Diffusion-Based Adaptation for Classification of Unknown Degraded Images

Supplementary Material

1. Experiment settings

1.1. Classifier training

We discuss in detail the hyperparameters utilized for training different classifier models in our study. For C_{clean}, we use pre-trained models provided by MMCV [1]. We conduct training for C_{adapt} and C_{deg} models spanning 100 epochs, utilizing the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.001 and weight decay set to $1e-4$. Furthermore, we apply a multi-step learning rate scheduler with milestones at epochs 50 and 75, adjusting the learning rate by 0.1 at each milestone.

For the training of distilled classifiers, *i.e.*, DC_{adapt} and DC_{deg}, we initially train all individual teacher networks ($Teacher$) on known degradations for 50 epochs. We also employ the SGD optimizer with identical settings as aforementioned, along with a multi-step learning rate scheduler set at epochs 15, 30, and 45, where each of these milestones sees an adjustment of the learning rate by a factor of 0.1. Subsequently, we train the student network with an identical optimizer and learning rate scheduler as the teacher network. We initialize teacher and student networks with C_{clean} pre-trained weights. To ensure a fair comparison between the training of C and DC classifiers, we set the number of epochs for C classifiers to match the combined epochs of the two-stage training process for both the teacher and student (DC) classifiers. Hyperparameter settings for all classifiers training are the same as discussed above over all datasets or backbones discussed in this study. Since we take inspiration from FusionDistill [2] for our distilled classifier, hyperparameter settings are similar to theirs.

Table 1. Hyperparameter tuning of parameter α on the Imagenet dataset.

Loss Weights	DC_{adapt}	DC_{deg}
$\alpha = 1$	72.80	81.67
$\alpha = 10$	**72.81**	81.86
$\alpha = 20$	72.79	81.98
$\alpha = 50$	72.75	**82.05**
$\alpha = 100$	72.75	82.02

1.2. Loss Weights Tuning

As discussed in our study, we only have a single hyperparameter in our proposed method, *i.e.*, α weight of the consistency loss function. We optimize consistency loss for synthetically prepared degraded and adapted images on the Imagenet dataset separately for DC_{adapt} and DC_{deg} as shown in Table 1. Alpha values vary between 1 and 100, given that the cosine similarity loss values are typically very small. Optimal values of $\alpha = 50$ for DC_{adapt} and $\alpha = 10$ for DC_{deg}.

2. Ablation Study: Experimental Results

2.1. Single Degradations

We provide additional experimental results for single degradation, *i.e.*, CIFAR-10-C [5] dataset on ResNet-18 [4] backbone as shown in Table 2 to support our claims further. Expectedly, the C_{clean} method performs the worst following DDA [3], given that both methods contain only clean image classifiers. Still, C_{clean} performs well on some degradations such as brightness, fog, and snow. On the other hand C_{deg}, DC_{deg} and DiffAUD outperforms C_{clean} and DDA [3] on most of the degradations. Our proposed method, DiffAUD, performs best for 10 out of 15 corruptions and outperforms other methods. The performance of ResNet-18 overall is quite similar to ResNet-50 on the CIFAR-10-C dataset compared to the existing methods.

2.2. Sequential Degradations

Moreover, we provide the experimental results for sequential degradations on the CIFAR-10-SEQ-C dataset as shown in Table 3. Similar to the Imagenet-SEQ-C dataset, DiffAUD outperforms other existing methods. However, known degradations-based methods outperform C_{clean} by significantly improving compared to Imagenet-SEQ-C. For example, C_{deg} DC_{deg} improves the performance by about 24%-25% on ResNet-18 and subsequently by about 23% over all degradations. In addition, we can see roughly

Table 2. Classification accuracy for CIFAR-10-C dataset with different corruptions averaged over all severities on ResNet-18 backbone.

