Noise or Signal: The Role of Image Background in Object Recognition

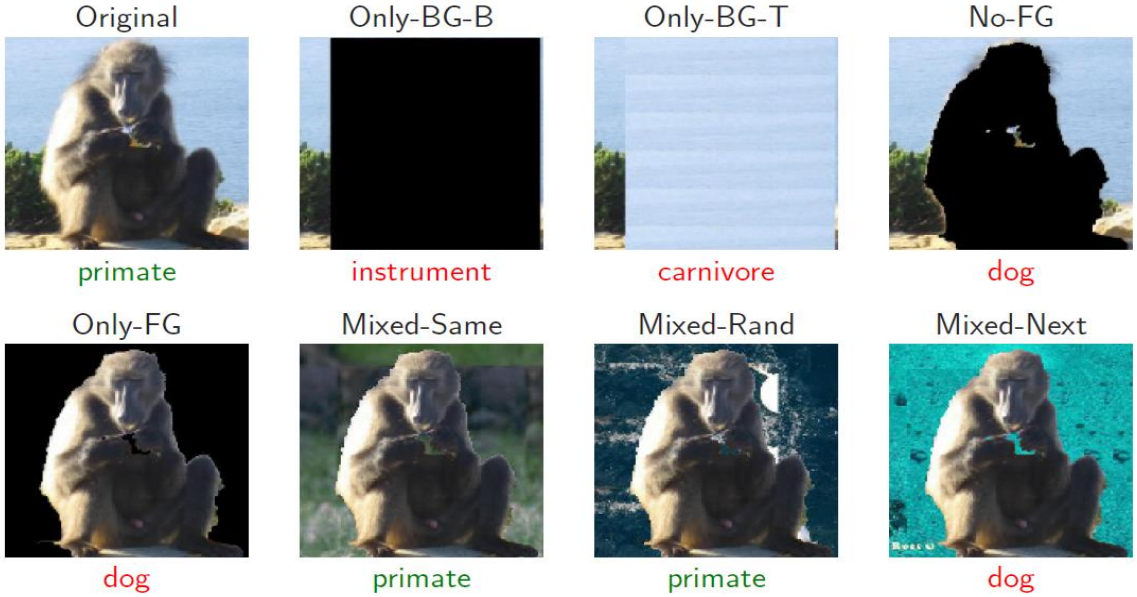
Kai Xiao, Logan Engstrom Aleksander Madry, Andrew Ilyas

International Conference on Learning Representations 2021

경영과학연구실 김윤석

Introduction

- A deep learning-based image classification model is biased towards the training dataset.
- For this reason, unexpected results may occur when testing with modified data.



Related work

Contextual bias

- Antonio Torralba and Alexei Efros. “Unbiased look at dataset bias”. In: CVPR 2011. 2011, pp. 1521–1528.
- Aditya Khosla et al. “Undoing the damage of dataset bias”. In: European Conference on Computer Vision (ECCV). 2012.
- Choi Myung Jin, Antonio Torralba, and Alan S. Willsky. “Context models and out-of context objects”. In: Pattern Recognition Letters. 2012.

Background dependence

- Jianguo Zhang et al. “Local features and kernels for classification of texture and object categories: A comprehensive study”. In: International Journal of Computer Vision (IJCV). 2007.
- Shiori Sagawa et al. “Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization”. In: International Conference on Learning Representations. 2020.

