

Article

Robust Visual Recognition in Poor Visibility Conditions: A Prior Knowledge-Guided Adversarial Learning Approach

Jiangang Yang ^{1,2}, Jianfei Yang ³, Luqing Luo ¹, Yun Wang ⁴, Shizheng Wang ⁵ and Jian Liu ^{1,*}

¹ Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100029, China; yangjiangang18@mails.ucas.ac.cn (J.Y.)

² University of Chinese Academy of Sciences, Beijing 101408, China

³ School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639798, Singapore

⁴ Guangdong Greater Bay Area Institute of Integrated Circuit and System, Guangzhou 510535, China

⁵ R&D Center for Internet of Things, Chinese Academy of Sciences, Wuxi 214200, China

* Correspondence: liujian@ime.ac.cn

Abstract: Deep learning has achieved remarkable success in numerous computer vision tasks. However, recent research reveals that deep neural networks are vulnerable to natural perturbations from poor visibility conditions, limiting their practical applications. While several studies have focused on enhancing model robustness in poor visibility conditions through techniques such as image restoration, data augmentation, and unsupervised domain adaptation, these efforts are predominantly confined to specific scenarios and fail to address multiple poor visibility scenarios encountered in real-world settings. Furthermore, the valuable prior knowledge inherent in poor visibility images is seldom utilized to aid in resolving high-level computer vision tasks. In light of these challenges, we propose a novel deep learning paradigm designed to bolster the robustness of object recognition across diverse poor visibility scenes. By observing the prior information in diverse poor visibility scenes, we integrate a feature matching module based on this prior knowledge into our proposed learning paradigm, aiming to facilitate deep models in learning more robust generic features at shallow levels. Moreover, to further enhance the robustness of deep features, we employ an adversarial learning strategy based on mutual information. This strategy combines the feature matching module to extract task-specific representations from low visibility scenes in a more robust manner, thereby enhancing the robustness of object recognition. We evaluate our approach on self-constructed datasets containing diverse poor visibility scenes, including visual blur, fog, rain, snow, and low illuminance. Extensive experiments demonstrate that our proposed method yields significant improvements over existing solutions across various poor visibility conditions.

Keywords: robust visual recognition; poor visibility conditions; unsupervised domain adaptation; image restoration



Citation: Yang, J.; Yang, J.; Luo, L.; Wang, Y.; Wang, S.; Liu, J. Robust Visual Recognition in Poor Visibility Conditions: A Prior Knowledge-Guided Adversarial Learning Approach. *Electronics* **2023**, *12*, 3711. <https://doi.org/10.3390/electronics12173711>

Academic Editor: Donghyeon Cho

Received: 12 August 2023

Revised: 29 August 2023

Accepted: 30 August 2023

Published: 2 September 2023

Corrected: 23 January 2024



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent advances in deep learning have led to remarkable success in computer vision tasks, including object recognition [1], object detection [2], and semantic segmentation [3]. Despite the achievements, deep learning models still face significant challenges when being applied to real-world scenarios. One of the most critical issues is the presence of poor visibility conditions, which introduce natural perturbations that can degrade image quality in various ways. These include the loss of texture information, object shape distortion, and partial occlusion. Studies have shown that deep models are susceptible to significant performance degradation under poor visibility conditions [4,5].

To address this problem, previous studies utilize image restoration techniques [6–9] to recover the damaged visual content, which is then fed into deep discriminative models for high-level tasks. These restoration methods are often limited in their generalization

due to their specificity to certain types of scenes, such as image de-raining for rainy scenes. Another popular solution is unsupervised domain adaptation. In this setting, labeled clear images are employed as the source domain, whereas unlabeled images from scenes with poor visibility are utilized as the target domain. By learning domain-invariant features, unsupervised domain adaptation can improve the model performance in the target domain. While some studies have investigated unsupervised domain adaptation in a limited number of poor visibility scenarios, such as day-to-night variation [10], the majority of poor visibility scenes remain largely unexplored. Additionally, other works have designed various image augmentation strategies [11–13] to improve the model robustness against multiple visual perturbations.

