Multi-fidelity Data Augmentation for U-Net Retinal Vessel Segmentation

John Lee, Lauren Partin, Daniele E. Schiavazzi

Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, US.

Motivation

In the field of image recognition and segmentation, deep learning and model-based approaches have already exceeded humans in accuracy and efficacy. However, these models are significantly reliant upon the presence of high fidelity (or high resolution) images to perform well. Since the cost and time consumption of acquiring high quality data scales quickly with resolution, one would be tempted to augment a dataset consisting of a limited number of costly high-fidelity images with a large number of inexpensive low-fidelity images.

In this study, we train 12 identically parametrized deep learning model (U-Net) on datasets with a varying number of images at different fidelities, evaluating how the size and fidelity distribution in the input data affects accuracy and prediction variability.

Background

U-Net - The U-Net is a fully convolutional encoder-decoder architecture designed to produce fast, accurate image segmentation without the need for thousands of annotated training samples. It relies heavily upon extensive data augmentation to maximize available annotations [1]. Below is the structure of the original U-Net, consisting of an encoder (down-sampling) and a decoder (up-sampling).

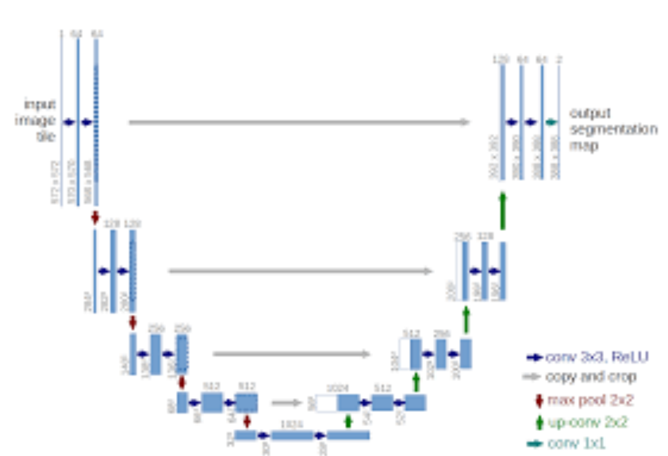


Figure 1: Structure of U-Net [1]

DropBlock – Many neural networks include a regularizing dropout layer which randomly selects neurons and zeros them, allowing for better generalization and reducing overfitting. However, dropout layers typically underperform when used with image tasks, dropping pixels randomly without attention to spatial locality. Our model incorporates DropBlock layers which drops pixels in N by N blocks and has shown to be more effective in image tasks [2].

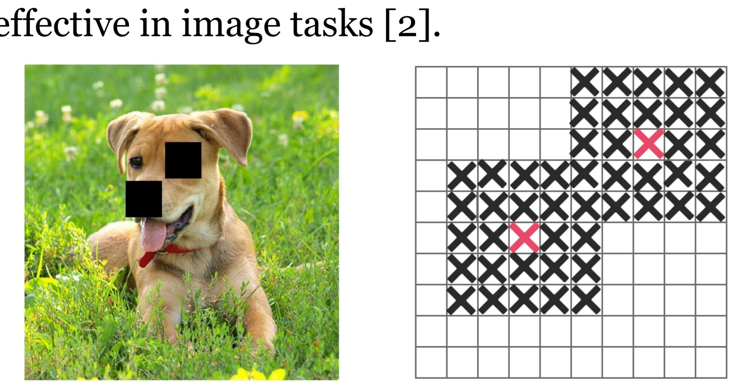
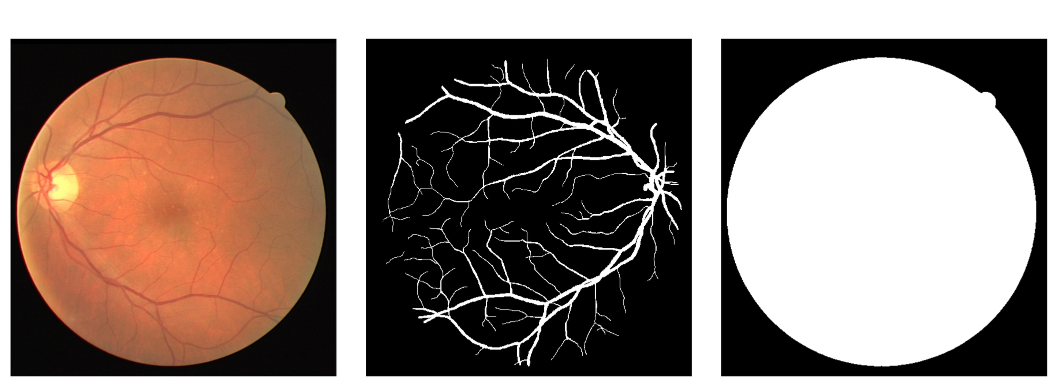


