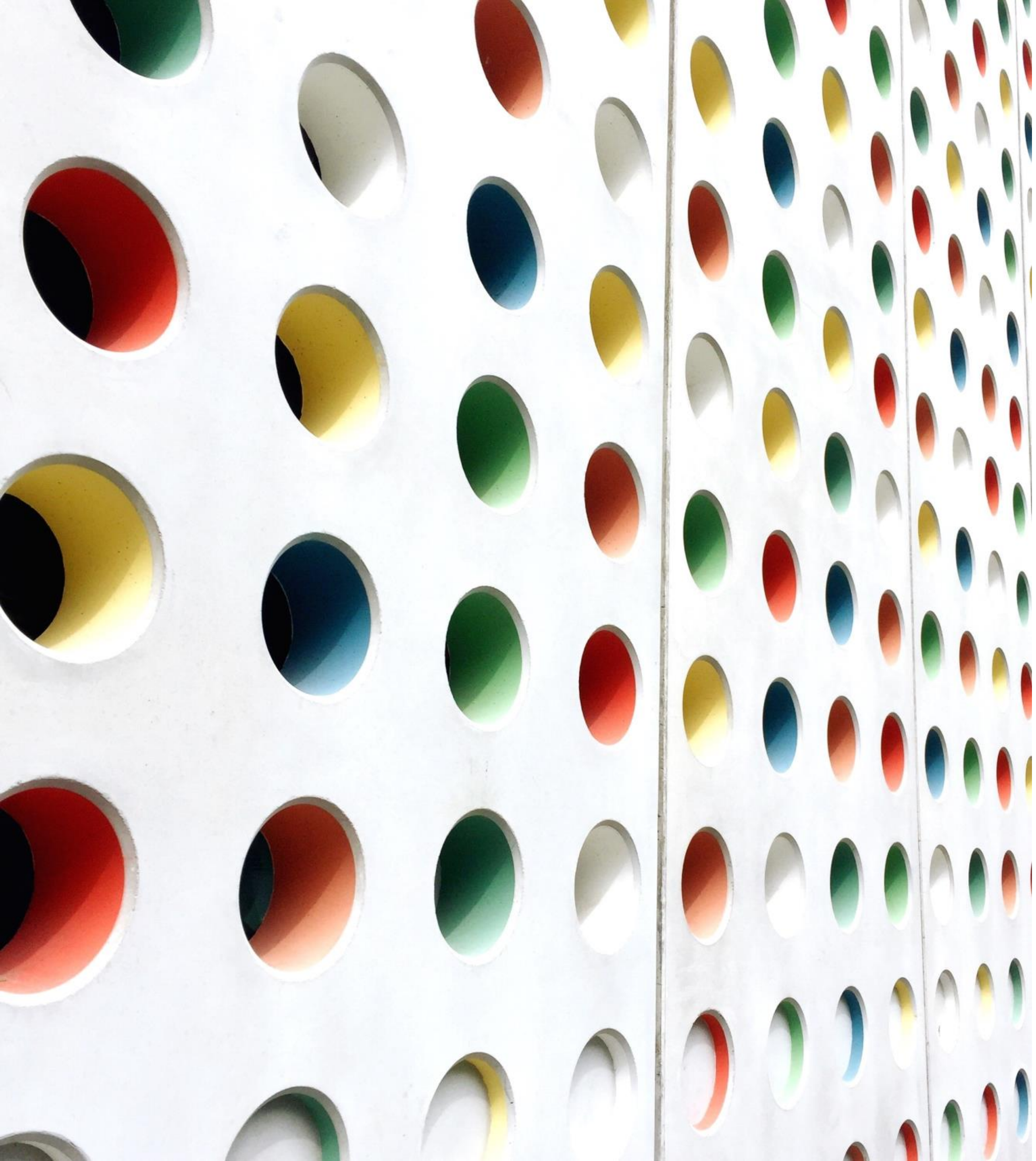


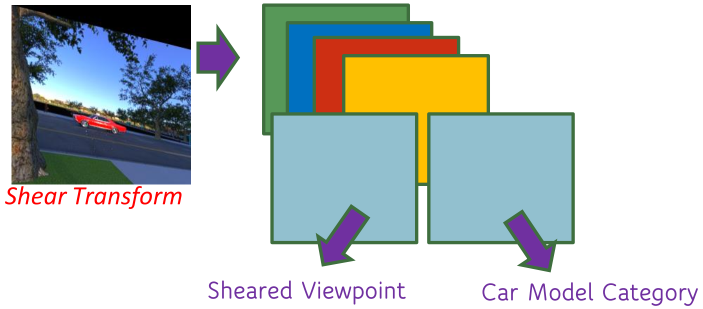
Deep Learning Group Project

Investigating Techniques to Improve CNN Generalization on Out of Distribution Samples

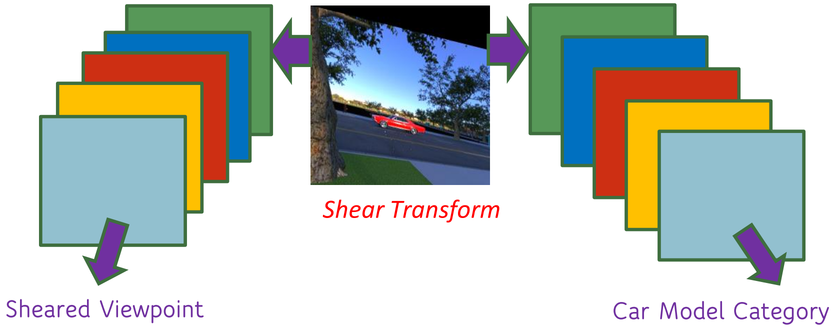
ALEXANDRE RAEVEL
ZIXIANG LOH
JOHN BLACKWELDER

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Introduction



Shared Architecture



Separate Architecture

- Deep Neural Networks (DNNs) used for object detection with respect to different viewpoints (illuminations, rotations and orientations, different conditions and backgrounds) often fail to generalize well to images containing a combination of object category and viewpoint not contained in the training set; we refer to such images as the Out-of-distribution (OoD) dataset [1,2]

- DNNs accuracy in recognizing OoD objects is shown to increase when more in-distribution (InD) category-viewpoint combinations are introduced in the training data (also known as data diversity).

- DNN OoD sample accuracy is also shown to improve when batch normalization parameters are tuned and the DNN is trained over a larger number of epochs after convergence (late-stopping).

- The type of DNN architecture (separate vs shared) is shown to seriously impact the DNN OoD sample accuracy. Separating viewpoint and category estimation into independent branches improves the OoD sample test accuracy significantly, as opposed to estimating both these features together.

