A Proof of Corollary 1

Proof. A sample $\langle \mathbf{x}^0, \mathbf{y}^0 \rangle$ follows A2, and we have its corresponding $\mathbf{x}_l^0$. Since $\epsilon \geq d(\mathbf{x}^0, \mathbf{x}_l^0)$ (as required by the Corollary), $\mathbf{x}_l^0$ is one of the perturbed examples that will be tested by Equation 2. Therefore, for model θ:

- If $\mathbf{y}^0 \neq f(\mathbf{x}^0; \theta)$, Equation 1 will be evaluated as 0, thus not accurate.
- If $\mathbf{y}^0 = f(\mathbf{x}^0; \theta)$, then $\mathbf{y}^0 \neq f(\mathbf{x}_l^0; \theta)$ (because of A2). Equation 2 will be evaluated as 0 (because $\mathbf{x}_l^0$ is one of the perturbed examples, as mentioned above), thus not robust.

□

B Extra Empirical Results

B.1 Colorful Images

We also experimented with a subset of ImageNet, which was constructed by merging related images into nine major classes, leading to a nine-class classification dataset with roughly 1350 images per class. We used a ResNet-50 for the experiments.

We repeated all the experiments with the the above experiment set-up and show the results. Overall, the results demonstrate a similar result.

B.1.1 Rethinking Data before Rethinking Generalization

LFC			HFC		
r	train acc.	test acc.	r	train acc.	test acc.
25	0.9315	0.5477	25	0.8730	0.1908
50	0.9681	0.6477	50	0.7655	0.1515
100	0.9759	0.8300	100	0.6476	0.1531
150	0.9717	0.8415	150	0.2416	0.1123

Table 3: table1

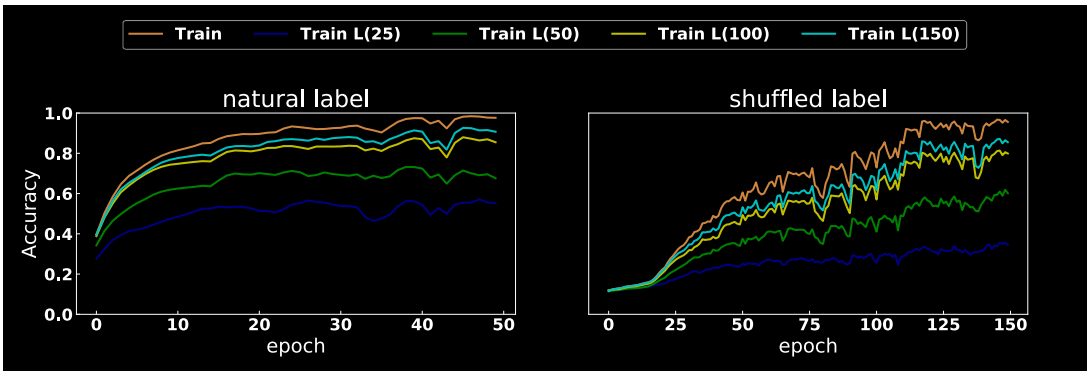


Figure 10: Training curves of the original label case (50 epoches) and shuffled label case (150 epoches), together plotted with the low-frequent counterpart of the images. All curves in this figure are from train samples.

B.1.2 Heuristics

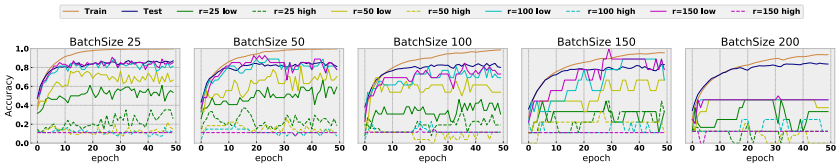


Figure 11: Plots of accuracy of different batch sizes along the epoches for train

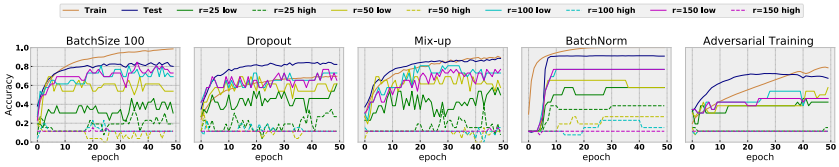


Figure 12: Plots of accuracy of different heuristics along the epoches for train, test data, as well as LFC and HFC with different radii.

