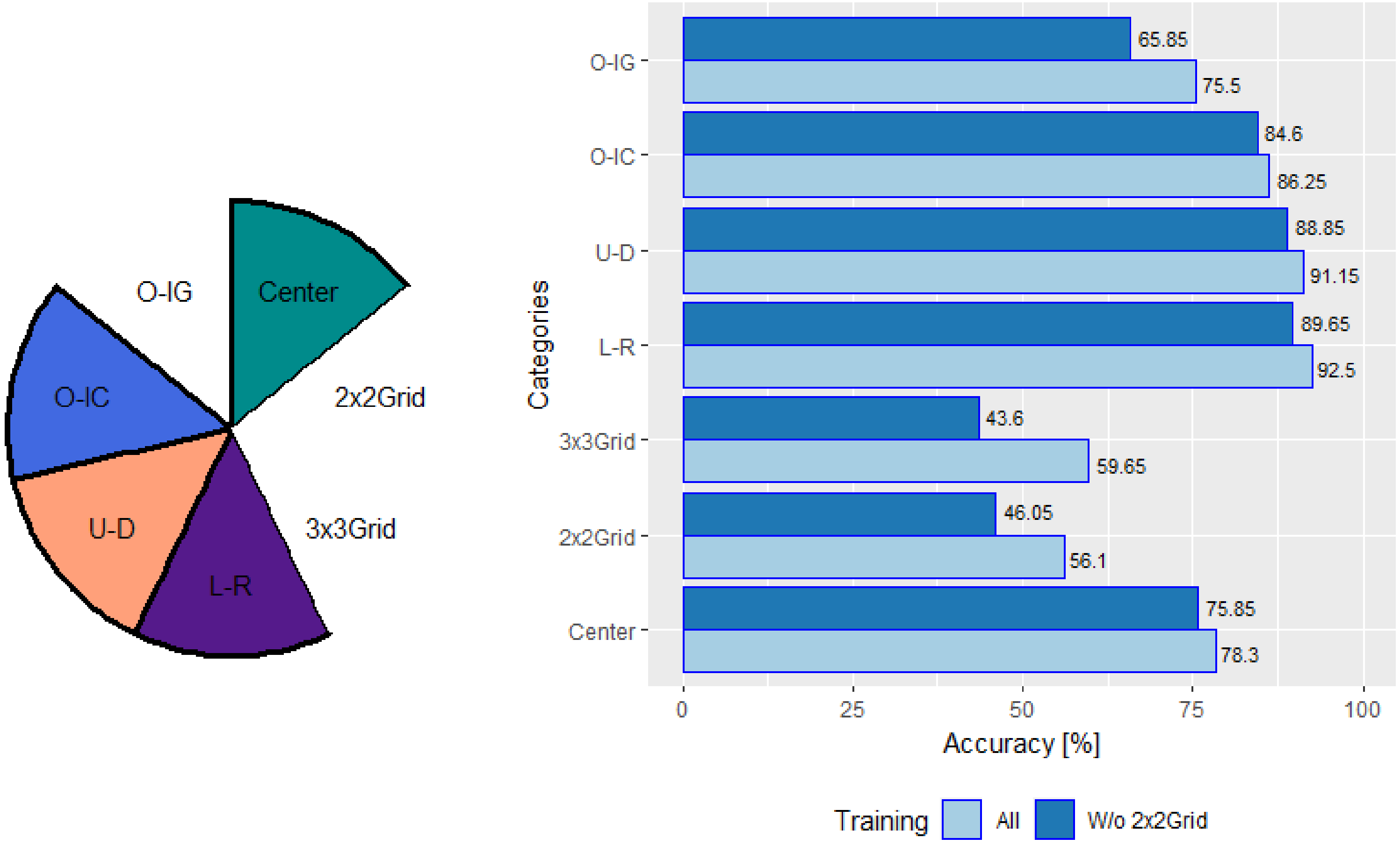
Training w/o 2x2Grid and 3x3Grid



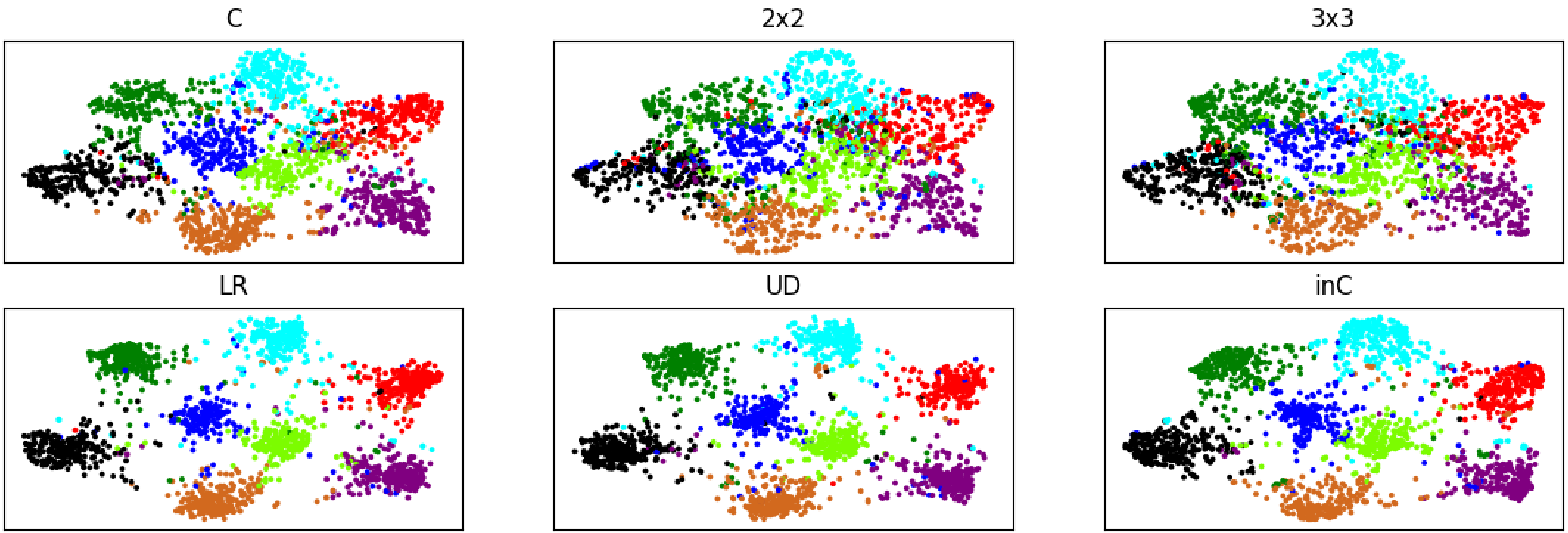


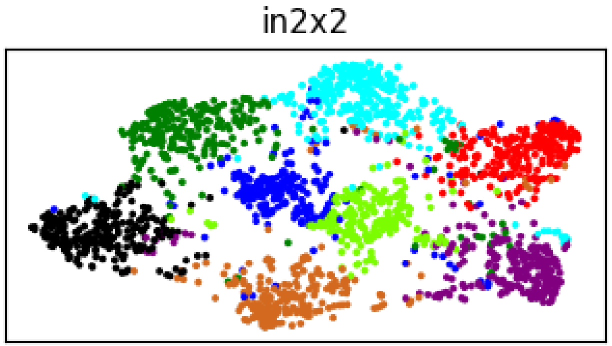
Trained without 2x2Grid, 3x3Grid and Out-InGrid

Training w/o 2x2Grid, 3x3Grid and O-IG



VISUALIZATION





UMAP is a general purpose manifold learning and nonlinear dimension reduction algorithm.[1]

1. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv–1802.

Proposed future advancement

- A metrics to quantify
  - the similarity among the problem types
  - the similarity among the answer set options
- Decreasing the training set size by keeping a certain portion of each problem category

THANK YOU FOR YOUR ATTENTION

Otázky a Komentáre

1. Popri textovom výpise veľkostí vrstiev a parametrov siete ResNet-18 na obrázku 2.1 by sa hodil aj názorný obrázok architektúry ResNet s vysvetlením.
2. V časti 3.1.1 píšete, že ste vo vašich experimentoch použili sieť s váhami predtrénovanými na ImageNete, ale že výsledná úspešnosť nie je väčšia ako v experimentoch v [Zhuo and Kankanhalli, 2020] bez predtrénovania. Čím si to vysvetľujete? V časti 2.5 píšete: „The input layer of the standard ResNet-18 model is 3-dimensional, which is suitable for our input, although it is originally meant to take in RGB images. The images in the RAVEN dataset are grayscale, therefore they have only 1 color dimension, which means that we can fit in the whole row as the input in one step." Bol tento trik použitý iba pri učení bez učiteľa? Pretože ak by bol použitý aj pri učení s učiteľom reportovanom v časti 3.1.1, to by vysvetľovalo, prečo by bolo predtrénovanie na ImageNete bezcenné, keďže tam vstupné RGB dimenzie farebného obrázku dostávajú nový význam (pokiaľ som to správne pochopil).

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	7×7, 64, stride 2
conv2_x	$56 \times 56 \times 64$	3×3 max pool, stride 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	7×7 average pool
fully connected	1000	512×1000 fully connections
softmax	1000	