Introduction

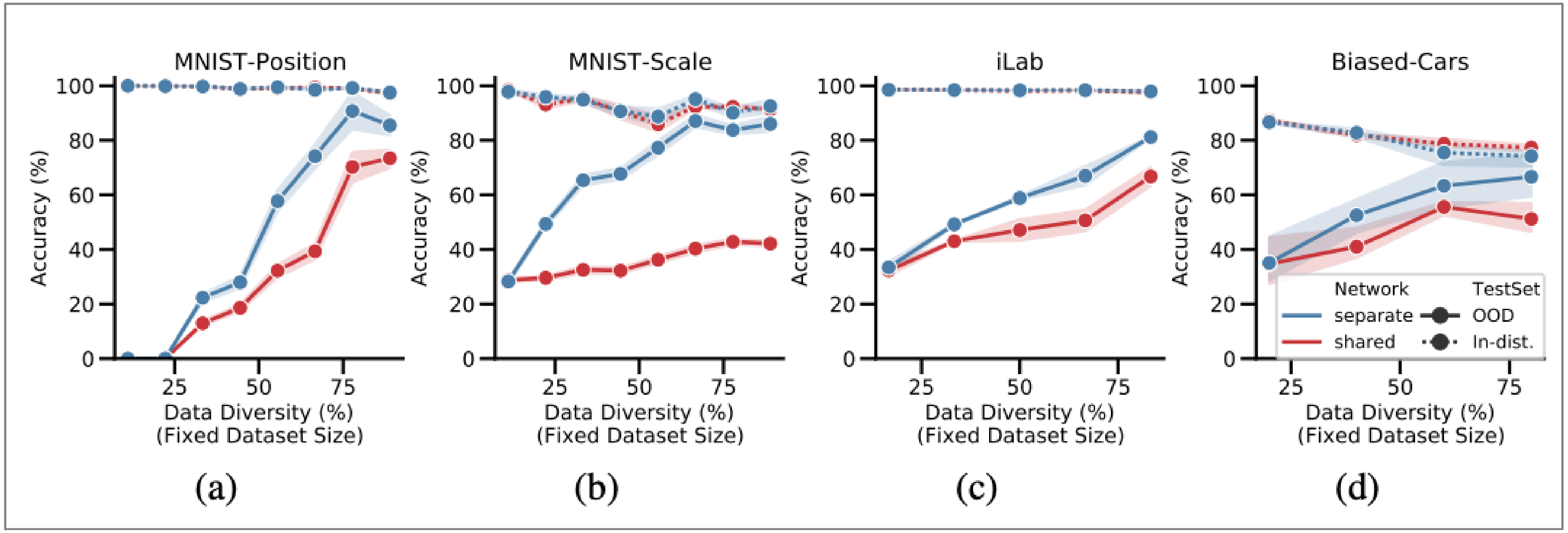
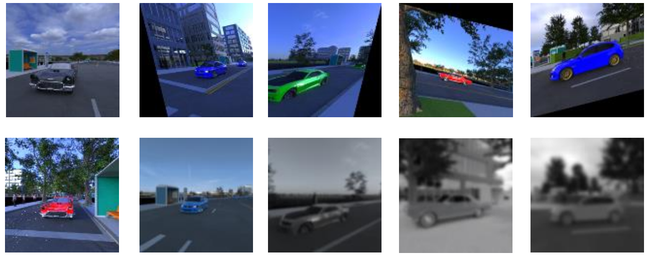
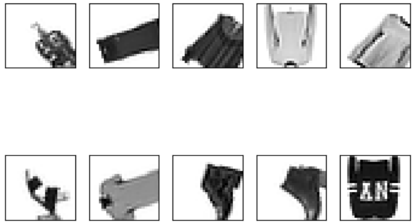


Fig. 1 - Results from Madan et al. [3] that show, across datasets, the role data diversity plays in improving OoD accuracy

Previous Work



Shear/Blur Transform Viewpoint Cars



Rotated MNIST Clothing (Deformable Object)