Method	bright	contrast	defocus	elastic	fog	frost	*gauss*	glass	impulse	*jpeg*	motion	pixel	shot	snow	zoom	mean
C_{clean}	**93.53**	78.89	83.93	85.33	**88.69**	79.40	51.42	53.16	56.59	80.71	78.68	77.94	62.88	82.91	80.32	75.62
C_{deg}	91.09	82.22	90.07	86.15	87.62	86.95	87.81	77.69	83.99	88.18	85.24	87.29	88.99	85.60	88.48	86.49
DC_{deg}	91.48	**84.24**	**90.82**	86.56	88.48	**87.41**	87.91	76.17	83.95	88.17	86.25	86.80	89.15	85.89	89.36	86.84
DDA [3]	89.35	68.91	86.15	84.44	77.07	80.24	80.88	78.38	80.95	84.28	82.77	84.91	82.83	82.66	84.30	81.87
Ours	90.87	82.08	90.81	**87.40**	86.81	86.46	**88.67**	**82.50**	**86.30**	**88.66**	**87.18**	**88.73**	**89.62**	**86.23**	**89.65**	**87.46**

Table 3. Classification accuracy for CIFAR-10-SEQ-C sequential degradation dataset with different severity levels on ResNet-18 and ResNet-50 backbones.

Method	ResNet-18				ResNet-50			
	weak	medium	strong	mean	weak	medium	strong	mean
C_{clean}	76.49	54.37	41.09	57.32	78.77	57.86	43.67	60.10
C_{deg}	89.08	82.92	74.50	82.17	89.60	84.05	75.56	83.07
DC_{deg}	89.19	82.57	73.43	81.73	89.59	83.42	74.19	82.40
DDA [3]	83.70	79.42	76.12	79.75	84.05	79.79	76.11	79.98
Ours	**89.30**	**85.53**	**80.48**	**85.10**	**90.13**	**85.95**	**80.71**	**85.60**

5% improvement from DiffAUD comparison to DDA. It demonstrates that the known degradation methods significantly help improve the performance of sequential degradations. At the same time, please note that the single degradations part of the training of the classifiers is the same degradations applied in the sequential degradations, *i.e.*, JPEG, GBlur, and AWGN. It shows that even if we apply single degradations on images for training, it can still provide robustness against unknown sequential degradations.

3. Sample training images

To show images utilized for training in our study, we show a few sample images in Figures 1 to 3. All sample figures include clean images, and pairs of their corresponding known degraded and adapted images from the diffusion model. Diffusion models adapt AWGN degradation images appropriately to the clean image domain; sometimes, the adapted images look sharper than the original clean image. On the other hand, images degraded with GBlur and JPEG are not perfect. JPEG compressed image adaptation quality is not so bad in case of low severity; however, with higher severity, adapted images lose some details, such as the eyes of the bird in Figure 1 for JPEG quality factor of 20. Moreover, the diffusion model struggles to adapt the images from GBlur degraded images even with low severities. It demonstrates the need for specialized classifiers, as we proposed in DiffAUD, to classify imperfect adapted images from the diffusion models.