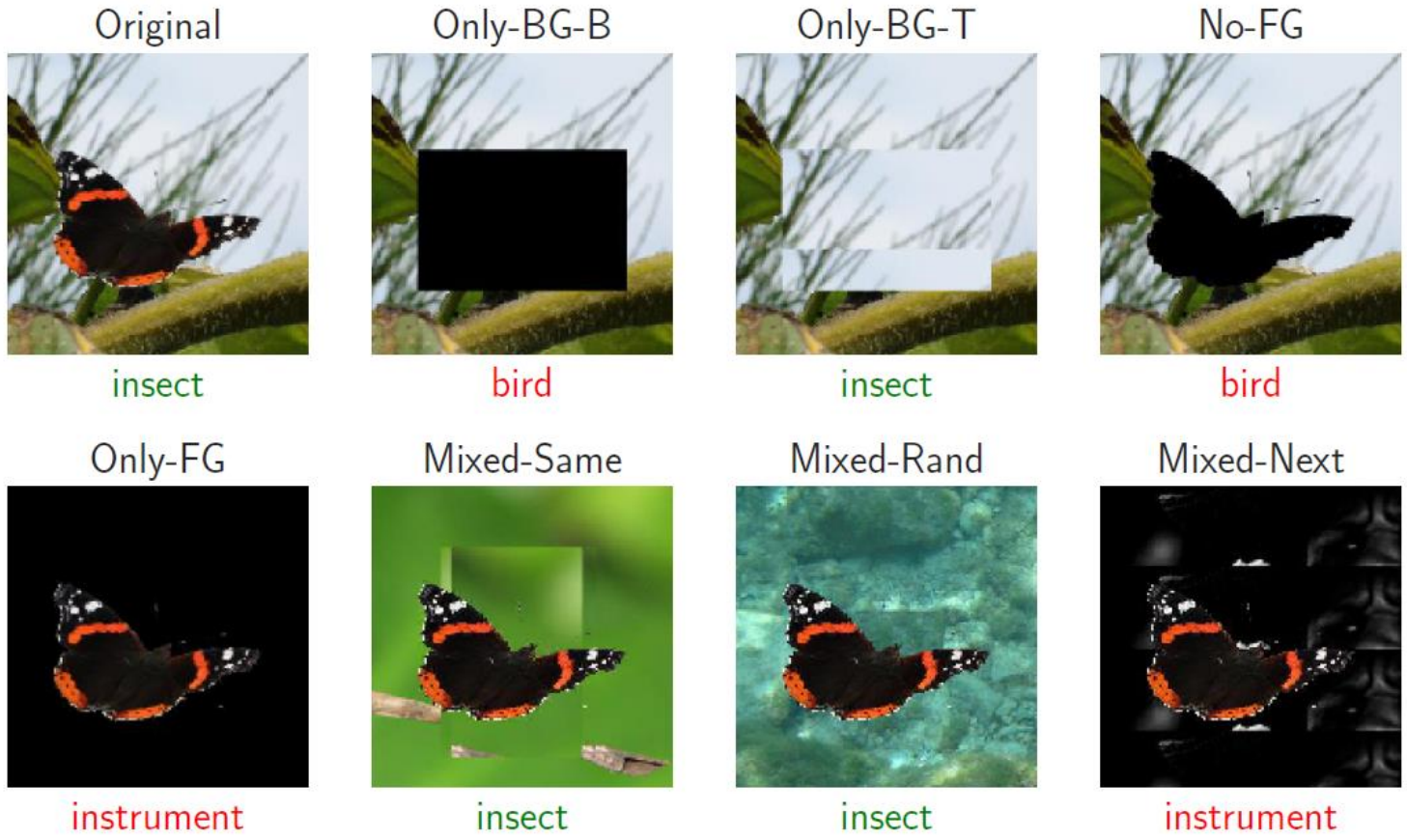
Problem statement & key idea

- Problem statement
  - The goal of this study is to achieve a deeper understanding of how recent image classifiers utilize image backgrounds.

- Key idea
  - They created a dataset called ImageNet-9 by removing or modifying backgrounds or foregrounds in ImageNet, and trained and evaluated the models using this dataset.

ImageNet-9

- ImageNet-9 consists of Original, Only-BG-B, Only-BG-T, No-FG, Only-FG, Mixed-Same, Mixed-Rand, and Mixed-Next.