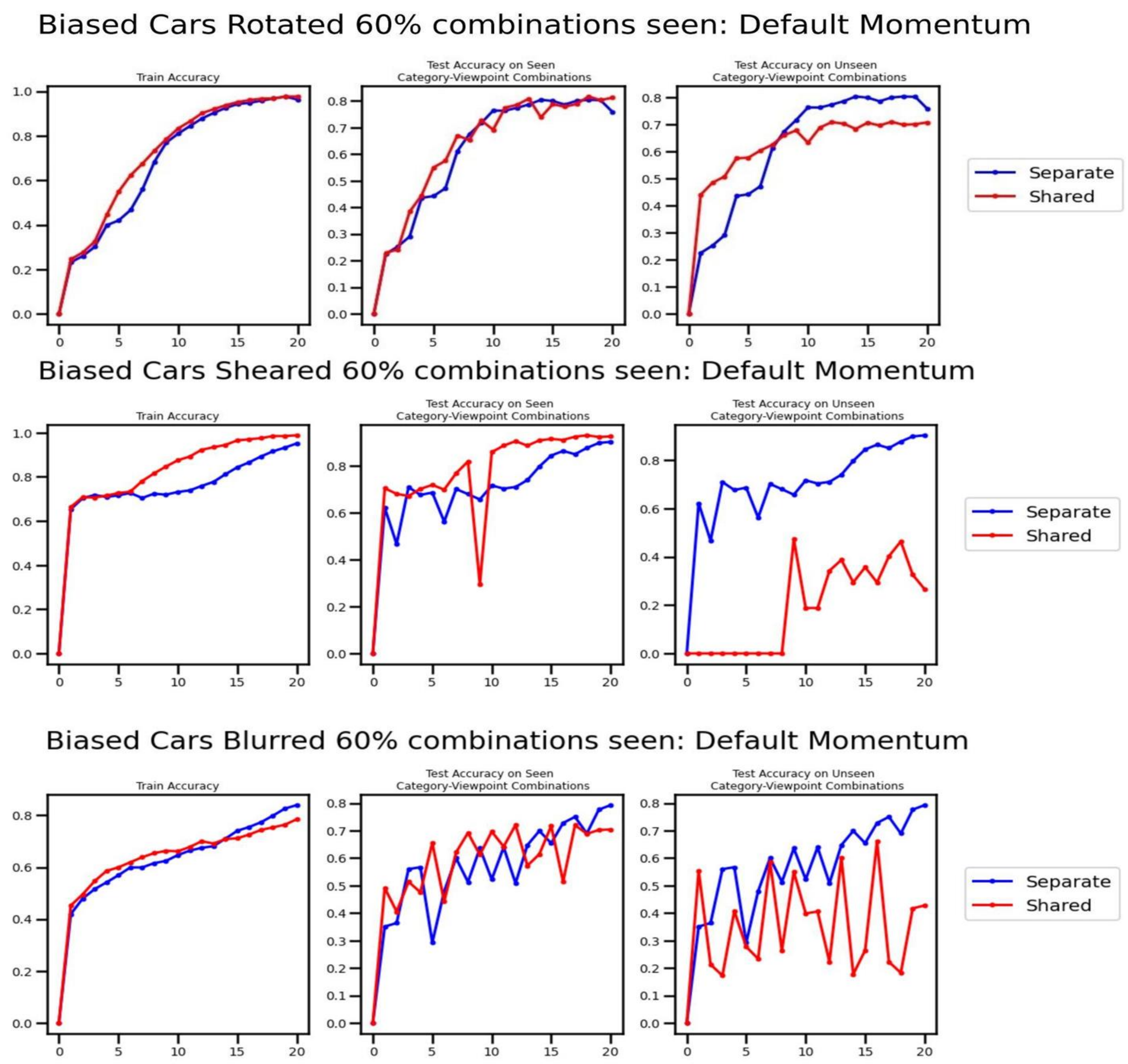
- Madan et al., acknowledge that their work was not validated on datasets that feature combinations of viewpoints and deformable object categories. Furthermore, only rotation image data transformation is considered to create different viewpoints. Hence, this work investigates different image data transformations across 2 datasets (biased car dataset and MNIST clothing dataset).

- Sakai et al. demonstrate that tuning batch normalization parameters improved OoD object recognition, no mention is made of the kind of relationship that underlies OoD object recognition accuracy with respect to batch normalization parameters. As a result, this work uncovers to what degree batch normalization parameters can play a role in ameliorating OoD object recognition.

- Sakai et al. do show that late-stopping can be used as a technique to improve OoD object recognition accuracy on DNNs built with a shared architecture, but the technique was not applied to the separate model architecture. Thus, this work assesses how the type of architecture can influence the impact of late-stopping on OoD object recognition accuracy.