Figure 2: Dropblock Layer [2]

DRIVE Dataset – The DRIVE dataset contains retinal images of resolution of 584 by 565, manual annotations by individuals trained by an experienced ophthal-mologist and masks of the retinal region [3].



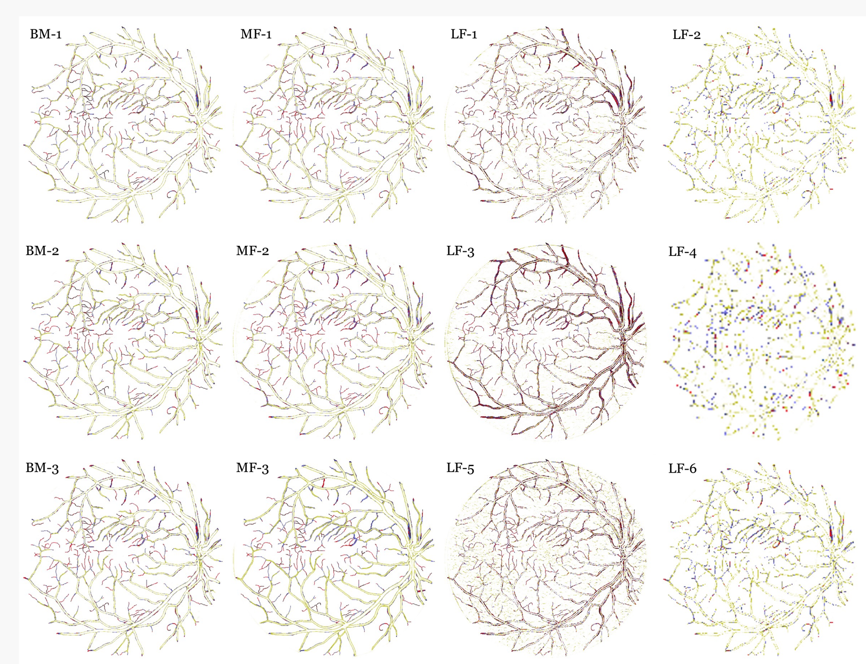
(a) Retinal Image (b) Manual Annotation (c) Bit Mask

Figure 3: DRIVE Dataset [4]

