

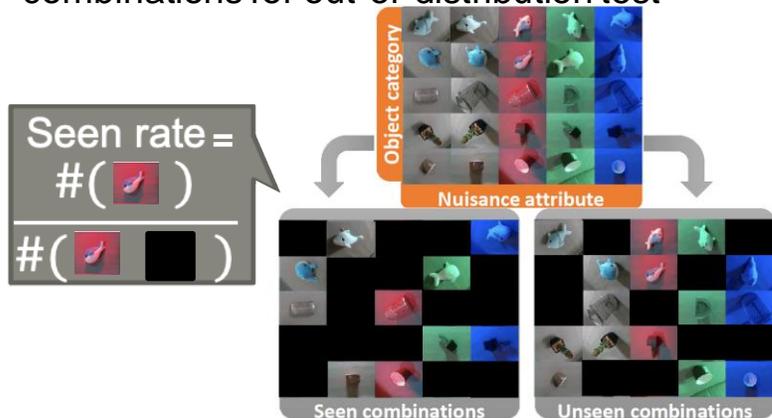
- We propose a novel regularization method to improve generalization performance of DNNs on the out-of-distribution setting
 - Spurious correlations in training data caused by nuisance attribute or factor cause major performance degradation
 - Matching neural activity for pairs of images belonging to the same object category during supervised learning prevents it (ENAMOR)
 - ENAMOR is shown to enforce the invariance to the nuisance factors

Experimental setting

- Datasets for object classification with nuisance factor labels

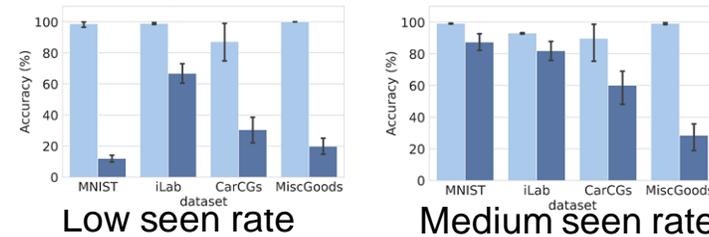
MNIST	iLab	CarCGs	MiscGoods
position	orientation	orientation	illumination

- Bias controlled experiments: split a dataset to seen combinations for training and unseen combinations for out-of-distribution test



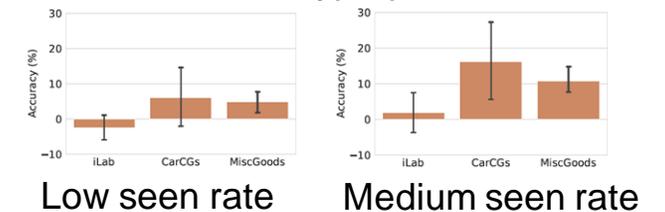
Performance Degradation

- Spurious correlations by nuisance factor caused up to **80%** degradation in ResNet-18



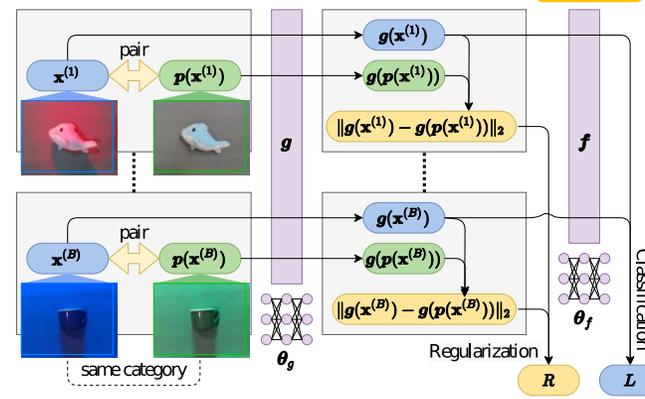
Improvement of performance

- ENAMOR achieved up to **20%** recovery (MNIST was used for hyperparameter search)



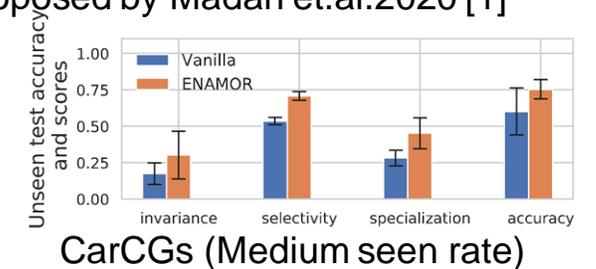
ENAMOR

- Using pairs of images (x , $p(x)$) to calculate Regularization term R



Improvement of invariance

- ENAMOR increased the invariance score proposed by Madan et.al.2020 [1]



Acknowledgement

We thank Tomaso Poggio, Pawan Sinha and Hanspeter Pfister for their insightful advice and warm encouragement. We have had a technical support by Shinichi Matsumoto and QNET engineering team.

[1] Madan et. al., On the Capability of Neural Networks to Generalize to Unseen Category-Pose Combinations Technical Report CBMM Memo No. 111, Center for Brains, Minds and Machines, 2020.