Deep Degradation Prior for Low-quality Image Classification

Yang Wang[1], Yang Cao[1,*], Zheng-Jun Zha[1,†], Jing Zhang[2], Zhiwei Xiong[1]
[1]University of Science and Technology of China, Hefei, China
[2]UBTECH Sydney Artificial Intelligence Centre, The University of Sydney, Sydney, Australia
ywang120@mail.ustc.edu.cn, forrest,zhazj,zwxiong@ustc.edu.cn, jing.zhang1@sydney.edu.au

Abstract

State-of-the-art image classification algorithms building upon convolutional neural networks (CNNs) are commonly trained on large annotated datasets of high-quality images. When applied to low-quality images, they will suffer a significant degradation in performance, since the structural and statistical properties of pixels in the neighborhood are obstructed by image degradation. To address this problem, this paper proposes a novel deep degradation prior for low-quality image classification. It is based on statistical observations that, in the deep representation space, image patches with structural similarity have uniform distribution even if they come from different images, and the distributions of corresponding patches in low- and high-quality images have uniform margins under the same degradation condition. Therefore, we propose a feature de-drifting module (FDM) to learn the mapping relationship between deep representations of low- and high- quality images, and leverage it as a deep degradation prior (DDP) for low-quality image classification. Since the statistical properties are independent to image content, deep degradation prior can be learned on a training set of limited images without supervision of semantic labels and served in a form of "plugging-in" module of the existing classification networks to improve their performance on degraded images. Evaluations on the benchmark dataset ImageNet-C demonstrate that our proposed DDP can improve the accuracy of the pre-trained network model by more than 20% under various degradation conditions. Even under the extreme setting that only 10 images from CUB-C dataset are used for the training of DDP, our method improves the accuracy of VGG16 on ImageNet-C from 37% to 55%.

1. Introduction

Recent years witness remarkable progresses of image classification task [46, 39], mainly with deep convolutional neural network trained on large-scale datasets like ImageNet [8]. Since the collected images of the existing datasets are usually free of degradation, such as low visibility [41, 12, 34], color cast [42, 2], and overexposure, evaluation of standard neural network-based methods for image classification shows a significant drop in classification accuracy when applied to low-quality images [13].

In practice, one solution for this problem is to firstly improve the visibility of degraded images by image enhancement methods, and then to perform image classification, called two-stage methods in this paper [7, 17, 18]. However, the main reason why image degradation affects image classification is that the structural and statistical properties of pixels in the neighborhood are obstructed by image degradation. Since the existing image enhancement methods are devised to achieve pleasing visual effect, they cannot guarantee that the regions with similar structure in the image can be enhanced uniformly, leading to uncovered and incomplete feature representation for classification. A typical example is shown in Figure 1. As can be seen, there is still a margin between the feature distributions of enhanced image and clear image.

Another feasible solution for low-quality image classification is to transform it into a domain adaptation problem and to match the degraded and clear images with adversarial learning or kernelized training. This method is based on the assumption that the marginal distributions of low- and high- quality images can be aligned in the learned feature space by a deep network. Therefore, after decreasing the divergence between the marginal distributions in the learned feature space, the classifier trained with high-quality images tends to perform well on the low-quality images.

While promising, most existing domain adaptation approaches require either complex neural network architectures [3, 27, 14] or fine-tuning the target domain [31, 21, 26]. Different from these methods, this paper proposes to learn a transferrable mapping relationship between deep representations of low- and high- quality images, and leverage it as a deep degradation prior (DDP) for image classifier. This method is based on statistical observations that, i)

[*]co-first author. [†]Corresponding author.