Impact of different image data transformations on model accuracy

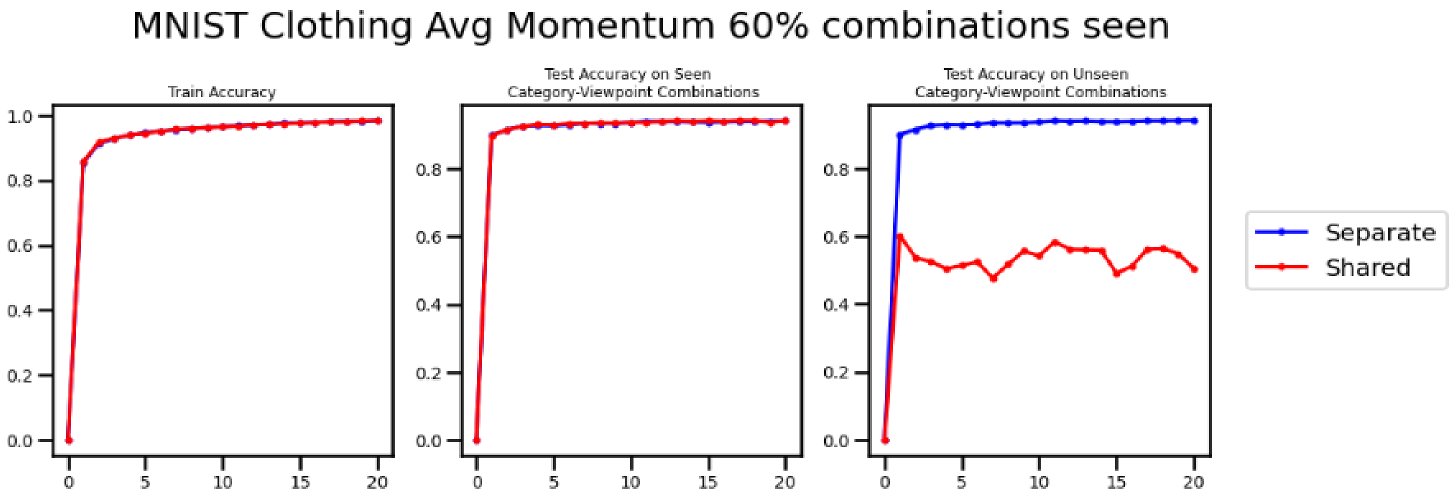
- Comparing separate, shared model architecture performance over
  - Original biased cars dataset
  - Biased car with shearing applied
  - Biased car with blurring applied
- Separate model continues to perform well over the OoD test dataset even after data transformations are applied
- Shared model struggles to learn when these transformations are applied to the training set



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Impact of changing the dataset on model accuracy