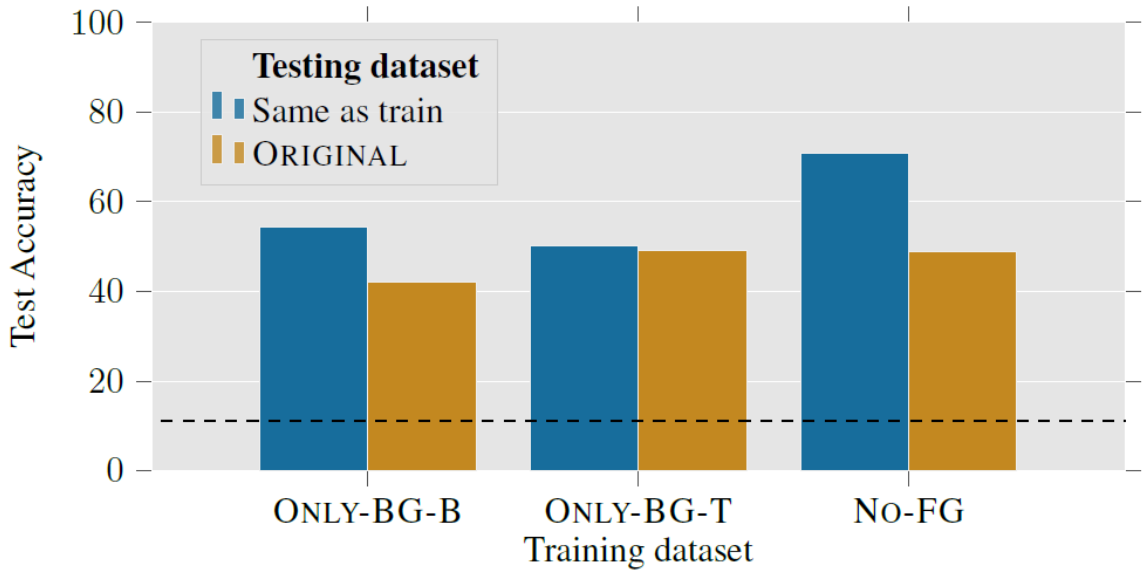
Method

- They evaluated the performance of ImageNet pre-trained models on ImageNet-9 and confirmed the dependence on backgrounds.
- They conducted various training using ImageNet-9 to investigate the role of backgrounds.

1. Backgrounds suffice for classification
2. Models exploit background signal for classification
3. Models are vulnerable to adversarial backgrounds
4. Training on Mixed-Rand reduces background dependence
5. A fine grained look at dependence on backgrounds

Backgrounds suffice for classification

- Performance evaluation experiment on two datasets after training on each individual data item.
- The model trained using only background images shows a performance of over 40% on original images.
- Deep learning-based image classification models do not solely extract features from objects.



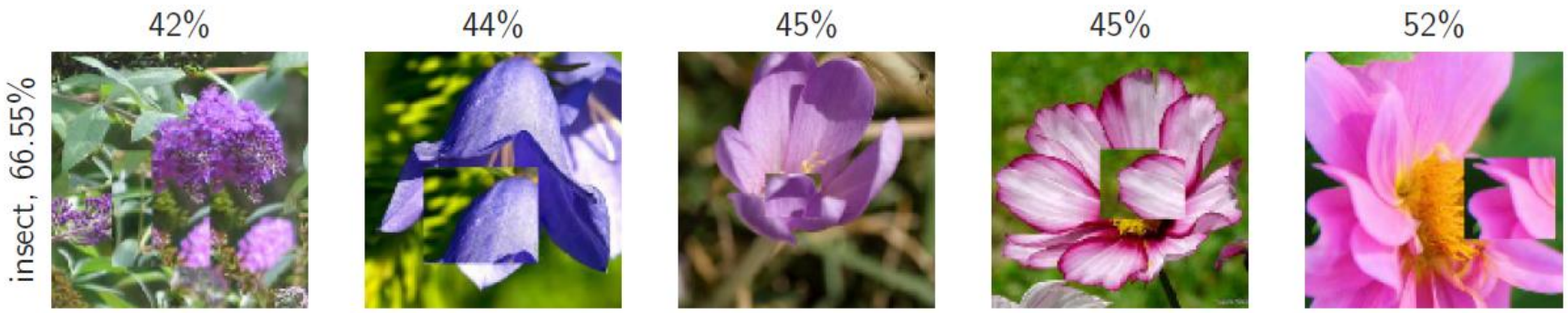
Models exploit background signal for classification

- The performance degradation when the background class changes was confirmed through BG-gap.
- Even if the foreground and label match, performance can decrease due to changes in the background.
- The experiment showed that the deep learning model's training data has a dependence on the background.

