Supplementary Material of Feature Quantization for Defending Against Distortion of Images

Anonymous CVPR submission

Paper ID 1001

1. Analysis of Divergence of Distributions under Deformations

In this section, we provide additional results for analysis of divergence of higher moments of feature distributions under deformation. As argued in the Figure 1 of the main text, deformations will result in shift of skewness and kurtosis as well as that of mean and variance. In addition, these types of shift cannot be successfully reduced by normalization methods. In Figure 1, we depict the amount of shift of skewness and kurtosis between feature distributions obtained using original and deformed images. We employ VGG16 [1] for evaluation of the models, and measure the amount of change of skewness and kurtosis for neurons in 6 different layers. 5,000 original and deformed images are used to calculate feature distributions. Note that the skewness of a normal distribution is defined on $(-1, 1)$, and the results indicate a decent amount of shift of skewness and kurtosis.

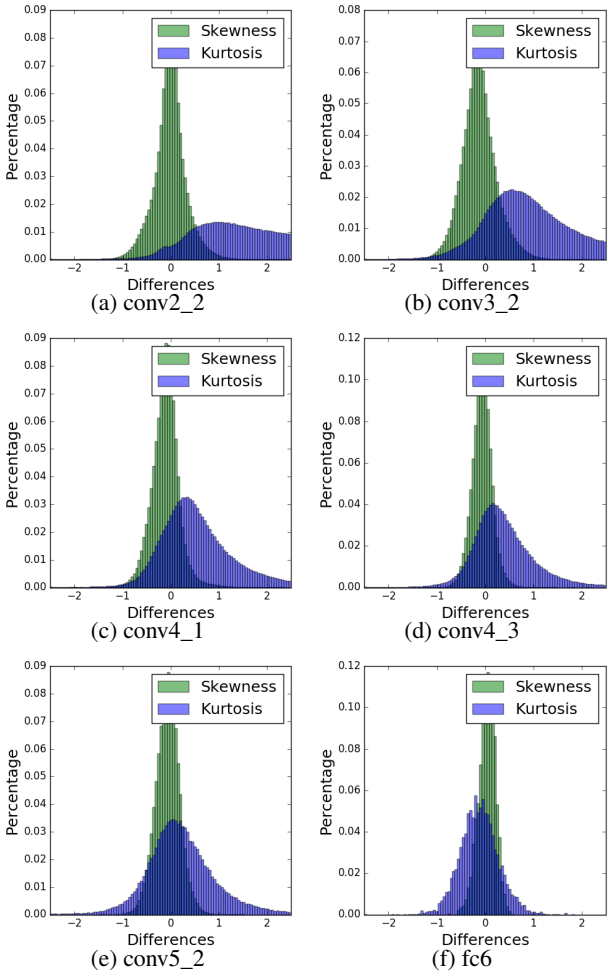


Figure 1: Change of skewness and kurtosis for neurons used at different layers. Horizontal axis shows difference of skewness and kurtosis between distributions obtained from original and deformed images.

2. Analysis of Effect of Regularization to Classification Error

In this section, we examine the effect of L1/L2 regularization towards exponent α used in the proposed power function. We employ a small network (details are given in Table 2) to perform standard classification tasks on the Cifar-10/100 datasets. The average classification errors with their std. deviations are given in Table 1. It is observed that, L2 regularization provides better results for the Cifar-10, while models that employ L1 regularization perform slightly better for the Cifar-100. Consequently, we infer that L1 regularization enforces a sparse distribution on , where majority of the channels act as an identical mapping. Therefore, we observe a reduction of over-fitting in the Cifar-100 (larger number of classes with less training samples per class) compared to the Cifar-10.

Table 1: Classification error (%) obtained using different types of regularization on the exponent α for the Cifar-10/100.