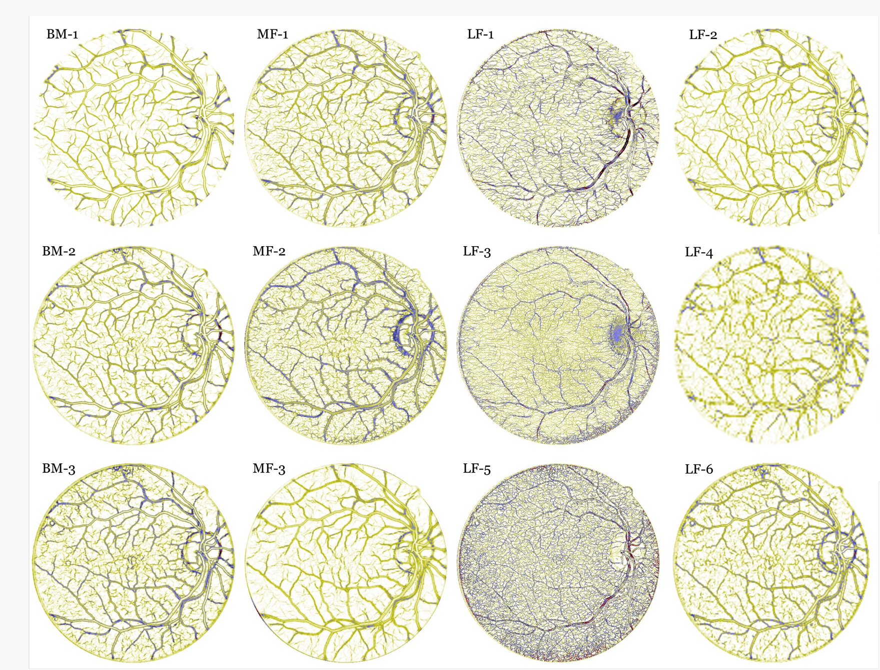
Results

Model	Test ID	F1-score (Dice)	AUROC Score	Accuracy Score	Test ID DropBlock	F1-score (Dice)	AUROC Score	Accuracy Score
Base	BM-1	.8076	.9787	.9539	BM-1-DB	.8073	.9797	.9546
	BM-2	.7952	.9735	.9504	BM-2-DB	.7853	.9742	.9508
	BM-3	.7876	.9711	.9487	BM-3-DB	.7739	.9715	.9488
Multi-Fidelity	MF-1	.7961	.9744	.9520	MF-1-DB	.7903	.9743	.9519
	MF-2	.7756	.9640	.9470	MF-2-DB	.7628	.9620	.9468
	MF-3	.7831	.9741	.9506	MF-3-DB	.7767	.9735	.9500
Low-Fidelity	LF-1	.6967	.9111	.9301	LF-1-DB	.6821	.9108	.9322
	LF-2	.7363	.9865	.9501	LF-2-DB	.7550	.9868	.9565
	LF-3	.5501	.7917	.8976	LF-3-DB	.4983	.7963	.9043
	LF-4	.7301	.9814	.9506	LF-4-DB	.7593	.9817	.9593
	LF-5	.5810	.8691	.8834	LF-5-DB	.5791	.8620	.8971
	LF-6	.7110	.9826	.9451	LF-6-DB	.7439	.9826	.9559

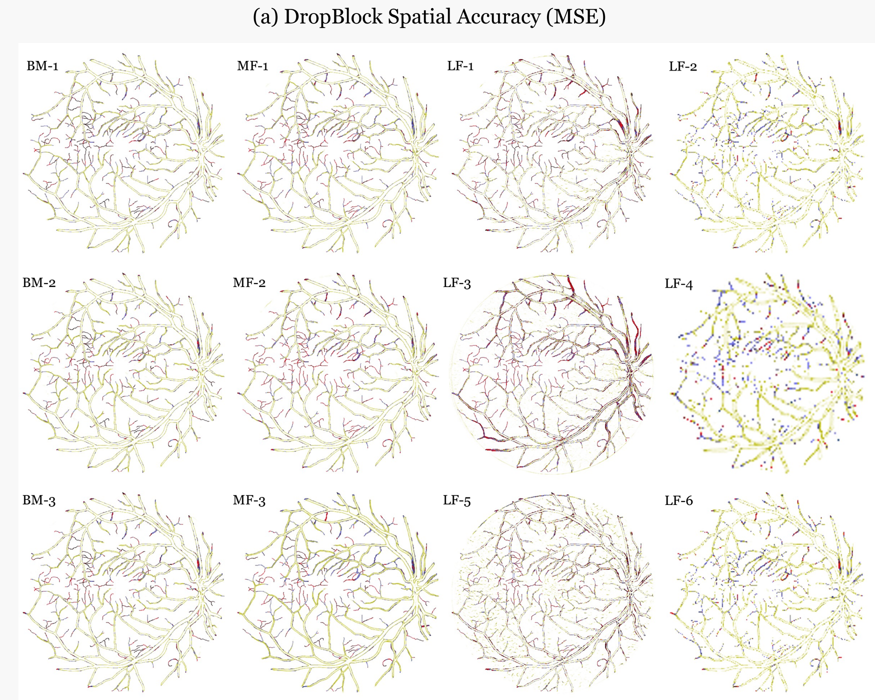
Figure 5: F1 (Dice) , AUROC, and Accuracy scores for each model and its DropBlock mean version.