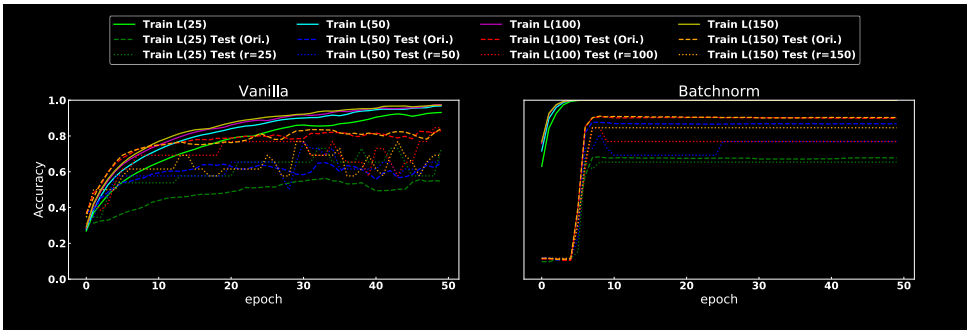
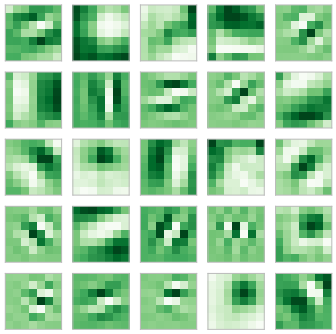
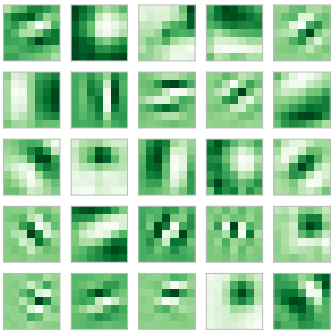


Figure 13: Comparison of models with vs. without BatchNorm trained with LFC data.

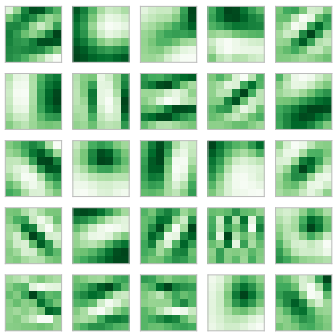
B.1.3 Adversarial Attack and Defense



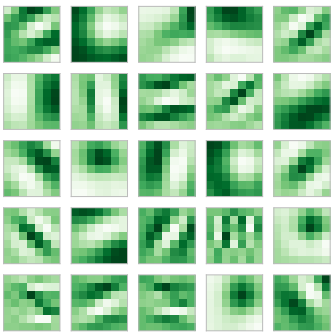
(a) convoluational kernels of $\mathsf{M}_{\mathsf{natural}}$



(b) convoluational kernels of $\mathsf{M}_{\mathsf{adversarial}}$



(c) convoluational kernels of $\mathsf{M}_{\mathsf{natural}}(r{=}1.0)$



(d) convoluational kernels of $\mathsf{M}_{\mathsf{adversarial}}(r{=}1.0)$

Figure 14: Visualization of convolutional kernels (64 kernels each channel $\times$ 3 channels at the first layer) of models.

	Clean	FGSM			PGD		
		$\epsilon = 0.03$	$\epsilon = 0.06$	$\epsilon = 0.09$	$\epsilon = 0.03$	$\epsilon = 0.06$	$\epsilon = 0.09$
$\mathsf{M}_{\mathsf{natural}}$	**0.821**	0.004	0.002	0.002	0.010	0.011	0.013
$\mathsf{M}_{\mathsf{natural}}(\rho = 0.10)$	0.813	0.009	0.005	0.003	0.015	0.010	0.007
$\mathsf{M}_{\mathsf{natural}}(\rho = 0.25)$	0.786	0.016	0.010	0.008	0.020	0.016	0.012
$\mathsf{M}_{\mathsf{natural}}(\rho = 0.50)$	0.767	0.119	0.119	0.118	0.127	0.127	0.125
$\mathsf{M}_{\mathsf{natural}}(\rho = 1.0)$	0.760	**0.370**	**0.368**	**0.368**	**0.382**	**0.382**	**0.378**
$\mathsf{M}_{\mathsf{adversarial}}$	**0.684**	0.122	0.033	0.012	0.034	0.038	0.029
$\mathsf{M}_{\mathsf{adversarial}}(\rho = 0.10)$	0.651	0.134	0.050	0.030	0.043	0.037	0.033
$\mathsf{M}_{\mathsf{adversarial}}(\rho = 0.25)$	0.613	0.15	0.08	0.07	0.032	0.031	0.019
$\mathsf{M}_{\mathsf{adversarial}}(\rho = 0.50)$	0.603	0.189	0.126	0.114	0.060	0.047	0.036
$\mathsf{M}_{\mathsf{adversarial}}(\rho = 1.0)$	0.590	**0.241**	**0.184**	**0.164**	**0.086**	**0.084**	**0.081**

Table 4: Prediction performance of models against different adversarial attacks with different ϵ.