	Pre-trained on ImageNet					Trained on IN-9L				
Test dataset	MobileNet-v3	EfficientNet-b0	ResNet-50	WRN-50x2	DPN-92	AlexNet	ShuffleNet	ResNet-50	WRN-50x2	VGG16-BN
ImageNet	67.9%	77.2%	77.6%	78.5%	80.0%			——		
ORIGINAL	91.0%	95.6%	96.2%	95.8%	96.8%	86.7%	95.7%	96.3%	97.2%	97.6%
IN-9L	90.0%	94.3%	95.0%	95.5%	96.0%	83.1%	93.2%	94.6%	95.2%	96.0%
ONLY-BG-T	15.7%	11.9%	17.8%	20.7%	20.6%	41.5%	43.6%	43.6%	45.1%	45.7%
MIXED-SAME	69.7%	79.7%	82.3%	81.7%	85.4%	76.2%	86.7%	89.9%	90.6%	91.0%
MIXED-RAND	56.1%	67.8%	76.3%	73.0%	77.6%	54.2%	69.4%	75.6%	78.0%	78.0%
BG-gap	13.6%	11.9%	6.0%	8.7%	7.8%	22.0%	17.3%	14.3%	12.6%	13.0%

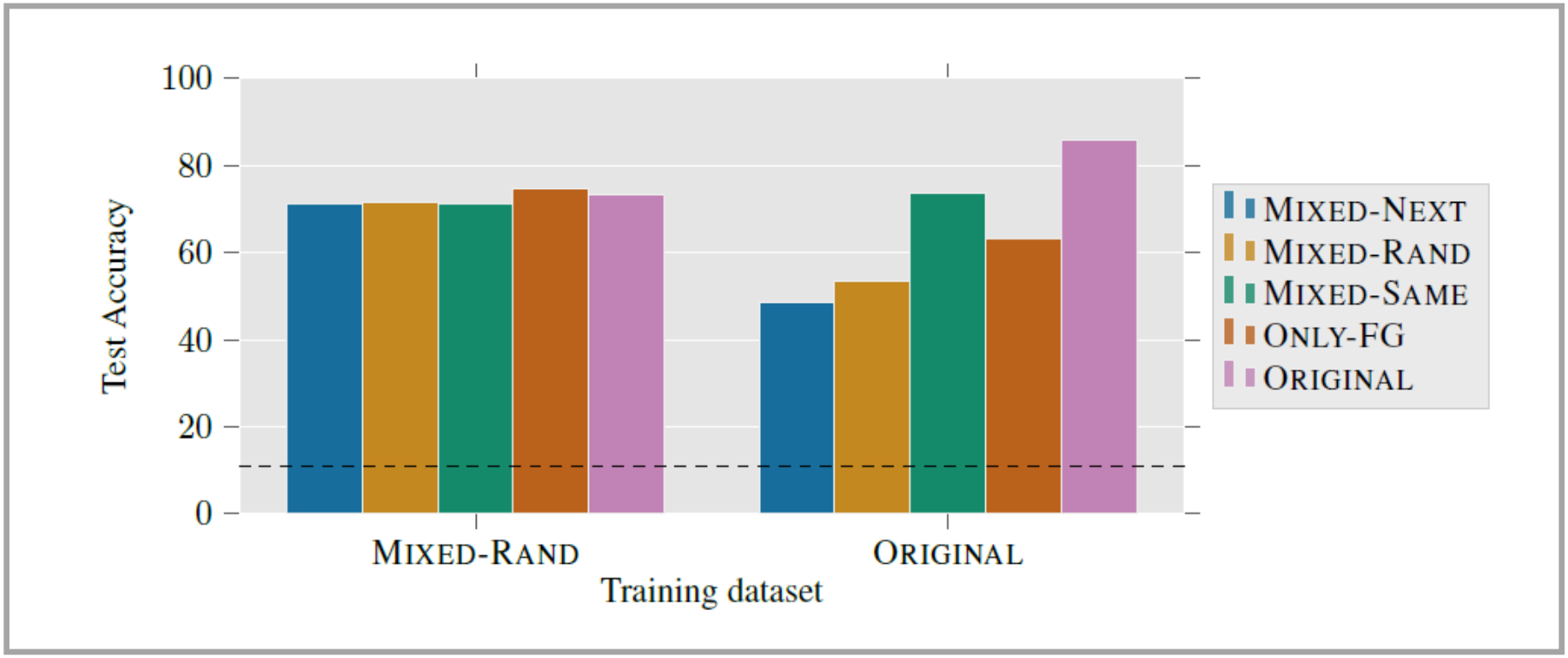
Models are vulnerable to adversarial backgrounds