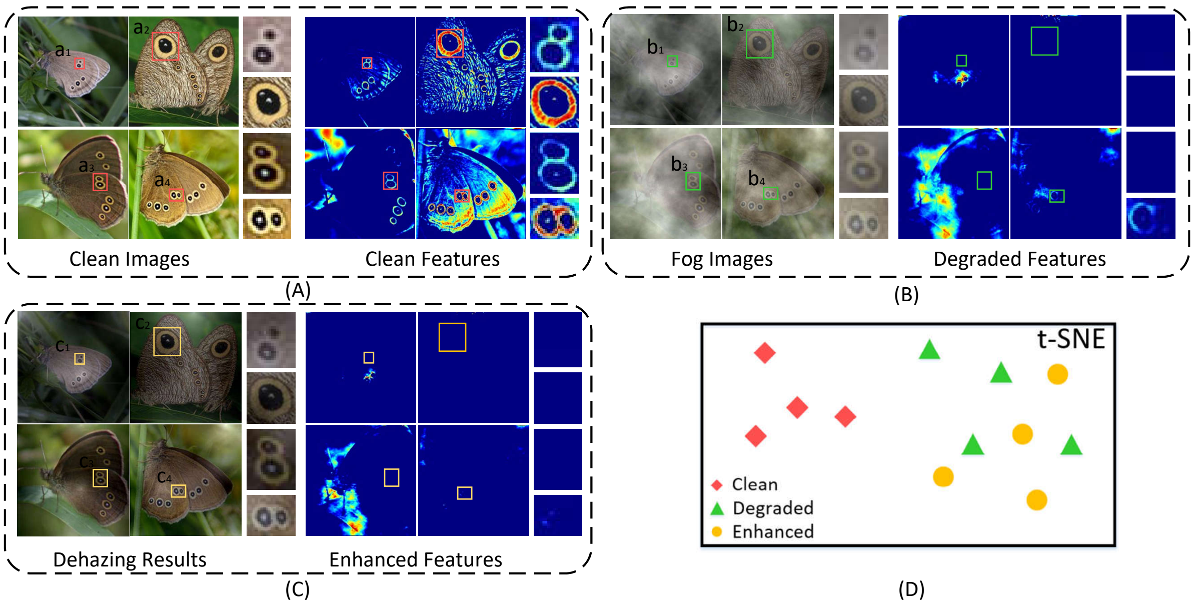

Figure 1. We choose four similar patches $a_1 \sim a_4$ from different clean images and visualize their features using VGG16 trained on ImageNet, as shown in (A). We visualize the feature distribution by using t-SNE [20]. As shown in (D), the clean patches have uniform distribution in feature space. The fog obstructs the statistical properties of pixels within local patches, resulting in feature drifting, as shown in (B) and (D). The visibility is well improved after dehazing [25]. However, the phenomenon of feature drift still exists, as in (C) and (D).

image patches with structural similarity have uniform distribution in the deep representation space, even if they come from different images; ii) the distributions of the corresponding structure-similar patches in low- and high-quality images have uniform margins under the same degradation condition. Since the statistical properties are irrelevant to image content, our proposed deep degradation prior can be learned on a small number of images and do not need supervision of semantic labels.

Specifically, we propose a feature de-drifting module (FDM) to learn the mapping relationship between deep representations of low- and high- quality images, and leverage it as a deep degradation prior (DDP) for low-quality image classification. The FDM is devised to compensate for the attenuation effect of image degradation on features, in which a four layers forward network is adopted to simulate the visual processing mechanism of non-classical receptive field model. After trained on an arbitrary degraded image dataset, the FDM can be easily plugged into the existing classification networks to improve their performance on degraded images. Evaluations on the benchmark dataset ImageNet-C [13] demonstrate the performance of our proposed method on low-quality image classification. The proposed DDP can be learned on only 10 images from CUB-C dataset, and achieve the accuracy improvement of VGG16 on ImageNet-C from 37% to 55%. The contributions of this paper are summarized as follows:

(1) This paper finds that, in the deep representation space, the distributions of low- and high-quality image patches have uniform margins under the same degradation condition, and accordingly proposes a novel deep degradation prior for low-quality image classification.

(2) This paper proposes a feature de-drifting module (FDM) to learn the mapping relationship between deep representations of low- and high- quality images. After trained on a small number of degraded image dataset, the FDM can be easily plugged into the existing classification networks to boost their performance on degraded images.

(3) Evaluations on the benchmark dataset ImageNet-C demonstrate the performance of our proposed method under various degradation conditions. Even under the extreme setting that only 10 images from CUB-C dataset are used for the training, our method still achieves the accuracy improvement of VGG16 from 37% to 55%.

2. Related Work

Differences in the representations of degraded images and clean images would shift the feature distribution, resulting in a performance drop when a pre-trained classification model is evaluated cross-domain [29]. Image enhancement algorithms can restore the degraded image to a clear version such that human vision can identify object and structure details. For example, B. Li *et al.* [17] uses the dehazing network to improve the object detection performance in the foggy environment. D. Da *et al.* [7] evaluates the influence of image super-resolution for high-