B.2 Greyscale Images

We also repeat the major results with the greycale images (MNIST and FashionMNIST).

B.2.1 Rethinking Data before Rethinking Generalization

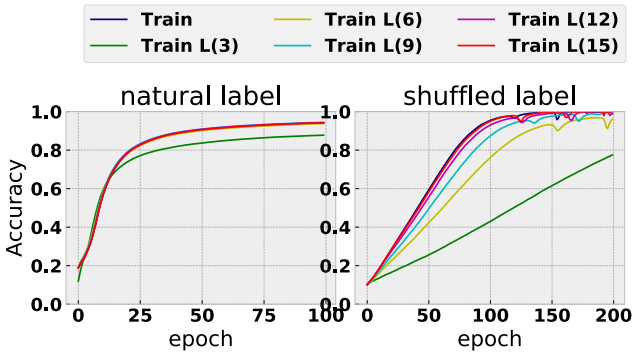


Figure 15: Experiments of natural label and shuffled labeled over MNIST data set

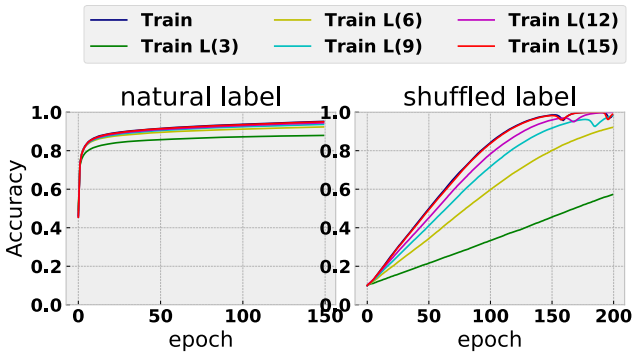


Figure 16: Experiments of natural label and shuffled labeled over FashionMNIST data set

B.2.2 Heuristics

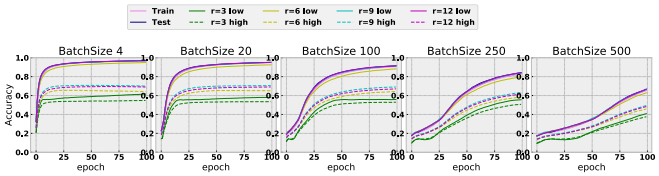


Figure 17: Experiments of batchsize over MNIST data set

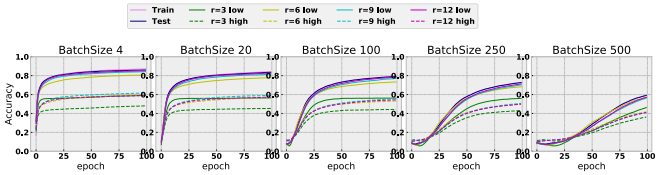


Figure 18: Experiments of batchsize over FashionMNIST data set

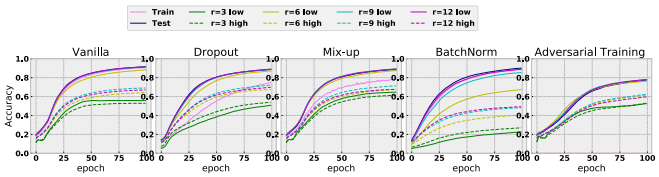


Figure 19: Experiments of heuristics for MNIST data set.

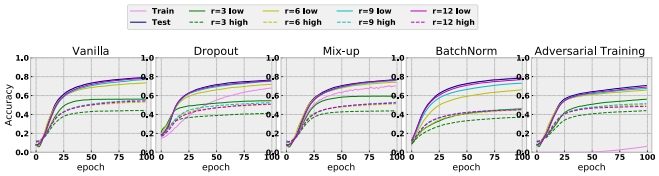


Figure 20: Experiments of heuristics for FashionMNIST data set.

B.2.3 Adversarial Attack & Defense

Figure 21: Convolutional Kernel of the models trained over MNSIT

Figure 22: Convolutional Kernel of the models trained over FashionMNSIT