The above-mentioned studies have demonstrated the efficacy of image restoration and unsupervised domain adaptation in enhancing the performance of object recognition tasks in single poor visibility scenarios. However, there is a lack of research exploring the integration of key ingredients from these two directions. For instance, the prior knowledge widely employed in image restoration tasks, such as the dark channel prior used for deblurring and dehazing [14,15], is seldom utilized in high-level semantic tasks. Most unsupervised domain adaptation research focus on the consistency of deep semantic representations across domains while ignoring the essential prior information embedded in shallow features. Therefore, we aim to leverage traditional prior knowledge effectively within the paradigm of unsupervised domain adaptation, thereby elevating the robustness of object recognition models across diverse types of poor visibility scenarios.

In this paper, we propose a new deep learning-based approach called Prior Knowledge-guided Adversarial Learning (PKAL) to enhance visual recognition in multiple poor visibility conditions. Based on our observations, the prior knowledge of intermediate features varies between clean and blurry images, and the dark and bright channel priors [15] from the intermediate features lose many meaningful contents due to visual blurs, as shown in Figure 1. Henceforth, it is difficult to extract sufficient semantic representations, which leads to a performance drop in high-level tasks. This phenomenon is widely witnessed in other poor visibility conditions. Moreover, prior studies have attempted to capture prior knowledge of low-level features and align these pivotal visual cues, ultimately enhancing the performance of downstream tasks [16–18]. To this end, we design a Feature Priors Matching Module (FPMM) to discern the discrepancies of prior knowledge-based features between clean and low-quality images and to suppress them during training. Deep models preserve more meaningful information from shallow layers of low-quality data under the constraints imposed by FPMM, thus enhancing the model robustness in poor visibility conditions.

Considering that the FPMM works in the shallow layers for robust and generic features, we propose a novel Mutual Information-based Robust Training (MIRT) strategy to improve the robustness of task-specific features. Concretely, MIRT establishes an adversarial learning mechanism between two feature generators and one class discriminator to enhance robust deep representations and to refine decision boundaries simultaneously. At the beginning of training, one feature generator is equipped with FPMM while the other is not. Thus, the former generator extracts more robust features than the latter one. The class discriminator accepts the robust features and rejects the other ones. Mutual information [19] is employed to quantify the receptive strength of deep features. Under the adversarial mechanism, MIRT encourages the generator with FPMM to generate robust features continuously, whereas the discriminator refines decision boundaries to reject sensitive features. Ultimately, the feature generator with FPMM and the class discriminator comprise a robust model for visual recognition under poor visibility conditions.