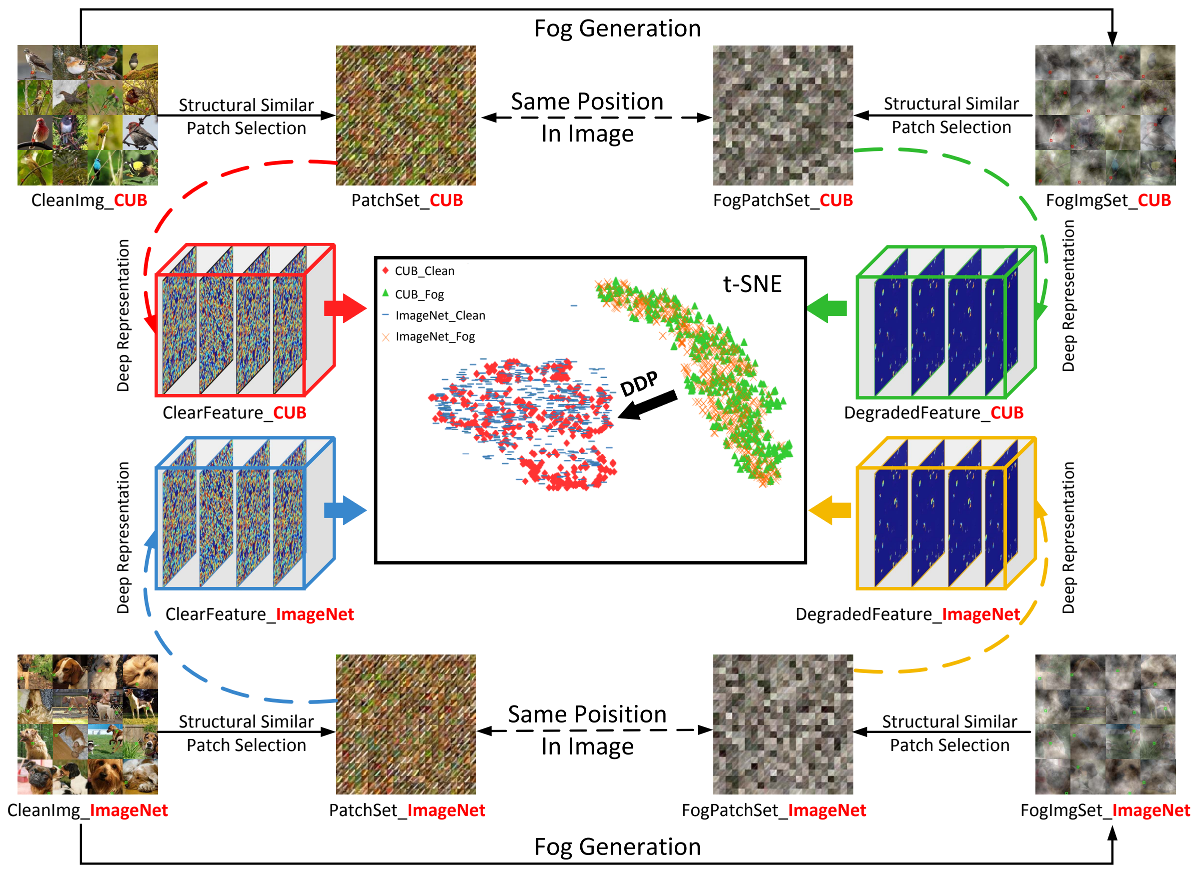

Figure 2. The distributions of corresponding structure-similar patches in the deep representation space. The high-quality images are from ImageNet dataset [8] and CUB dataset [33], and the low-quality images are generated by the method in [13].

level vision tasks. D. Liu *et al.* [18] employs the denoising algorithm to improve image segmentation performance. However, these enhancing methods can't always promise an improved performance for down-stream high-level vision tasks, especially when the degradation is heavy. Fine-tuning the network using degraded images is another popular strategy to improve performance. However, fine-tuning based methods need semantic labels to supervise network training. When the semantic labels are unavailable for degraded images, it will not be applicable. Domain adaption [22] transfers the models across different domains by solving the domain shift problem. Enormous of domain adaption methods [45, 31, 21, 26, 3, 27, 14] are proposed in recent years. However, these methods usually have complex neural network architectures. Besides, there is no work specifically tackling the classification problem on degraded images. By contrast, we propose a light-weight "plugging-in" feature de-drifting module, which significantly improves the generalizability of networks pre-trained on the ImageNet.

The most related work with us are [30, 29], where W. Tan *et al.* proposes a feature super-resolution method to achieve the semantic-level alignment between high-resolution and low-resolution images, and Z. Sun *et al.* proposes a feature quantization method to map the distorted features into new space with less divergence. Different from them, our method is more general since: 1) it can be trained on a dataset while testing on another, *i.e.*, CUB-C $\rightarrow$ ImageNet-C; 2) it can achieve fairly good performance even if it is only trained on a small scale dataset, *e.g.*, 10 images from CUB-C; 3) it does not require the semantic labels of target domain, reducing the effort for collecting and annotating degraded images.

3. Deep Degradation Prior

Prior work on sparse coding evidence that local patches in natural images can be reconstructed based on a learned over-complete visual dictionary, where the representation coefficient is sparse, called *sparse representation* [35, 19, 24, 32]. The visual dictionary is learned from many arbitrary local patches sampled from natural images as long as they contain diverse local structures, *e.g.*, smooth areas, edges, textures, *etc.* In other words, the visual dictionary learned from a dataset (*e.g.*, CUB) can be used to represent images from another dataset ($e.g.$, ImageNet). This property comes from the following fact that the local patches are statistically irrelevant to specific image content ($e.g.$, semantics) and can be embedded into a low-dimensional manifold [35]. Besides, recent progress in deep convolutional neural network also witnesses a similar phenomenon that features learned in shallow layers are mainly low-level ones, such as edges, colors, textures, etc [9, 38]. Accordingly, the learned convolutional weights can be regarded as visual dictionaries to extract (represent) these low-level features (local patches).

Motivated by the above work, we propose a novel deep degradation prior which is simple and effective for downstream high-level vision tasks, such as classification on degraded images. Specifically, we argue that: 1) the clear image patches from different dataset share a similar distribution in the feature embedding space, resulting in an indistinguishable cluster; 2) the degraded image patches have the similar property while they are separated from the clear ones due to their distinct local statistics; 3) if we can learn a mapping between the clear features and degraded features, it could be used for arbitrary natural images.

To illustrate the above claims, we conducted a statistical experiment on the clear and foggy images, synthesized according to the hazy model [13] on the ImageNet [8] dataset and CUB [33] dataset. Some examplar hazy images are shown in the right corners of Figure 2. First, for the ImageNet dataset, we sampled hundreds of clear patches and the corresponding foggy patches from the same positions, denoted S_I and S_{IF}, respectively. The same procedure was carried out on the CUB dataset, and the sampled patches are denoted S_C and S_{CF}, respectively. Referring to [35], the patch size was set 10$*$10 in the experiment. To better visualize their distributions and avoid messy clusters, we filtered the patches to enable that they share the similar local structures. Mathematically, they were subjected to the following constraint:

$$SSIM\,(p_i, p_j) > T,\ \forall\, p_i,\ p_j \in S_I\ or\ S_C, \qquad (1)$$

where SSIM denotes the structural similarity measurement [4]. T is a threshold and was set to 0.7. We kept 500 patches for both clear and foggy cases on each dataset, as shown in the up/bottom middle part of Figure 2.

Then, to obtain their feature representations in the embedding space [40], we used the VGG16 network pre-trained on the ImageNet dataset as the feature extractor. Specifically, the features of each patch were extracted from the "conv2_2" layer, since it's respective field is 9*9, almost the same with the patch size. These features are shown in the red, blue, green, and yellow cubes in Figure 2. Finally, we leveraged t-SNE [20] to visualize them, as shown in the central part of Figure 2.