- An experiment to verify the success rate of attacks on the foreground.
- 87.5% of foregrounds are affected by adversarial backgrounds.
- Model has a high dependence on the background to the extent that it ignores the object.



Training on Mixed-Rand reduces background dependence

- Experiment to investigate background dependence when training on the Mixed-Rand dataset.
- When trained with the Mixed-Rand dataset, model's dependence on background decreases.



Training on Mixed-Rand reduces background dependence

- Examining the weight allocation for inputs to a model trained using Mixed-Rand.
- It can be observed that the model trained with the Mixed-Rand dataset emphasizes foreground pixels more than the model trained with the Original dataset.

Saliency map

- A saliency map is a heatmap that visualizes important information in an input image
- Grad-CAM method is commonly used to generate saliency maps..

- Grad-CAM

  c: index of class, k: index of feature map, (i, j): index of pixel

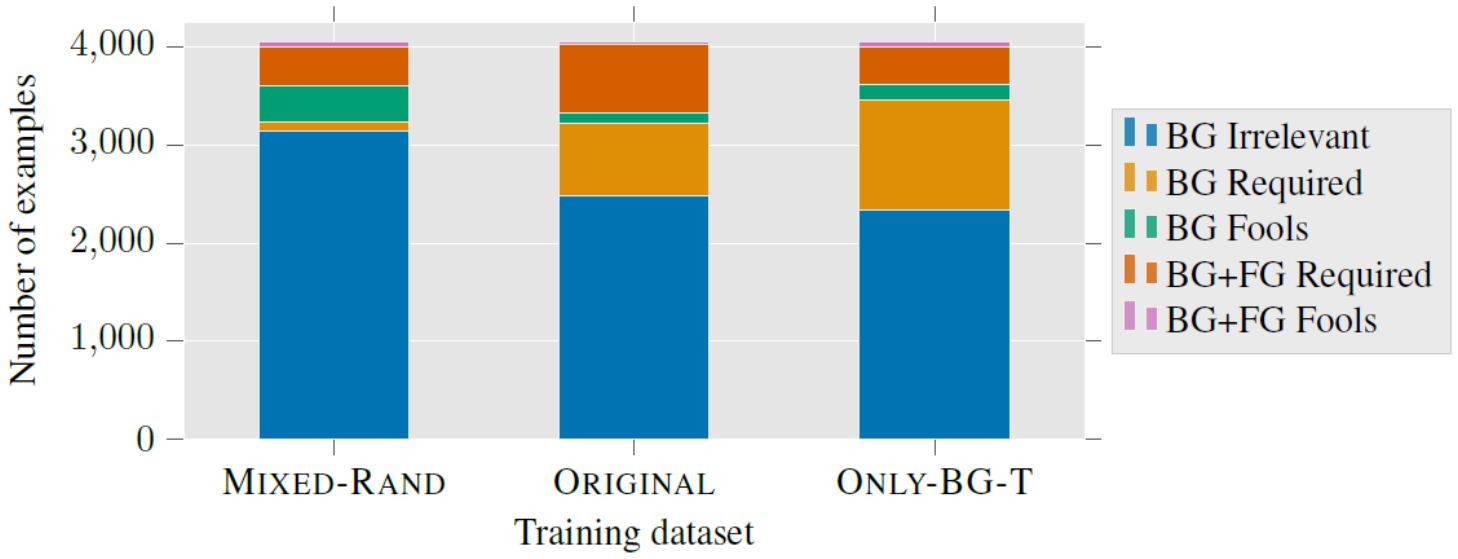
  f: feature map, S^c: Classification score for the c-th class

$$L^c_{Grad-CAM}(i,j) = ReLU(\sum_k a^c_k f_k(i,j))$$

$$a^c_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial S^c}{\partial f_k(i,j)}$$