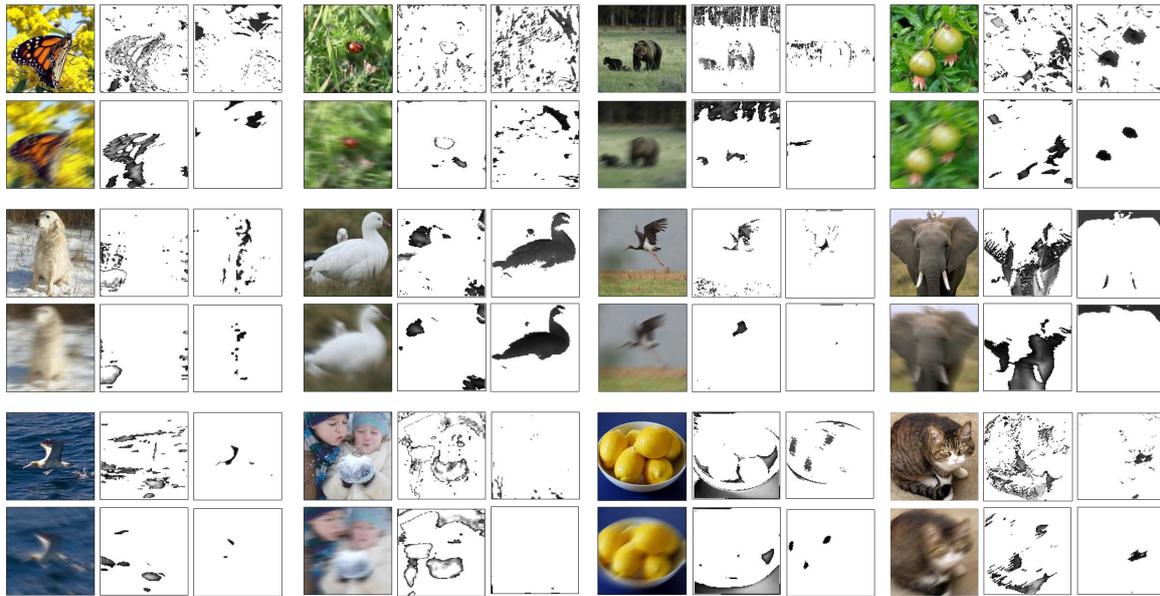


Figure 1. From left to right for each image, we show the raw data, dark channel features, and bright channel features. We can observe that the reduction of bright and dark channel features induced by visual blur may remove the significant object details and thus lose semantic information.

To validate the efficacy of our proposed approach, we build a comprehensive dataset of poor visibility with several common perturbations, including visual blur, fog, rain, snow, and low illuminance. The dataset comprises both real-world samples and synthetically generated data that simulate realistic settings. Furthermore, we conduct a comparative analysis between our approach and 16 established methods renowned for their effectiveness in image restoration, domain adaptation, and model robustness. Through extensive experimentation, we demonstrate that our approach outperforms the majority of existing methods in various poor visibility scenarios, while achieving comparable performance to the remaining ones. In summary, the main contributions of this paper are as follows:

- We propose a novel deep learning-based approach, PKAL, to enhance the model robustness for visual recognition under various poor visibility conditions.
- The proposed feature matching module, FPMM, transfers typical prior knowledge widely used in low-level tasks to high-level ones.
- We design an adversarial min–max optimization strategy to enhance robust task-specific representations and to refine decision boundaries simultaneously.
- We evaluate our proposed approach on a diversity of poor visibility scenarios, including visual blur, fog, rain, snow, and low illuminance. The experiments demonstrate the efficacy and adaptability of our approach.

2. Related Work

2.1. Image Restoration for Poor Visibility Conditions

Image restoration is a long-standing research interest aimed at enhancing visual data in poor visibility conditions. This includes image deblurring [20–22], de-raining [23,24], defogging [8,25], and other related areas. While these approaches focus on improving visual effects for human perception, they do not necessarily consider the needs of high-level machine vision. Early approaches mainly rely on seeking image prior knowledge [6,14,15,26–28], which represents special salient features in poor visibility scenes. These image priors are commonly incorporated as key regularization terms into the optimization process for image restoration. Despite achieving better visual results, these algorithms require high computation costs during testing and do not generalize well to similar poor visibility scenarios, thus limiting their application.

Recent advances in deep learning-based image restoration have aimed to address the aforementioned limitations by employing various loss functions, learning paradigms, and network architectures [20–22,29]. With the explosion of visual data, deep learning-based restoration algorithms have demonstrated superior performance and efficiency compared to prior-based algorithms. In addition to improving the visual quality of images, several studies investigate the potential benefits of image restoration for high-level vision tasks. For instance, restoration algorithms are evaluated as pre-processing modules in object recognition, detection, and segmentation [5,30].

2.2. Data Augmentation for Poor Visibility Conditions

Image augmentation has become a popular technique for training deep neural networks [31]. Early data augmentation methods were designed to improve model performance and alleviate overfitting. In more recent years, image augmentation has also been used to enhance model robustness by modifying visual attributes in various ways while preserving semantic content. One way of doing this is by altering patch-level pixel regions. For instance, Cutout [32] randomly masks out square regions of the inputs to simulate occlusions that occur in real-world scenarios. CutMix [33], on the other hand, replaces the masked regions with patches of other images to mitigate information loss. Another approach is to augment the training examples with varying image styles. Ref. [34] shows that applying AdaIN [35] to add stylized training data can promote robust deep models. DeepAugment [12] randomly adjusts the weights in the image-to-image network to produce the same content under distinct textural variations. In addition, various augmentation techniques, such as AutoAugment [13] and RandAugment [36], have been developed to identify the optimal combination of image processing operations to train deep models. AugMix [37], a new combination, randomly selects data processing operations and their severity levels in a fixed and parallel pipeline. It has shown great potential in enhancing model robustness against synthetic image corruptions [11].