It is clear that: 1) features from clear images are clustered together regardless which dataset they come from. It is the same for the foggy case; 2) there is a large gap between the clear features and foggy ones. Therefore, if we find a mapping between them, we can bridge them together. In this sense, restoring the degraded image patch to a clear one in the feature embedding space will certainly benefit the down-stream classification and detection tasks. In this paper, we call such a mapping as the *deep degradation prior (DDP)*. In the next part, we will present an efficient solution to show how to learn an effective DDP.

4. Learning DDP by Deep Neural Network

4.1. Overview of the Network

Given a clear dataset and its paired degraded dataset without semantic labels, our goal is to learn an effective DDP, which can be plugged in existing convolutional neural networks seamlessly to enhance their generalizability on degraded images. To this end, we propose an effective learning method by reconstructing the clear features from the degraded ones under the supervision of a simple Mean Square Error (MSE) loss. As shown in Figure 3 (a), during the training phase, we first use a pre-trained model to extract the low-level features of both degraded and clear images, for example, "conv2_2" in VGG16 [28], and the first layer in AlexNet [15], respectively. This part of network is fixed during the training phase and is called Shallow Pretrained Layers (SPL) in this paper. The degraded and clear features are denoted "DF" and "CF" in Figure 3 (a).

Then, we propose a novel feature de-drifting module (FDM) to accomplish the feature reconstruction. Taking the hazy image as an example, it is concatenated with the DF from SPL together and fed into FDM. Leveraging the residual learning idea, the output feature from FDM is fused with the original DF by an element-wise sum. The resulting enhanced feature (EF) is compared with its paired CF to calculate the MSE loss and the error is back-propagated to FDM to update its parameters.

During the testing phase, we insert the trained FDM into an existing classification network, $i.e.$, between its SPL and subsequent deep pre-trained layers (DPL), as shown in Figure 3 (b). It is noteworthy that the learned weights in FDM serve as the learned DDP to map the degraded features to the clear ones (see Figure 2).

4.2. Feature De-drifting Module

The degradation changes the statistics of a local patch, which results in a biased feature response by SPL compared with the original clear one. In order to correct the drifted feature response, we propose a novel feature de-drifting module. It is inspired by the non-classical receptive field of human vision [6] as shown in Figure 3 (c). Non-classical receptive field is very useful to enhance the high frequencies

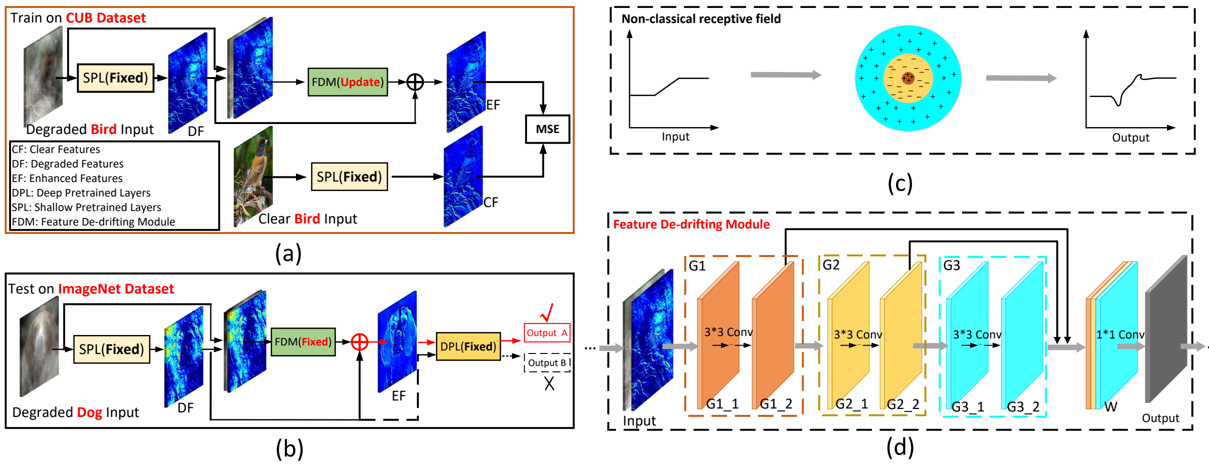

Figure 3. (a)-(b) Diagram of the training/testing phase of the proposed method. (c) Illustration of the non-classical receptive field [6] for enhancing the high frequencies while maintaining the low frequencies. (d) The network structure of our feature de-drifting module (FDM).

while maintaining low frequencies. Mathematically, given an input I, a filter with a non-classical receptive field generates an output f as follows:

$$f = A_1(I * G_1(\sigma_1)) + A_2(I * G_2(\sigma_2)) + A_3(I * G_3(\sigma_3)), \tag{2}$$

where $G_1 \sim G_3$ denote three Gaussian filters with different filter bandwidths, $i.e.$,

$$G(\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2+y^2}{2\sigma}) \tag{3}$$

$A_1 \sim A_3$ represent the coefficient in the central, surrounded, and marginal frequency areas, respectively. $*$ denotes the convolution operation. $\sigma_1 \sim \sigma_3$ are the scale parameters determining the filter bandwidths accordingly. Eq. (2) can be reformulated as:

$$\begin{array}{c} f = A_1(I * G_1) + A_2((I * G_1) * {G_2}') \\ +A_3((I * G_1) * {G_2}') * {G_3}' \end{array}, \tag{4}$$

where, ${G_2}'$ and ${G_3}'$ are also Gaussian filters with scale parameters $\sqrt{\sigma_2^2 - \sigma_1^2}$ and $\sqrt{\sigma_3^2 - \sigma_2^2}$. In this way, the convolutional result in the first (second) term can be used as the input of the second (third) Gaussian filter ${G_2}'$ (${G_3}'$).