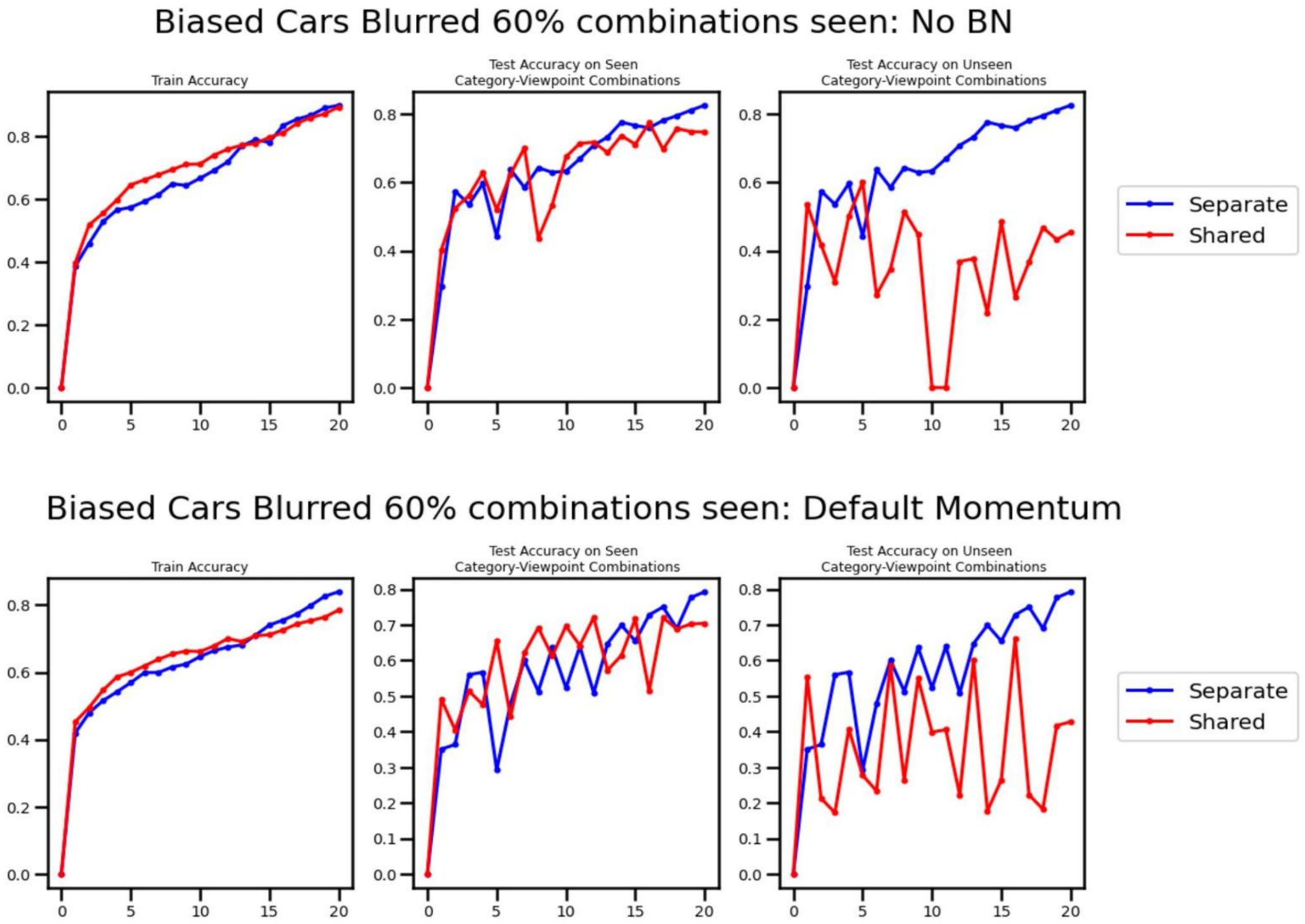
- Comparing separate, shared models on deformable objects (*i.e.* the rotated MNIST Clothing dataset)
- Separate model outperforms shared model
- Late-stopping does not appear to improve performance



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Impact of including and removing batch normalization on model accuracy