2.3. Unsupervised Domain Adaptation

Machine vision under poor visibility conditions inevitably captures a large number of unlabeled, low-quality images. These low-quality images are often accompanied by visual perturbations, leading to a distribution shift from clean data. Unsupervised Domain Adaptation (UDA) is well suited to handle such distribution shift problems, aiming to improve the generalization of deep models trained in the label-abundant domain to the unlabeled domain with different data distributions. Based on their learning paradigm, UDA methods can primarily be categorized into two groups. The first one focuses on aligning feature statistics to learn invariant representations across domains. DDC [38] minimizes high-dimensional distribution discrepancies by measuring Maximum Mean Discrepancy (MMD) on deep representations. DAN [39], an advanced version of DDC, estimates feature discrepancies on multiple layers using MMD with a non-linear kernel. Deep CORAL [40] aligns the second-order statistics of deep features between the source and target domains. The second group pursues domain-invariant features through the adversarial learning principle. DANN [41], the pioneer in this area, uses an adversarial learning framework to force the discriminator to fail in recognizing the domain label of deep features. MCD [42] introduces a novel adversarial training strategy between two label classifiers to refine the task-specific decision boundary. CDAN [43] incorporates a conditioning strategy into adversarial learning to achieve better discriminability and transferability. VADA and DIRT-T [44] implement a two-stage adversarial learning pipeline to penalize violations of the cluster assumption [45]. Our proposed approach adheres to the UDA settings, seeking to maximize the use of unlabeled data.

3. Preliminaries

3.1. Setup

We consider a source domain $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ with N_s annotated clean examples and a target domain $\mathcal{D}_t = \{x_j^t\}_{j=1}^{N_t}$ containing N_t unlabeled low-quality data. Both domains are associated with K categories and sampled from joint distributions $\mathbb{P}(x^s, y^s)$ and $\mathbb{Q}(x^t, y^t)$, respectively. The ground-truth label y_i^s belongs to the set $1, 2, \dots, K$. We define a deep classifier F for the K -class classification problem. Given an input x , the deep classifier produces a K -dimensional vector such that $F(x) \in \mathbb{R}^K$. Using the softmax operation, we can compute the prediction probability for the k -th category as follows:

$$p(y = k|x, F) = \frac{e^{F_k(x)}}{\sum_{k=1}^K e^{F_k(x)}} \quad (1)$$

In this paper, we decompose the deep classifier F into a feature generator G and a label discriminator D , such that $F = G \circ D$. Therefore, we can rewrite (1) as follows:

$$p(y = k|x, G \circ D) = \frac{e^{D_k(G(x))}}{\sum_{k=1}^K e^{D_k(G(x))}} \quad (2)$$

3.2. Dark and Bright Channel Priors

In this paper, we revisit two types of classic image priors: the Dark Channel Prior (DCP) and the Bright Channel Prior (BCP), both of which are employed in our proposed approach. The DCP and BCP are derived from the observation that image perturbations can significantly affect the number of dark pixels (the smallest value in an image patch) and bright pixels (the largest value in an image patch). As illustrated in Figure 2, the DCP and BCP are clearly visible in the image histogram, where the number of dark and bright pixels decreases as natural perturbations occur. The DCP and BCP have been extensively utilized in image deblurring and dehazing [15,46]. For a given image I , we extract the dark channel information T_d for a pixel location q as follows:

$$T_d(I)(q) = \min_{p \in \Omega(q)} (\min_{c \in \{r, g, b\}} I^c(p)), \quad (3)$$

where p, q are the pixel locations and c denotes the color channel. The local patch centered at q is represented by $\Omega(q)$. The bright channel prior, denoted by T_b , can be defined in a similar manner:

$$T_b(I)(q) = \max_{p \in \Omega(q)} (\max_{c \in \{r, g, b\}} I^c(p)), \quad (4)$$