Inspired by the non-classical receptive field and Eq. (4), we design our FDM as shown in Figure 3 (d). The first three blocks filled in orange, yellow and blue, represent the convolution process in the central, surrounded, and marginal frequency areas, respectively. Each block consists of two convolutional layers. The output features from the three blocks are then concatenated together and fed into a final 1*1 convolutional layer simulate the linear weighting between the central, surrounded and marginal frequency parts, as shown in Eq. (2). Details of FDM are summarized in Table 1 and Table 2 for VGG16 and AlexNet, respectively.

Table 1. The details of FDM with VGG16.

	InputSize	Num	Filter	Stride	Pad
G1_1	131*112*112	128	3	1	1
G1_2	128*112*112	128	3	1	1
G2_1	128*112*112	64	3	1	1
G2_2	64*112*112	64	3	1	1
G3_1	64*112*112	32	3	1	1
G3_2	32*112*112	32	3	1	1
W	224*112*112	128	1	1	0

Table 2. The details of FDM with AlexNet.

	InputSize	Num	Filter	Stride	Pad
G1_1	67*27*27	128	3	1	1
G1_2	128*27*27	128	3	1	1
G2_1	128*27*27	64	3	1	1
G2_2	64*27*27	64	3	1	1
G3_1	64*27*27	32	3	1	1
G3_2	32*27*27	32	3	1	1
W	224*27*27	64	1	1	0

5. Experiments

We evaluate the performance of our proposed deep degradation prior on ImageNet-C dataset [13], which is a rigorous benchmark with 50000 images in 1000 categories and widely used for robustness evaluation of image classifier. In this paper, we mainly focus on three kinds of degradation conditions, including fog, low contrast, and brightness. For each of the mentioned degradation conditions, we perform experiments on five levels of degradation. The examples of degraded images are shown in Figure 4. Besides, under each degradation condition, we test the influence of data size on deep degradation prior modeling.

Following the degraded image generation methods in [13], we use the clean images of CUB (total 11788 images)

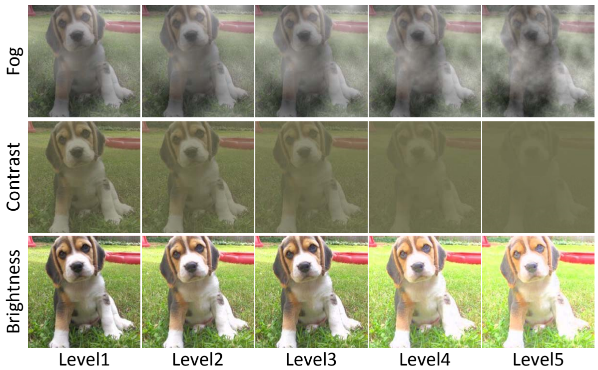

Figure 4. Some examples of degraded images from ImageNet-C.

dataset to synthesize the degraded images, named CUB-C, for the training of FD-Module. We employ AlexNet [15] and VGG16 [28] pre-trained on clean images as base models. The FD-Module is trained for 5000 iterations using SGD with batch size of 32. The initial learning rate is 0.001. The learning rate decreases to 0.0001 after 2500 iterations. In FD-Module, the filter weights of each layer are initialized using the MSRA initialization method.

5.1. Fog

Fog is very common for the images captured in outdoor and will cause image degradation due to atmospheric absorption and scattering. The high-frequency components and color fidelity are degraded in fog images. Since the degradation is spatial-variant, it will change the structure of local regions inconsistently and significantly increase the difficulty of feature extraction. The goal of image dehazing methods is to enhance the contrast and restore the structure details, which make the images more visual pleasing, but cannot eliminate the structure inconsistency. In this paper, we select five state-of-the-art image dehazing methods as baseline: DehazeNet [4], AoD-Net [17], FPCNet [42], FAMED-Net [43], and GFN-Net [25]. All of the test foggy images are dehazed by the baseline methods and then send to pre-trained image classifiers (VGG16 and AlexNet). Our proposed FDM is firstly trained on foggy images and then plugged into the same pre-trained image classifiers (VGG16 and AlexNet). The evaluations are performed on five degradation levels respectively.

As can be seen from Figure 5 (a), our proposed method significantly surpasses the two-stage methods in terms of classification accuracy, especially when the fog concentration is large. Moreover, with the increase of fog concentration, the performance of two-stage methods drops notably, while our method can also improve the accuracy of the pre-trained network by more than 30%. It demonstrates that FDM effectively boosts the fog-degraded features.

5.2. Low Contrast

The images captured under weak and colored illumination conditions often appear low contrast and color cast, which will degrade the recognition accuracy. The goal of contrast enhancement methods is to improve the dynamic range of images and remove color cast. In this paper, we select five contrast enhancement methods as baseline: HE [11], CVC [5], WAHE [1], LDR [16], OpenCE [36]. All of the low-contrast images are enhanced by the baseline methods and then send to pre-trained image classifiers (VGG16 and AlexNet). Our proposed FDM is firstly trained on low-contrast images and then plugged into the same pre-trained image classifiers (VGG16 and AlexNet). The evaluations are performed at five degradation levels.