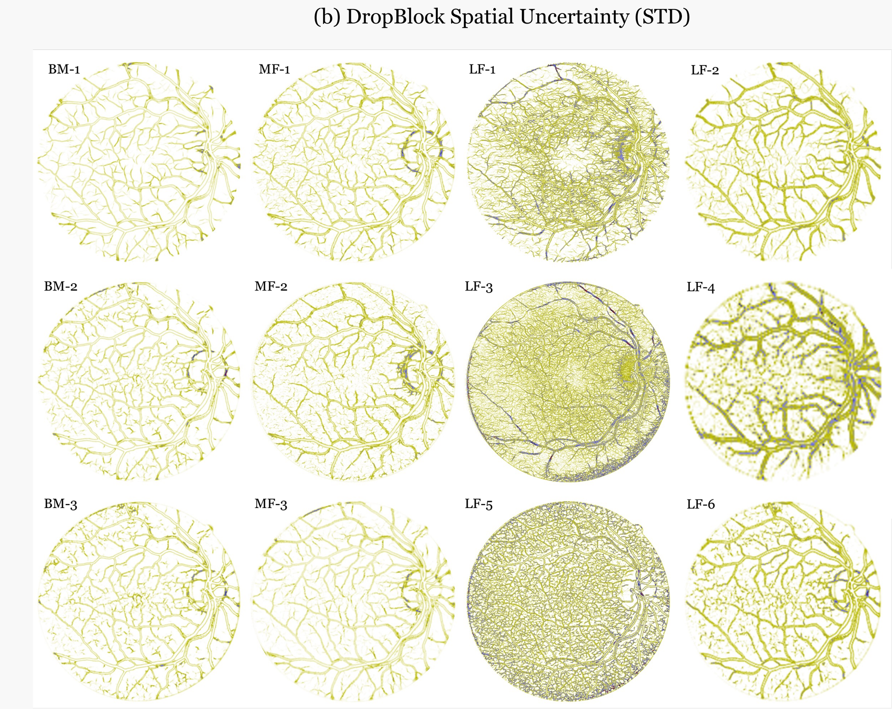
(a) DropBlock Spatial Accuracy (MSE)



(b) DropBlock Spatial Uncertainty (STD)



(c) Rotation Spatial Accuracy (MSE)



(d) Rotation Spatial Uncertainty (STD)

Figure 6: Measures of Accuracy and Uncertainty for DropBlock and Rotation. MSE (a, c) and STD (b, d) images are obtained from the image with the respective highest average measurement. From top to bottom, column (1) contains BM-1 to 3, column (2) contains MF-1 to 3, column (3) contains LF-1, 3, 5, and column (4) contains LF-2, 4, 6.

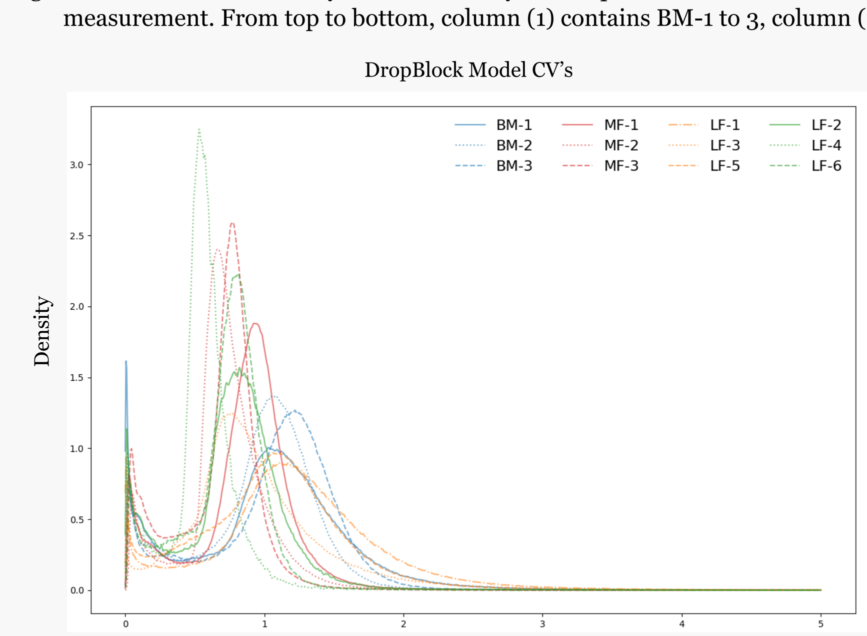


Figure 7: Distribution of pixel-wise aggregated Coefficient of Variation (CV) for DropBlock