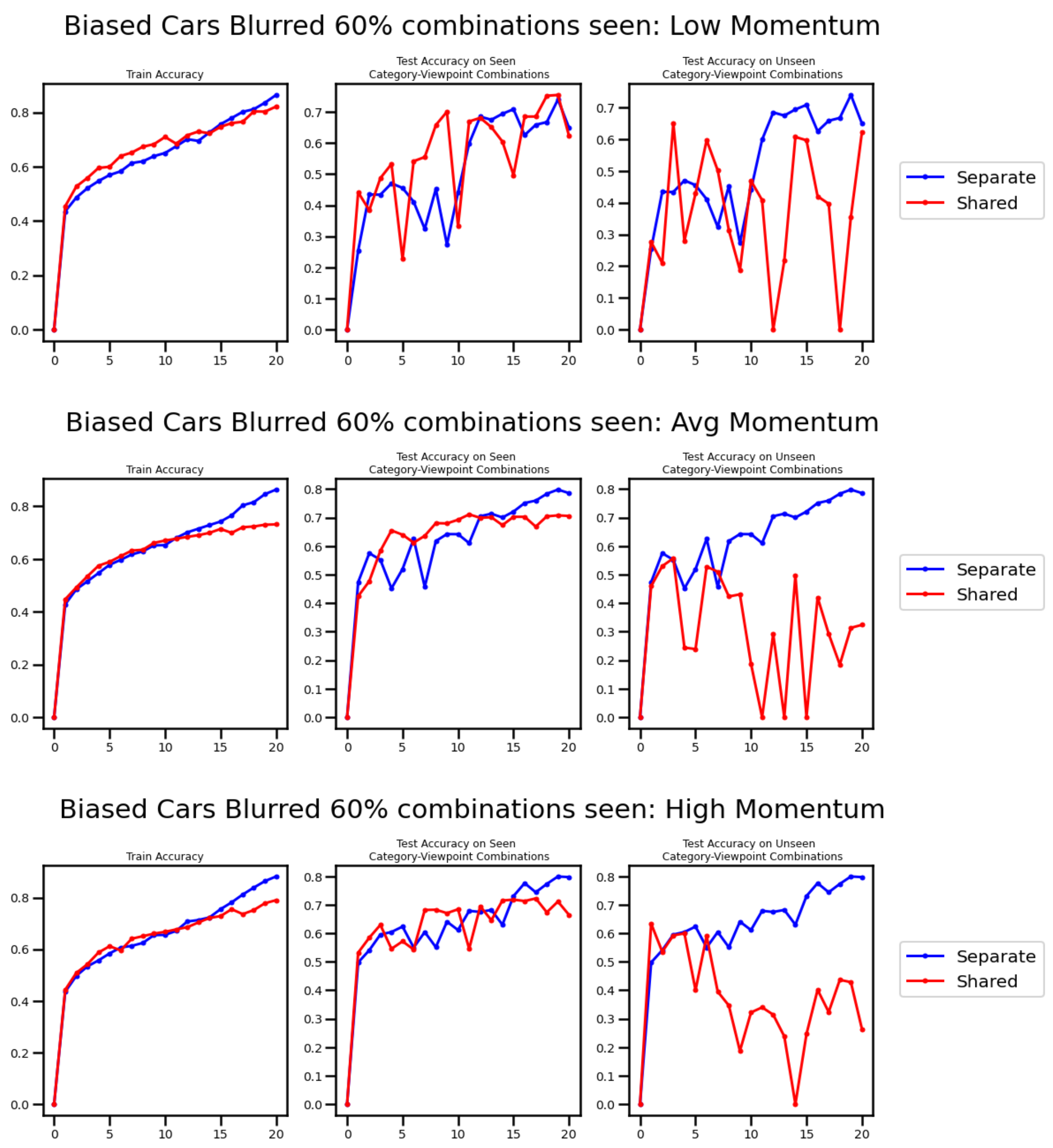
- Comparing how separate, shared models perform over the blurred dataset with no batch normalization, default batch normalization
- We do not observe any significant effects



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Impact of changing the batch normalization momentum parameters on model accuracy

- Comparing how separate, shared models perform over the blurred dataset after adjusting the batch normalization momentum parameter
- Decreasing the Pytorch batch normalization momentum parameter appears to increase oscillations in separate and shared InD and OoD test accuracy



COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

GitHub Repo

Conclusion

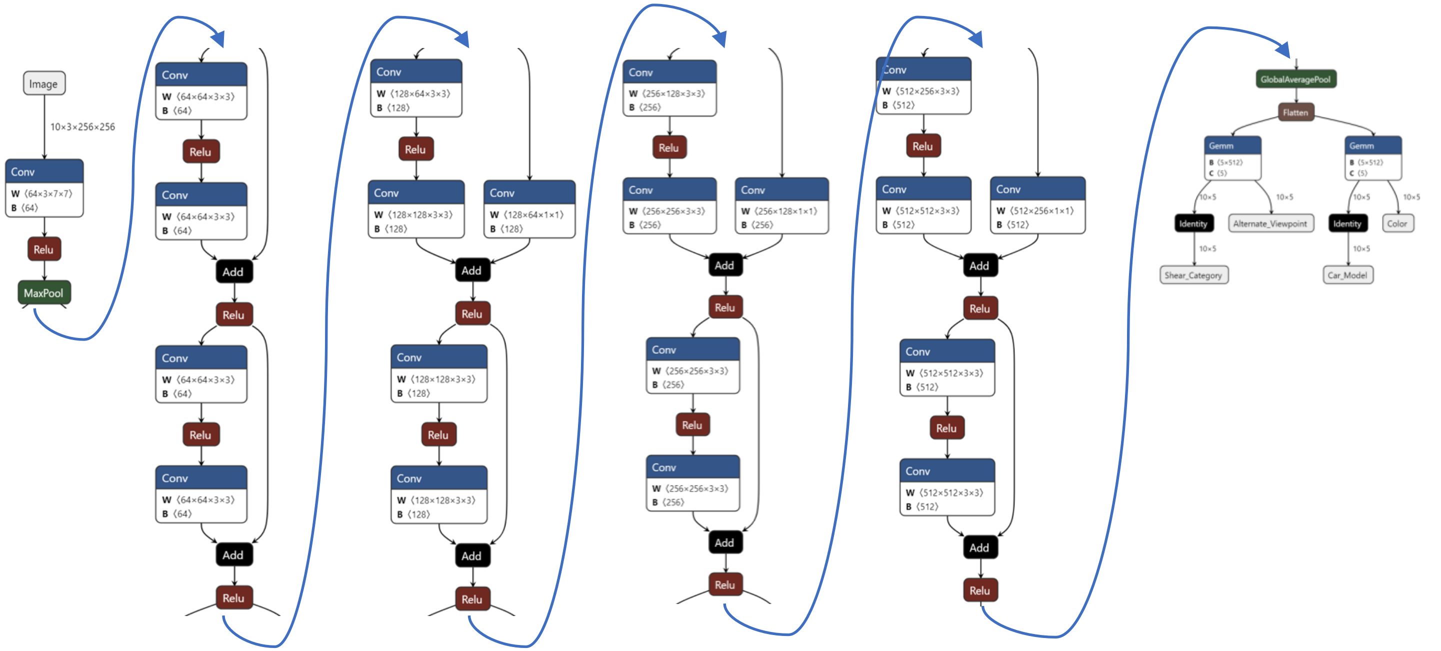
- The type of image data transformation (shearing, blurring, rotation) applied to create viewpoint-object combinations seriously impacts the training, InD and OoD accuracies of the CNN model
- The type of model implementation (shared versus separate) profoundly influences the model's accuracy for different image data transformations, with the separate architecture outperforming the shared one for the shearing and blurring image transformations
- With regards to batch normalization, its presence is inconclusive on improving training, InD and OoD accuracy irrespective of the kind of image transformation applied. However, when tuning the momentum parameter of the batch normalization layers, irrespective of the image transformation implemented, the separate architecture's OoD accuracy oscillates more as the Pytorch momentum parameter decreases.
- With regards to late-stopping, the improvement in training, InD and OoD accuracy clearly depends on the dataset considered

References

[1] A. Sakai, T. Sunagawa, S. Madan, K. Suzuki, T. Katoh, H. Kobashi, H. Pfister, P. Sinha, X. Boix, and T. Sasaki, Three approaches to facilitate DNN generalization to objects in out-of-distribution orientations and illumination

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, Deep residual learning for image recognition, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778

[3] S. Madan, T. Henry, J. Dozier, H. Ho, N. Bhandari, T. Sasaki, F. Durand, H. Pfister, and X. Boix, When and how CNNs generalize to out-of-distribution category-viewpoint combinations, Nature Machine Learning 4 (2022), 146–153.



RESNET18 with Late Branching Combined Architecture Visualized via Netron. We use the best model as the baseline for our evaluation

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science