As shown in Figure 5 (b), the two-stage method cannot consistently improve the classification accuracy. When the degradation is small (e.g., Contrast-1 and Contrast-2), the performance is even worse than directly using the origin images. As a contrast, our proposed method can consistently improve the classification accuracy of all five degradation conditions, which demonstrates the robustness of our DDP.

5.3. Brightness

Low lighting condition and limited dynamic range of digital imaging devices in image capturing can significantly degrade the image quality from many aspects, such as low contrast and amplified noise, which results in a biased feature distribution compared with the clear one and downgrades the classification accuracy. One common solution is to use HDR algorithm to balance the dynamic range and suppress noises. In this experiment, we compared our method with four HDR methods: LIME [12], HQEC [44], SRIE [10], MRINP [34]. The results are shown in Figure 5 (c). It is clear that the accuracy of classifying HDR enhanced images is even slightly lower than classifying the degraded images. Although the HDR methods could improve the visibility of degraded images, it also produced some artifacts, which consequently results in extra feature drifting as well. By contrast, our method improves the accuracy by correcting the feature drifting and aligning them with the clear ones in the embedding space. It demonstrates the effectiveness of the learned DDP by the FDM.

5.4. Performance Analysis

We first explore the effect of training set volume on deep degradation prior modeling. We use different proportions of degraded images from CUB-C to train the FDM, *e.g.*, Fog-3, Contrast-3, and Brightness-3. The results are plotted in Figure 6 (a)-(c), respectively. As can be seen, the performance of FDM is not sensitive to the training dataset volume. The classification accuracy only drops by 5% when using only 1% training data compared with using all the training data (total 11788 images). Besides, in the extreme

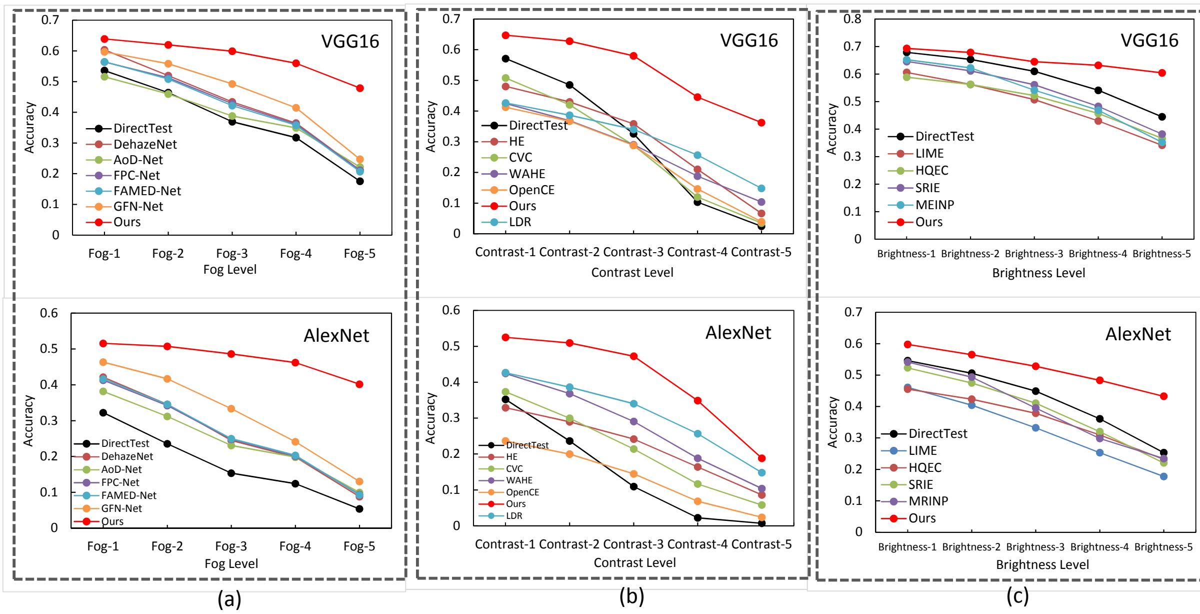

Figure 5. Comparison of DDP against the two-stage methods under different degradation conditions, (a) fog, (b) contrast and (c) brightness.

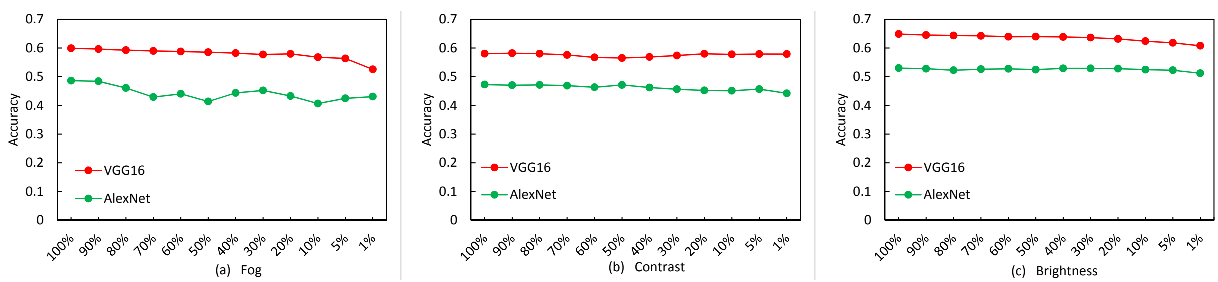

Figure 6. The influence of training dataset volume on DDP. The x-axis represents the proportion of images in CUB dataset used for training.

condition that only 10 foggy images from CUB-C dataset are used for training, our method still boosts the accuracy of VGG16 on ImageNet-C from 37% to 55%. It is very useful when large scale degraded images are difficult to collect.