Furthermore, prior research incorporates the bright channel prior (BCP) and the dark channel prior (DCP) into the Extreme Channels Prior (ECP), which serves as the regularization term in both optimization-based deblurring [26,28] and deep deblurring [6,14].

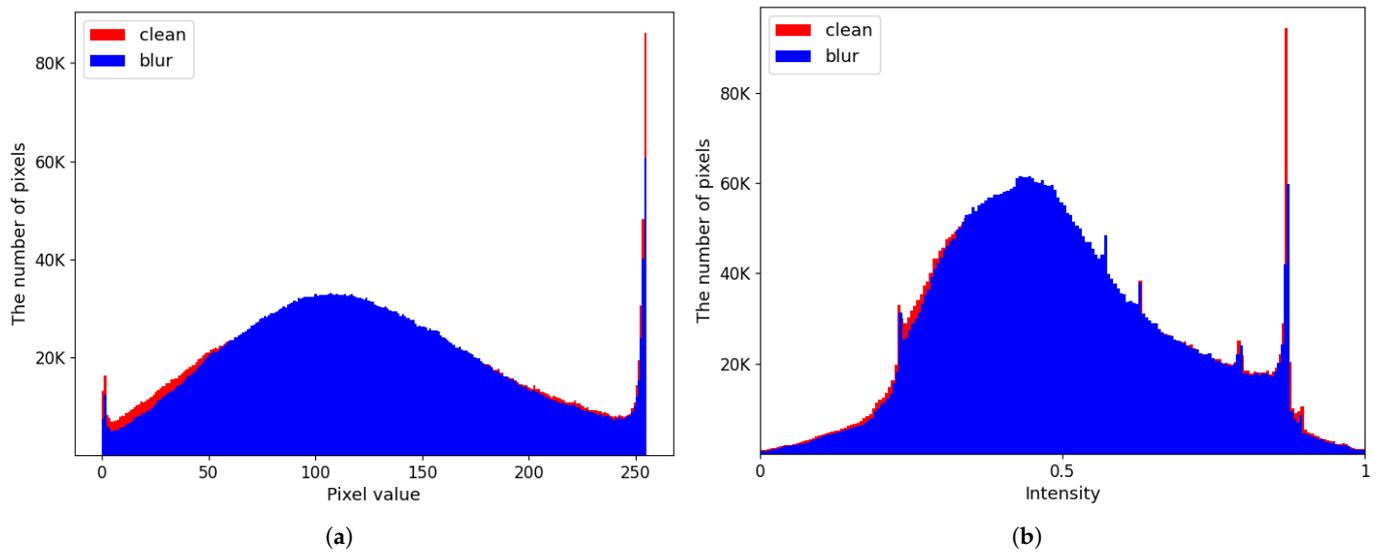


Figure 2. (a) The histogram comparison between clean and blurry images. (b) The histogram comparison between intermediate features from clean and blurry images.

3.3. The Estimation of Mutual Information

Mutual information is an entropy-based metric used to assess the mutual dependence between variables. Recent research demonstrates the effectiveness of mutual information as a regularization term in representation learning [19,47]. Let (X, Y) represent a pair of random variables. Its general form can be expressed as follows:

$$I(Y; X) = H(Y) - H(Y|X), \tag{5}$$

where $H(Y)$ is the marginal entropy of the variable Y and $H(Y|X)$ is the conditional entropy of Y given the variable X . Due to the infeasibility of obtaining an analytical solution for mutual information, recent studies employ deep learning techniques to estimate its lower or upper bounds [48,49]. In our approach, we present a straightforward method to estimate the mutual information between the inputs and their predicted outputs, which has demonstrated its efficacy in semi-supervised and unsupervised learning problems [47,50]. To provide a better understanding of this method, let us consider a K -class ($K \geq 2$) object recognition task, where x represents the input and y denotes its corresponding ground-truth label. We assume that F denotes the deep classifier. By using (1), we first calculate the conditional entropy of the prediction outputs given the inputs:

$$\begin{aligned} \mathbb{E}_{x \sim \mathbb{P}_X} [H[p(y|x, F)]] = \\ \frac{1}{N} \sum_{i=1}^N - \sum_{k=1}^K p(y = k|x_i, F) \log p(y = k|x_i, F), \end{aligned} \tag{6}$$

where H is the symbol of entropy and N denotes the sample number. The marginal entropy of the prediction outputs can be estimated as:

$$\begin{aligned} H[\mathbb{E}_{x \sim \mathbb{P}_X} [p(y|x, F)]] = \\ - \sum_{k=1}^K \left[\frac{1}{N} \sum_{i=1}^N p(y = k|x_i, F) \cdot \log \frac{1}{N} \sum_{i=1}^N p(y = k|x_i, F) \right], \end{aligned} \tag{7}$$

The conditional entropy gives an indication of the average uncertainty degree of the deep classifier on each input sample. In contrast, the marginal entropy reflects the total uncertainty of the deep classifier on the entire input distribution. By utilizing (5)–(7),

we can estimate the mutual information between the inputs and their corresponding predicted outputs.

4. Prior Knowledge-Guided Adversarial Learning

This section provides a detailed description of our proposed approach, Prior Knowledge-guided Adversarial Learning (PKAL), along with its crucial components: the Feature Priors Matching Module (FPMM) and Mutual Information-based Robust Training (MIRT).

4.1. Feature Priors Matching Module

Image priors consist of latent image attributes that are imperceptible to human vision but capture significant differences between visual domains (e.g., clean vs. foggy images). In typical image processing, image priors are commonly integrated as regularization terms in optimization. However, as illustrated in Figure 1, our analysis shows that mismatches in image priors between clean and low-quality data persist not only in the raw pixel domain but also in the intermediate features of deep models. If ignored, these mismatches can lead to deep representations with incorrect semantic information, ultimately resulting in erroneous predictions. Thus, mitigating this feature-level mismatch is essential for enhancing the robustness of deep models in high-level discriminative tasks.

Based on our observations, we propose the Feature Priors Matching Module (FPMM) as a plug-and-play solution to address feature-level mismatches in the dark and bright channel priors on shallow features. FPMM constrains the statistical discrepancies between these priors during the training process. As shown in Figure 3, FPMM initially extracts the dark channel prior T_d and the bright channel prior T_b from the shallow layer ϕ_0 of the feature generator G . By using (3) and (4), we formulate T_d and T_b on the intermediate features as follows:

$$T_d(\phi_0(x))(q) = \min_{p \in \Omega(q)} (\min_{c \in C} \phi_0(x)^c(p)), \quad (8)$$

$$T_b(\phi_0(x))(q) = \max_{p \in \Omega(q)} (\max_{c \in C} \phi_0(x)^c(p)), \quad (9)$$

where $\phi_0(x)$ represents the intermediate feature map with C channels for the given input x . $\phi_0(x)^c$ denotes the c -th channel feature map, where C is the total number of channels. After extracting the feature priors, we employ the Maximum Mean Discrepancy (MMD) as the high-dimensional distribution distance metric to quantify the statistical discrepancy of these feature-level priors between the clean and low-quality domains. The empirical approximation of this distance metric can be expressed as follows:

$$\begin{aligned} \text{MMD}(T; \mathcal{D}_s, \mathcal{D}_t) = \\ \|\mathbb{E}_{x_i^s \sim \mathcal{D}_s} [T(\phi_0(x_i^s))] - \mathbb{E}_{x_j^t \sim \mathcal{D}_t} [T(\phi_0(x_j^t))]\|_{\mathcal{H}_k}^2, \end{aligned} \quad (10)$$