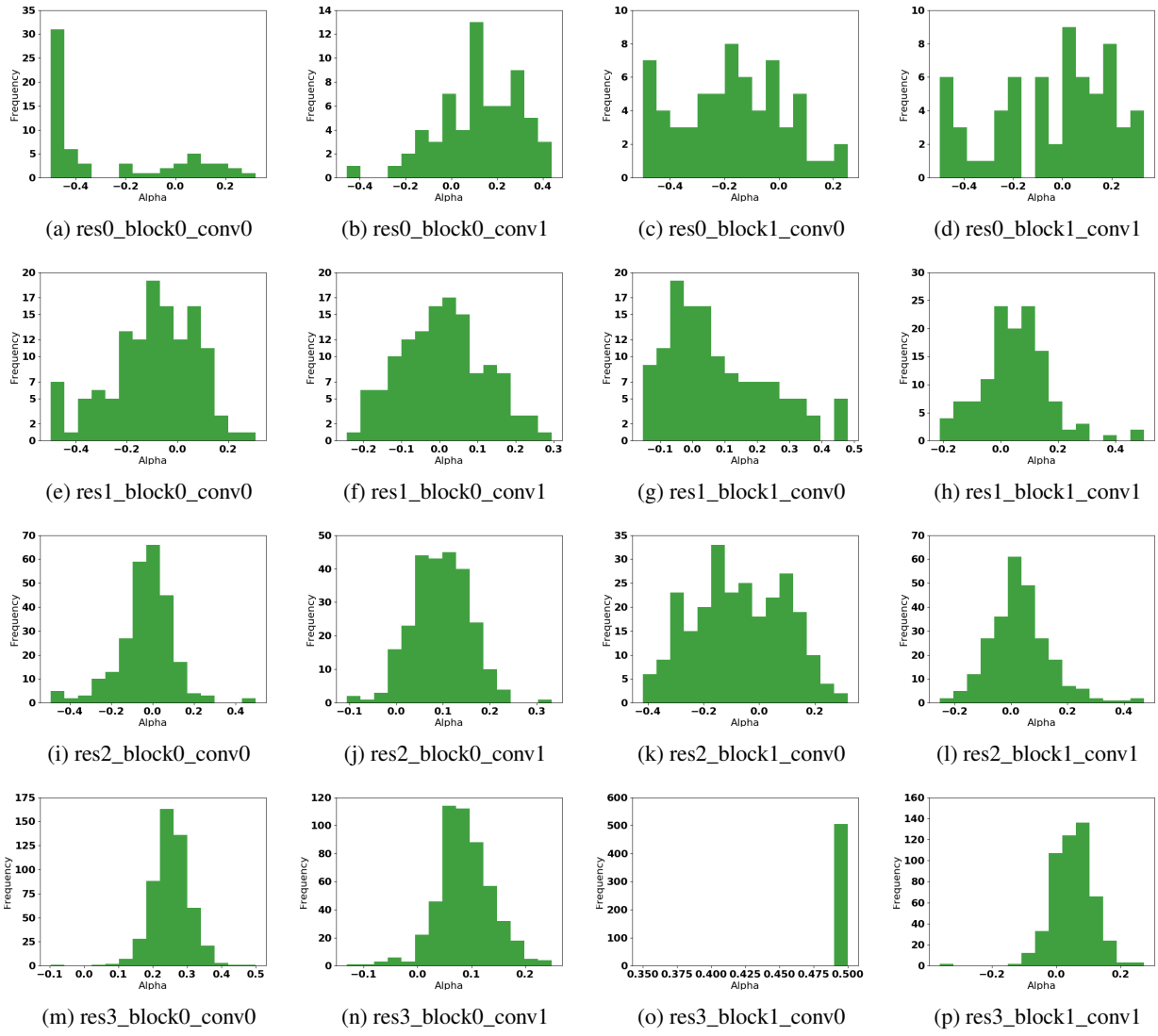
Models	**Cifar-10**	**Cifar-100**
Base	18.31 ± 0.15	46.63 ± 0.28
L1 regularization		
+POW-1	18.03 ± 0.19	46.21 ± 0.09
+POW-2	18.39 ± 0.18	46.44 ± 0.31
+POW-4	18.58 ± 0.34	46.45 ± 0.35
+POW-8	18.41 ± 0.24	45.99 ± 0.39
+POW-16	18.10 ± 0.07	46.43 ± 0.49
L2 regularization		
+POW-1	17.99 ± 0.17	46.52 ± 0.22
+POW-2	17.92 ± 0.14	46.53 ± 0.34
+POW-4	17.99 ± 0.17	46.52 ± 0.24
+POW-8	17.99 ± 0.17	46.46 ± 0.34
+POW-16	18.06 ± 0.32	46.39 ± 0.42

Table 2: CNN configurations for Cifar-10/100 used in Section 2. The convolution layer parameters are denoted by conv –<RF size>–<number of output channels>. All the conv. layers are set to be stride 1 equipped with pad 1. The conv. layers in the middle block are equipped with proposed power function.