Then, we present the feature visualization results to demonstrate the performance of feature de-drifting. Here we take the fog degradation as example. We visualize the feature maps of synthesized foggy images, clear images, dehazed images and the ones generated by our method in Figure 8. As can be seen, due to the influence of fog, the amplitude of feature response of object is attenuated inconsistently. The dehazing methods can recover some structural information of the image, but cannot consistently enhance the discriminative features for object. As a contrast, our method significantly boosts the feature responses on discriminative region while suppressing the interference. Furthermore, we also evaluate the performance of our method and two-stage method on real foggy images. The visualized feature maps are shown in Figure 7. We can see that, when the fog concentration is small, the two-stage method can also recover the structural features. However, with the increase of fog concentration, the discrimination of the features captured from the dehazed images drop significantly. As a contrast, the features extracted by our method can always reveal the structure of objects, even in the heavies fog condition. It demonstrates that our proposed method is robust and applicable to real foggy scenes.

Apart from above three kinds of degradation, we also evaluate the performance of DDP on the degraded images with motion blur. DPP can improve the accuracy of the VGG16 from 24% to 38% by using only 10 images for training. In addition, we conduct experiments on real-world images with two types of degradations (under-exposure and noise) from UG2 [37]. We train two DDP modules (*e.g.*, DDP$_c$ for contrast enhancement and DDP$_d$ for denoising), and incorporate them into the pre-trained networks (*e.g.*, VGG16) to boost the degraded features as shown in Figure 9. We can see that the results are also positive.

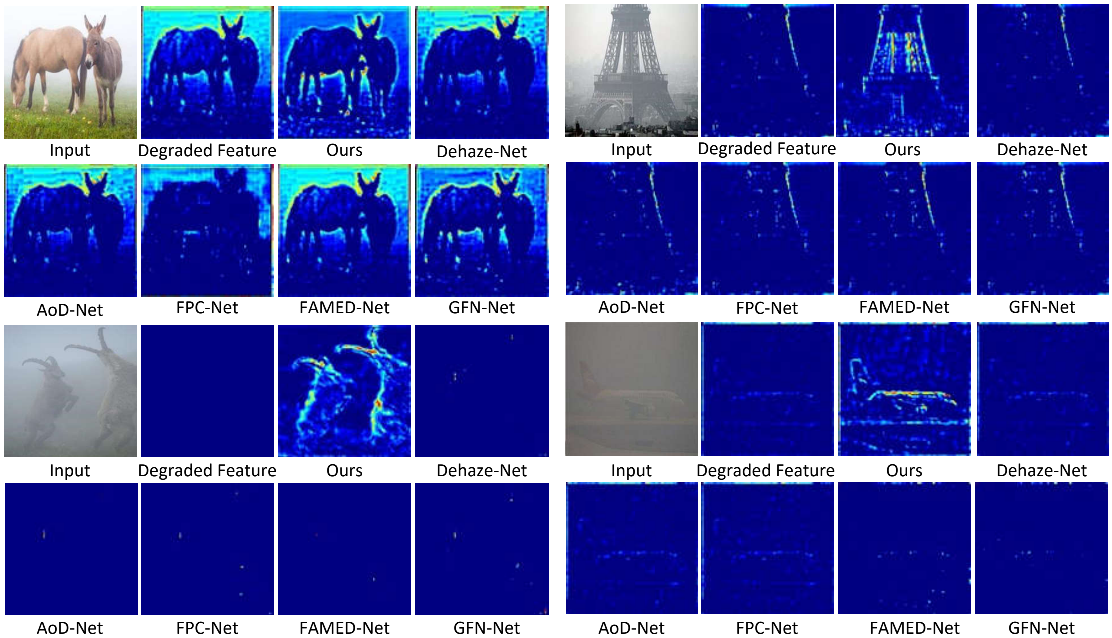

Figure 7. Visualization of the feature maps of real foggy images and de-foggy results. The foggy images are from Haze20 dataset [23]. The feature maps are extracted from the Conv2_2 layer of VGG16 pre-trained on ImageNet. With the increase of fog density, the discrimination of the features of dehazed images drop significantly, while the features extracted by our method can always reveal the structure of objects.

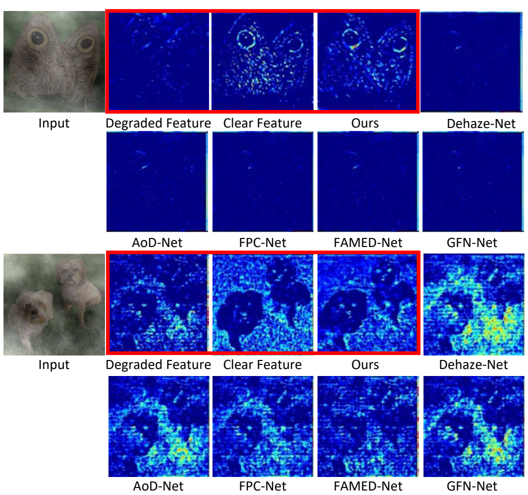

Figure 8. The feature maps of synthesized foggy images, clear images, dehazed images and the ones generated by our method. All the features are output from the same Conv2_2 layer of VGG16 trained on ImageNet.