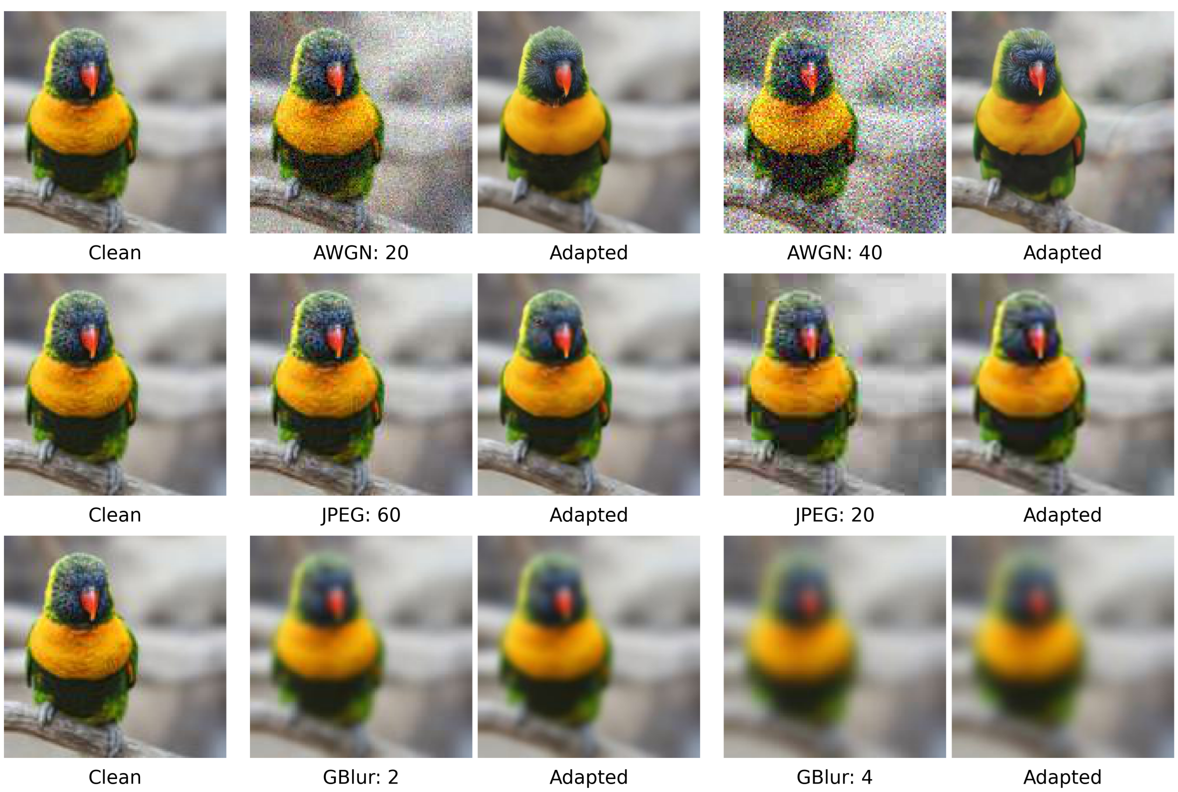

Figure 1. Sample figure of a bird with different known training degradations and severity levels.