where \mathcal{H}_k refers to the Reproducing Kernel Hilbert Space (RKHS) with a characteristic kernel k . Here, $T(\cdot)$ can be replaced with either $T_d(\cdot)$ or $T_b(\cdot)$ to indicate the extraction of either dark or bright prior-based features. Therefore, the final objective for FPMM can be formally defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{FPMM}}(G; \mathcal{D}_s, \mathcal{D}_t) = \\ \text{MMD}(T_b; \mathcal{D}_s, \mathcal{D}_t) + \text{MMD}(T_d; \mathcal{D}_s, \mathcal{D}_t), \end{aligned} \quad (11)$$

The final objective is to measure the statistical mismatch of two common feature-level priors between the clean and low-quality domains. By minimizing the above objective, we mitigate the negative effects of this mismatch and encourage deep models to learn robust features during the decision-making process.

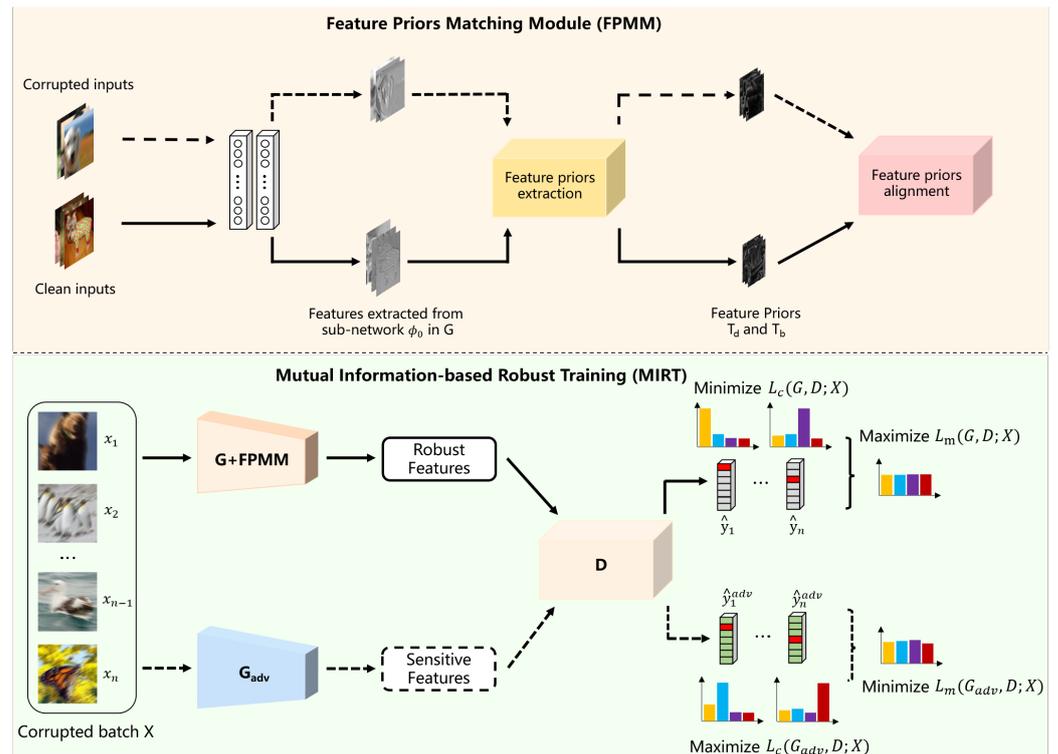


Figure 3. Top: The work flow of FPMM which extracts and aligns the feature-level priors from the clean and low-quality data. Bottom: The training detail of MIRT, in which the discriminator D receives the features from G with FPMM via mutual information maximization and rejects the features from G_{adv} by mutual information minimization.