Module
conv – 3×3 – 32 BN & ReLU
max-pooling – 2×2 – stride 2
conv – 3×3 – 64 BN & ReLU conv – 3×3 – 64 BN & ReLU
max-pooling – 2×2 – stride 2
conv – 3×3 – 10/100 global ave-pooling soft-max classifier

3. Analysis of Distribution of α

In this section, we examine distributions of learned exponents for +POW model used in Section 3.2 of the main text. We visualize the distributions within convolution layers of different blocks of the ResNet-18+POW-1 model, and the histograms are shown in Figure 2. It can be observed that, the majority of learned are distributed around 0. Strong quantization effects are obtained for layers such as res0_block0_conv0. On the other hand, for layers such as res3_block1_conv0, extracted features are usually sparser than those extracted in the lower layers. In addition, for these layers, employment of provides an amplification effect on features, that is, activations larger than 1 are amplified, while those smaller than 1 are quantized to 0.



(a) res0_block0_conv0 (b) res0_block0_conv1 (c) res0_block1_conv0 (d) res0_block1_conv1

(e) res1_block0_conv0 (f) res1_block0_conv1 (g) res1_block1_conv0 (h) res1_block1_conv1

(i) res2_block0_conv0 (j) res2_block0_conv1 (k) res2_block1_conv0 (l) res2_block1_conv1

(m) res3_block0_conv0 (n) res3_block0_conv1 (o) res3_block1_conv0 (p) res3_block1_conv1

Figure 2: Distributions of learned α for the ResNet-18+POW-1 model utilized in Section 3.2 of the main text. The *res_block_conv* denotes the index of the layers.

4. Analysis of Quantization Resolution and Number of Splits

In this section, we provide experimental results for employment of different quantization resolutions in +SF models and different number of splits for +POW models. For this propose, we quantize the ResNet-18/50 using floor function with a resolution of 10. In addition, we employ a small 10-layer plain network (given in Table 4) for examining the robustness of +POW model under different split configurations. We use the same approach for evaluating the robustness as utilized in Section 3.2 of the main text, and the results are provided in Table 3. The results show that quantization of the models with lower resolution is helpful in dealing with statistical noises. However, the lack of expressive power also limits their performance towards clean or Jpeg compressed images. On the other hand, the increased number of splits for +POW models seems to cause over-fitting problems and performance decrease in general.

Table 3: Classification accuracy (Top-5 accuracy(%)) obtained using distorted images.

Models	Clean	Motion Blur	Jpeg Comp.	Salt & Pepper	NGRN Noise	In-paint.	Tar. Occ.[a]
ResNet18	90.3	31.0	38.0	27.6	24.9	31.4	48.4
+SF-10	90.3	31.8	36.0	34.6	31.7	31.8	47.4
+SF-100	90.4	32.3	43.8	26.7	24.7	31.7	45.9
ResNet50	93.4	39.2	51.2	52.9	50.7	35.6	51.9
+SF-10	92.9	33.8	53.1	55.8	51.8	30.9	49.6
+SF-100	93.5	41.2	53.2	52.7	51.3	28.3	53.9
Plain-10	83.5	41.7	32.1	22.1	22.3	32.5	42.3
+POW-1	83.8	39.4	31.5	27.9	27.7	33.0	41.8
+POW-2	83.9	38.3	29.2	25.4	25.4	33.9	40.2
+POW-4	83.6	37.4	31.1	24.2	26.8	33.2	41.9
+POW-8	83.9	38.9	31.3	23.8	25.1	33.7	43.0

[a] Top-1 accuracy is reported.

Table 4: The configuration of the Plain-10 models used in Section 4. The convolution layer parameters are denoted by conv –<RF size>–<number of output channels>. All the conv. layers are set to be stride 1 equipped with pad 1. All the conv. layers are followed with a combination of BN-ReLU.