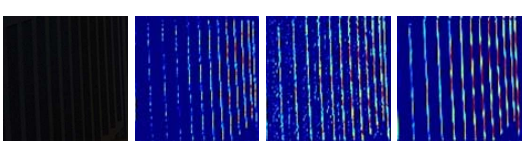

Figure 9. The feature maps of real low-light image with two kinds of degradations (*e.g.*, low-contrast and noise).

6. Conclusion

In this paper, we introduce a novel deep degradation prior (DDP) for low-quality image classification, which can be used to reduce the feature mismatch between clear images and various degraded ones. We devise a simple and effective module named feature de-drifting module (FDM) to learn the DDP by adaptively enhancing features at different frequencies. FDM can be trained on a small scale dataset, *e.g.*, only 10 images from CUB-C, but still remains robust and effective. Further, the learned DDP by FDM on a dataset can be generalized to another dataset with the same degradation, regardless the diverse image contents. Moreover, it is very easy to deploy FDM into existing classification networks as a "plugging-in" module. Extensive experiments on different types of degradation and two popular benchmark datasets, *i.e.*, ImageNet-C and CUB-C, validate the effectiveness of the learned DDP by FDM in degraded image classification tasks. It boosts the classification accuracy by large margins under various degradation conditions.

Acknowledgments: This work was supported by the National Key RD Program of China under Grant 2017YFB1300201, the National Natural Science Foundation of China (NSFC) under Grants 61872327, 61622211, 61806062, U19B2038 and 61620106009 as well as the Fundamental Research Funds for the Central Universities under Grant WK2380000001 and WK2100100030.

References

[1] T Arici, S Dikbas, and Y Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 18(9):1921–1935, 2009. 6

[2] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–894, 2017. 1

[3] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3722–3731, 2017. 1, 3

[4] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 4, 6

[5] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 6

[6] Li Chao-Yi, Pei Xing, Zhow Yi-Xiong, et al. Role of the extensive area outside the x-cell receptive field in brightness information transmission. *Vision Research*, 31(9):1529–1540, 1991. 4, 5

[7] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. Is image super-resolution helpful for other vision tasks? In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9. IEEE, 2016. 1, 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1, 3, 4

[9] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2016. 4

[10] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016. 6

[11] Rafael C Gonzales and Richard E Woods. Digital image processing, 2002. 6

[12] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 1, 6

[13] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 1, 2, 3, 4, 5

[14] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. *arXiv preprint arXiv:1711.03213*, 2017. 1, 3

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 4, 6

[16] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 6

[17] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 1, 2, 6

[18] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S Huang. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*, 2017. 1, 3

[19] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):171–184, 2012. 3

[20] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008. 2, 4

[21] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5715–5725, 2017. 1, 3

[22] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 3

[23] Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does haze removal help cnn-based image classification? In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 682–697, 2018. 8

[24] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2233–2246, 2012. 3

[25] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018. 2, 6

[26] Paolo Russo, Fabio M Carlucci, Tatiana Tommasi, and Barbara Caputo. From source to target and back: symmetric bi-directional adaptive gan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8099–8108, 2018. 1, 3

[27] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, 2018. 1, 3

[28] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4, 6

[29] Zhun Sun, Mete Ozay, Yan Zhang, Xing Liu, and Takayuki Okatani. Feature quantization for defending against distortion of images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7957–7966, 2018. 2, 3

[30] Weimin Tan, Bo Yan, and Bahetiyaer Bare. Feature super-resolution: Make machine see more clearly. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3994–4002, 2018. 3

[31] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4068–4076, 2015. 1, 3

[32] Andrew Wagner, John Wright, Arvind Ganesh, Zihan Zhou, Hossein Mobahi, and Yi Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):372–386, 2011. 3

[33] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 3, 4

[34] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, Zhiwei Xiong, Wei Zhang, and Feng Wu. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2015–2023. ACM, 2019. 1, 6

[35] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 3, 4

[36] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017. 6

[37] Ye Yuan, Wenhan Yang, Wenqi Ren, Jiaying Liu, Walter J. Scheirer, and Zhangyang Wang. 7

[38] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014. 4

[39] Zheng-Jun Zha, Xian-Sheng Hua, Tao Mei, Jingdong Wang, Guo-Jun Qi, and Zengfu Wang. Joint multi-label multi-instance learning for image classification. In *2008 ieee conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2008. 1

[40] Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuicheng Yan, and Tat-Seng Chua. Robust (semi) nonnegative graph embedding. *IEEE transactions on image processing*, 23(7):2996–3012, 2014. 4

[41] Jing Zhang, Yang Cao, Shuai Fang, Yu Kang, and Chang Wen Chen. Fast haze removal for nighttime image using maximum reflectance prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7418–7426, 2017. 1

[42] Jing Zhang, Yang Cao, Yang Wang, Chenglin Wen, and Chang Wen Chen. Fully point-wise convolutional neural network for modeling statistical regularities in natural images. *arXiv preprint arXiv:1801.06302*, 2018. 1, 6

[43] Jing Zhang and Dacheng Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *arXiv preprint arXiv:1906.04334*, 2019. 6

[44] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Wei-Shi Zheng. High-quality exposure correction of underexposed photos. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 582–590. ACM, 2018. 6

[45] Qiming Zhang, Jing Zhang, Wei Liu, and Dacheng Tao. Category anchor-guided unsupervised domain adaptation for semantic segmentation. In *Advances in Neural Information Processing Systems*, pages 433–443, 2019. 3

[46] Wei Zhang, Shengnan Hu, Kan Liu, and Zhengjun Zha. Learning compact appearance representation for video-based person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8):2442–2452, 2018. 1