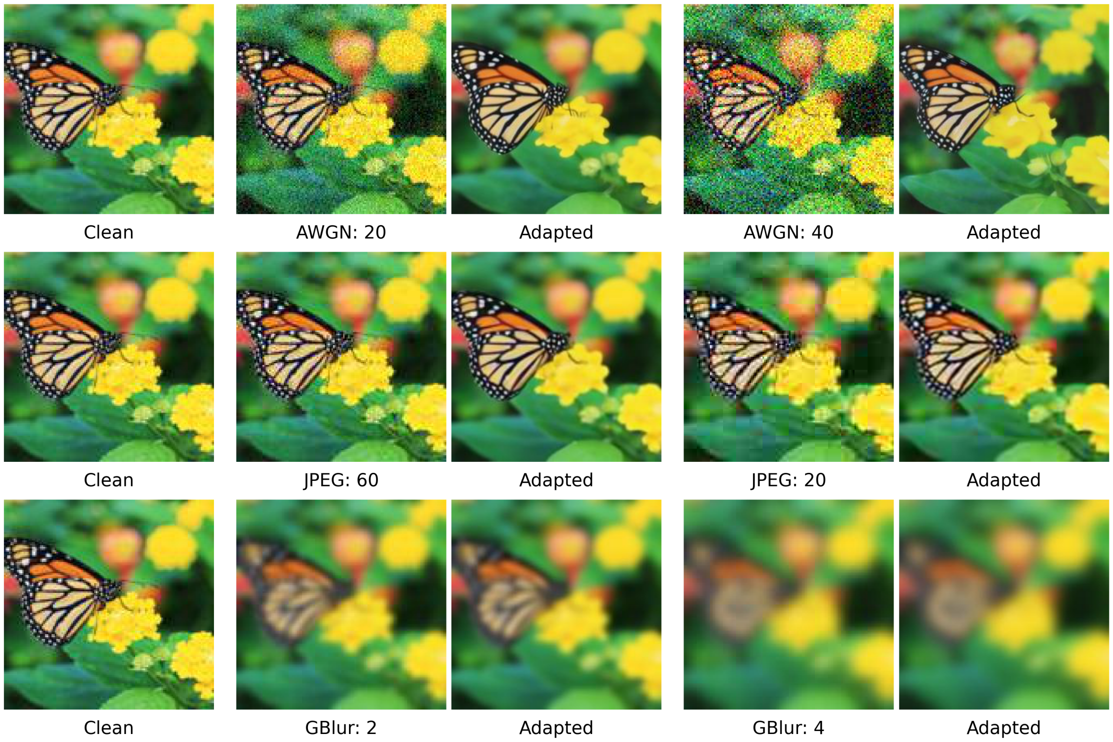

Figure 2. Sample figure of a butterfly with different known training degradations and severity levels.

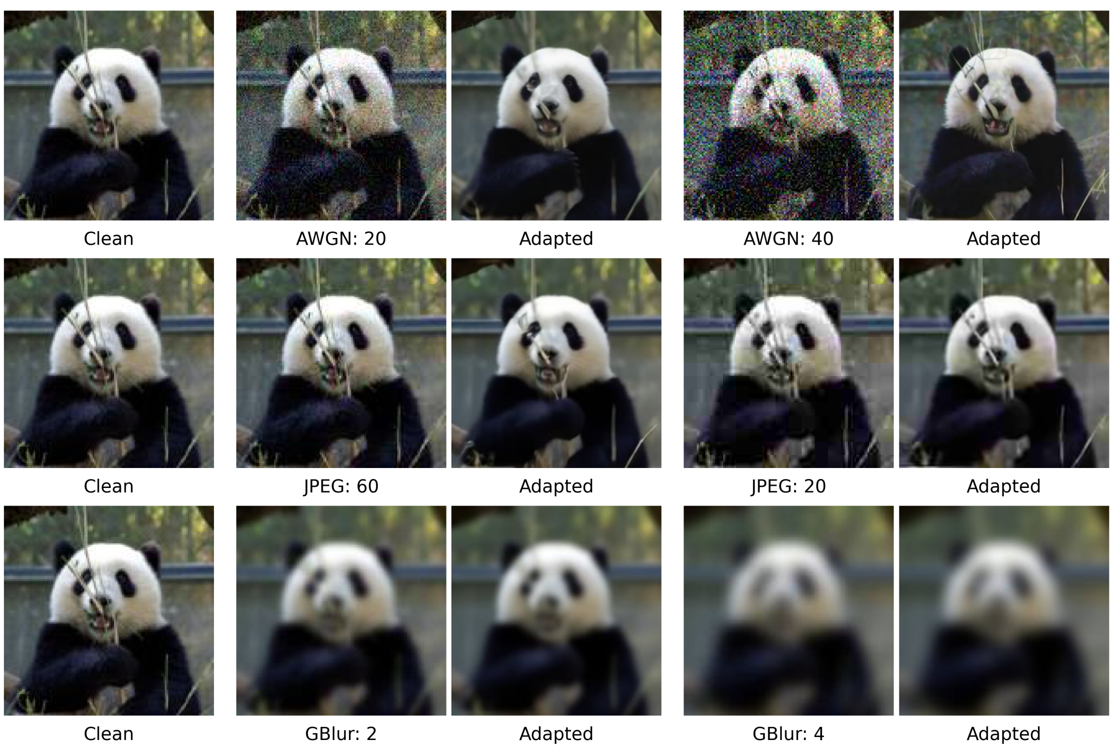

Figure 3. Sample figure of a panda with different known training degradations and severity levels.

References

[1] MMCV Contributors. MMCV: OpenMMLab computer vision foundation, 2018. 1

[2] Dinesh Daultani and Hugo Larochelle. Consolidating separate degradations model via weights fusion and distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pages 440–449, 2024. 1

[3] Jin Gao, Jialing Zhang, Xihui Liu, Trevor Darrell, Evan Shelhamer, and Dequan Wang. Back to the source: Diffusion-driven adaptation to test-time corruption. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11786–11796, 2023. 1, 2

[4] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2015. 1

[5] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *International Conference on Learning Representations (ICLR)*, 2019. 1