Module
conv – 7×7 – 64
max-pooling – 3×3 – stride 2
conv – 3×3 – 64 – stride 2 conv – 3×3 – 64
conv – 3×3 – 128 – stride 2 conv – 3×3 – 128
conv – 3×3 – 256 – stride 2 conv – 3×3 – 256
conv – 3×3 – 512 – stride 2 conv – 3×3 – 512
global ave-pooling fc – 1000 soft-max classifier

5. Additional Results on ResNet-50

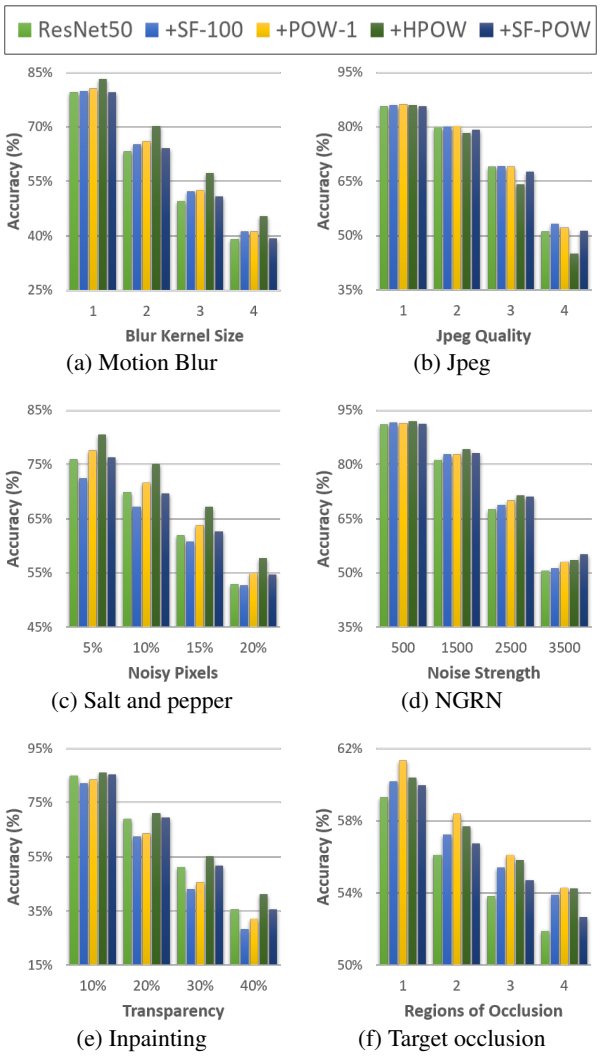
We provide additional results on classification performance for employment of distortion methods with different strength (see Figure 4 in the main text), using ResNet-50 as the base model. These figures depict that the improved robustness against both minor and heavy distortions can be verified in ResNet-50 based models as well. Moreover, we observe that for targeted occlusion, the proposed methods implemented in ResNet-50 out-performed the base ResNet-50 model regardless of the strength of distortion. On the other hand, the boost for Jpeg compression is relatively trivial compared to that observed with the ResNet-18 models. We argue that, since most of the training samples are encoded using Jpeg compression, a model with increased capacity would be helpful in dealing with the compression artifacts inherently (e.g. the ResNet-50 performs $\sim$13% better compared with the ResNet-18 for Jpeg5). As a consequence, the additional expressive power obtained from models equipped with power function may increase complexity of the space of hypothesis functions learned by the networks and cause over-fitting. We should also be aware that, the ResNet-50 is still fragile towards occlusions (targeted and inpainting), even its better generalization performance (accuracy) regarding clean images seems to be helpful for defending against statistical distortions. Therefore, we conjecture that the best approach which can be used to deal heavily distorted images is still to train with them, if we have the prior knowledge of the types of distortions in test images.

6. Additional Visualizations on Object Detection

We provide additional visualizations of object detection task using Voc Pascal 2007 dataset, the results are given in Figure 4.

References

[1] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1



(a) Motion Blur (b) Jpeg (c) Salt and pepper (d) NGRN (e) Inpainting (f) Target occlusion

Figure 3: Classification accuracy (Top-5 accuracy(%)) obtained using images with different strength of distortions.