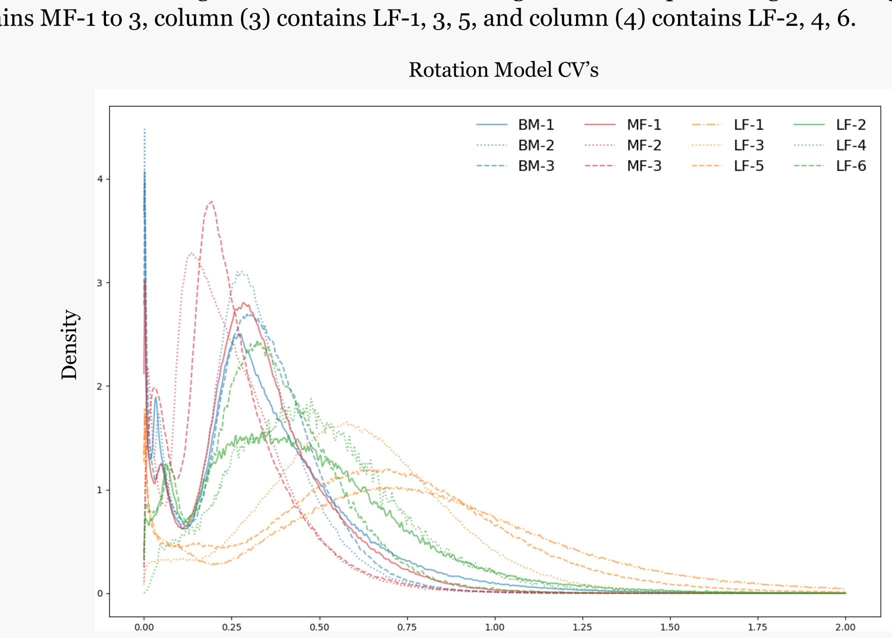


Figure 8: Distribution of pixel-wise aggregated Coefficient of Variation (CV) for Rotations

Key Findings

- In the MF networks, STDs on small vessels is amplified by inputs at smaller resolutions. The network is confused by the combination of HF and LF images, unable to associate a resolution to a given input and coherently analyze it. However, a comparison between the BM2-MF1 and BM3-MF2 accuracy, reveals that not much has been gained in terms of accuracy by using a multi-fidelity training dataset
- Rotational uncertainty leads to prediction uncertainties equal to roughly 25% - 75% with respect to those produced by DropBlock. It is also able to better differentiate models with different image resolutions.
- Reducing the resolution of the input images has a greater impact on the network accuracy than reducing the training set size. Two HF training images (72 augmented images) appear sufficient to produce satisfactory MSE results for both BM and MF networks.

Methodology

Data Augmentation – Beginning with 20 images, we adopt a 70/30 (14, 6) split for training/validation. During preprocessing, all images were gray-scaled. For each training image, 36 new augmented images were constructed using random rotations and vertical, horizontal, and diagonal flips. During model training, dataset reduction and image resizing was performed on the fly as needed, with image mappings to ensure consistent information loss.

Architecture Modifications – Due to memory limitations, effective training batch sizes were limited to 1. Rather than using BatchNorm which offers no practical effect at batch size 1, Group Normalization (GN) with a size of 32 was chosen due to its advantages at smaller batch sizes [5] . Furthermore, a DropBlock layer with block size 7, drop probability 0.15, and 1500 step linear scheduling was placed after each convolution between GN and ReLU and after skip connections as suggested in [2].

Training – Each model was trained using a Stochastic Gradient Descent (SGD) optimizer with a model-specific tuned learning rate (lr) and a momentum of 0.99. Each model was given 200 epochs to run, but lr reduction on plateau and early stopping were implemented allowing training to finish when validation losses failed to decrease. Model weights were saved when validation losses were lowest.

Table – We trained 12 models as described below.