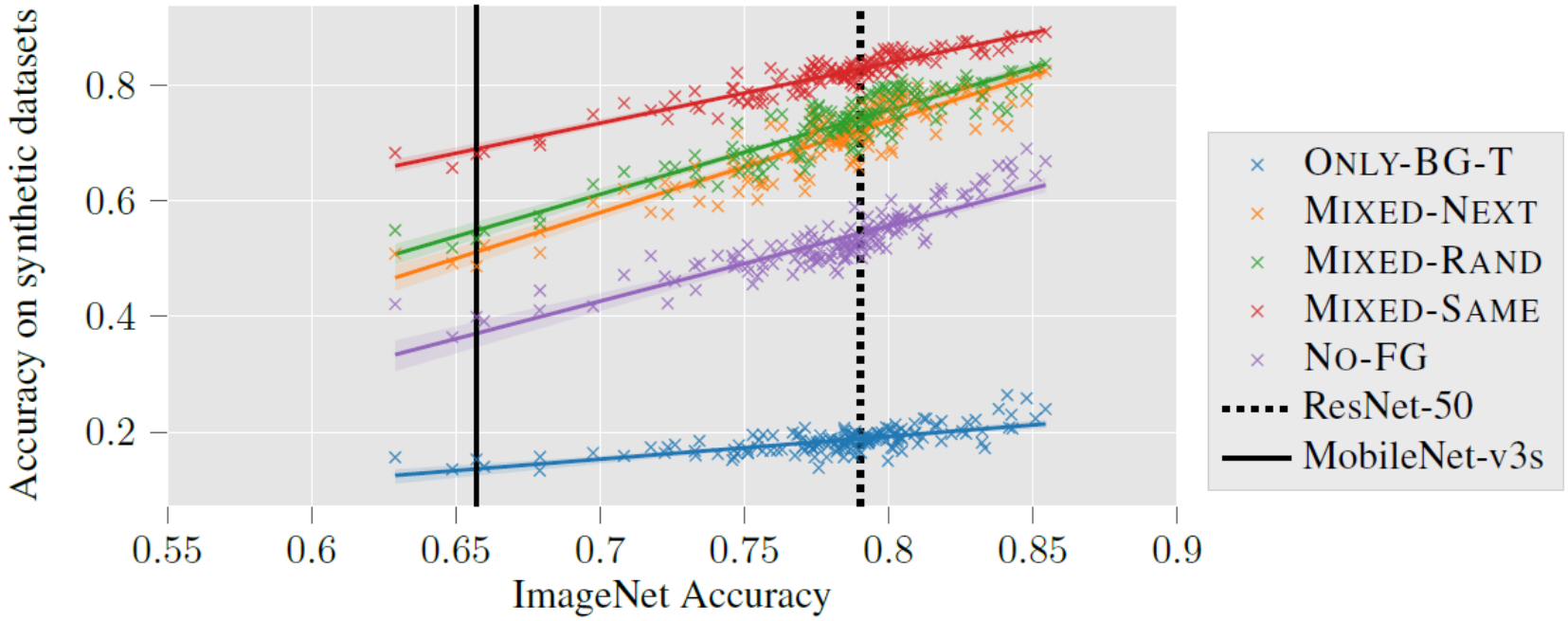
A fine grained look at dependence on backgrounds

- An experiment to analyze and understand background dependency.
- Foreground signal class prediction and background signal class prediction are used for Mixed-Rand and Only-BG-T predictions.
- For most images, the background is necessary for correct classification.



Benchmark Progress and Background Dependence

- They observe the performance change in ImageNet-9 according to the performance improvement on ImageNet.
- As performance increases, the performance of datasets that can be classified based on background information also increases.
- However, the ability to utilize foreground information increases more rapidly with the increase in performance..



Conclusion

- In image classification tasks, learning from a limited dataset can result in a dependency on the background

- Although a model's robustness can be reduced by its dependency on the background image, it can also be used as context, so further research is needed.

- They propose a data augmentation method to reduce background dependency.