Model	Test ID	Description	Resolution/Dataset Size	Tr/Val w.r.t
Base	BM-1	HF inputs, Full Dataset	14: 584 x 565	HF Truth
	BM-2	HF inputs, Reduced Dataset	4: 584 x 565	HF Truth
	BM-3	HF inputs, Reduced Dataset	2: 584 x 565	HF Truth
Multi-Fidelity	MF-1	Uniform HF/LF inputs, Full Dataset	4: 584 x 584, 5: 256 x 256, 5: 128 x 128	HF Truth
	MF-2	Ratioed HF/LF inputs, Full Dataset	2: 584 x 584, 4: 256 x 256, 8: 128 x 128	HF Truth
	MF-3	HF/Reduced Quality HF inputs, Full Dataset	Ratio like MF-2. LF images resized to 584 x 584	HF Truth
Low-Fidelity	LF-1	LF inputs, Full Dataset	14: 256 x 256	HF Truth
	LF-2	LF inputs, Full Dataset	14: 256 x 256	LF Truth
	LF-3	LF inputs, Full Dataset	14: 128 x 128	HF Truth
	LF-4	LF inputs, Full Dataset	14: 128 x 128	LF Truth
	LF-5	LF inputs, Reduced Dataset	4: 256 x 256	HF Truth
	LF-6	LF inputs, Reduced Dataset	4: 256 x 256	LF Truth

Uncertainty

Monte Carlo DropBlock Uncertainty – By enabling DropBlock layers during evaluation, the random drops enable a model to output slightly different results on identical inputs. This allows us to perform a Monte Carlo simulation and determine a distribution of outputs for each pixel. [3].

Monte Carlo Rotational Uncertainty – Since convolutional network are not rotationally equivariant, by evaluating each model on multiple rotations of the same image and rotating back by the corresponding degree, we achieve a distribution effect like Monte Carlo DropBlock.

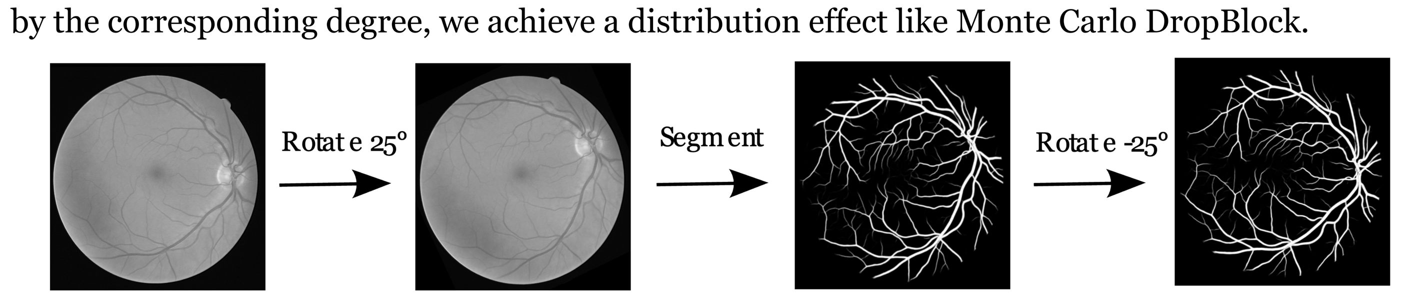


Figure 4: The steps for Rotational Uncertainty

Future Work

In the future, we intend to design a training technique on the U-Net that can effectively utilize multi-fidelity nature of inputs.

Acknowledgements

We would like to acknowledge the National Science Foundation for supporting us through a REU Supplement to NSF CAREER Award #1942662 “Bayesian Inference Networks for Model Ensembles under the supervision of Program Manager Dr. Alan Sussman. Special thanks to Rachel Shen for editing and reviewing this poster.

References

[1] - O. Ronneberger, P. Fischer, and T. Brox. **U-net: Convolutional networks for biomedical image segmentation**. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[2] - Ghiasi, T.-Y. Lin, and Q. V. Le. **Dropblock: A regularization method for convolutional networks**. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 31. Curran Associates, Inc., 2018.

[3] - M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. W. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi. **A review of uncertainty quantification in deep learning: Techniques, applications and challenges**. CoRR, abs/2011.06225, 2020.

[4] - Drive - grand challenge.

[5] - Y. Wu and K. He. **Group normalization.** CoRR, abs/1803.08494, 2018