4.2. Mutual Information-Based Robust Training

To enhance the discriminability of high-level semantic representations, we propose a novel min–max optimization strategy called Mutual Information-based Robust Training (MIRT), which complements the robustness of generic features in the shallow layers achieved by FPMM. Figure 3 provides details on the implementation of MIRT. Our approach relies on the intuition that, with the aid of FPMM, the feature generator G can obtain more robust representations from low-quality inputs than G_{adv} . For these inputs, the class discriminator D is motivated to differentiate between deep representations generated by G and those generated by G_{adv} . We employ mutual information between the inputs and their corresponding predicted outputs as the judgment criterion. Using (6) and (7), the estimation of mutual information over G and D can be calculated:

$$\begin{aligned} \mathcal{L}_{mi}(G, D; \mathcal{D}_t) = & \\ & H[\mathbb{E}_{x_j^t \sim \mathcal{D}_t}[p(y|x_j^t, G \circ D)]] - \mathbb{E}_{x_j^t \sim \mathcal{D}_t}[H[p(y|x_j^t, G \circ D)]], \end{aligned} \tag{12}$$

Similarly, we denote the estimation of mutual information over G_{adv} and D as $\mathcal{L}_{mi}(G_{adv}, D; \mathcal{D}_t)$. For low-quality inputs, G and D learn to maximize the mutual information in the same direction. As we discussed earlier, this also entails maximizing the marginal entropy and minimizing the conditional entropy. A lower conditional entropy compels D to widen the margin of decision boundaries for relatively robust features from G and encourages G to produce deep features far from these boundaries. A higher marginal entropy promotes a uniform distribution over the predictions of $G \circ D$. Additionally, the discriminator D engages in an adversarial game with the feature generator G_{adv} . Since G_{adv} extracts deep features that are more sensitive to low visibility conditions, D minimizes mutual information to refine the decision boundaries close to these sensitive features from G_{adv} . Consequently, D can successfully differentiate

between deep features generated by G or G_{adv} . G_{adv} optimizes in the opposite direction. For labeled clean inputs, we train G , G_{adv} , and D by minimizing the cross-entropy loss. The cross-entropy loss over G with D can be expressed as follows:

$$\mathcal{L}_y(G, D; \mathcal{D}_s) = \mathbb{E}_{(x_i^s, y_i^s) \sim \mathcal{D}_s} [-\log p(y = y_i^s | x_i^s, G \circ D)], \quad (13)$$

Similarly, $\mathcal{L}_y(G_{adv}, D; \mathcal{D}_s)$ denotes the cross-entropy loss over G_{adv} with D . Therefore, the final objective in MIRT can be formulated as:

$$\begin{aligned} & \min_{G_{adv}, D} \mathcal{L}_y(G_{adv}, D; \mathcal{D}_s), \\ & \max_{G_{adv}} \min_D \gamma \cdot \mathcal{L}_{mi}(G_{adv}, D; \mathcal{D}_t), \\ & \min_{G, D} \mathcal{L}_y(G, D; \mathcal{D}_s) + \alpha \cdot \mathcal{L}_{FPMM}(G; \mathcal{D}_s, \mathcal{D}_t) \\ & \quad - \beta \cdot \mathcal{L}_{mi}(G, D; \mathcal{D}_t), \end{aligned} \quad (14)$$

where α , β , and γ are the weight factors. The detailed PKAL procedure is provided in Algorithm 1. We note that G and D comprise the final model for inference in poor visibility conditions.

Algorithm 1: Prior Knowledge-guided Adversarial Learning

Input: $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, $\mathcal{D}_t = \{x_j^t\}_{j=1}^{N_t}$, Batch Size N .

Output: Learned model parameters $(\theta_G, \theta_{G_{adv}}, \theta_D)$.

Initialize $\theta_G, \theta_D, \theta_{G_{adv}}$ randomly.

for iter = 1 **to** MaxIter **do**

$B_s = \text{RANDOMSAMPLE}(\mathcal{D}_s, N)$;

$B_t = \text{RANDOMSAMPLE}(\mathcal{D}_t, N)$;

$\theta_{G_{adv}} \leftarrow \theta_{G_{adv}} - \nabla_{\theta_{G_{adv}}}(\mathcal{L}_y(G_{adv}, D; B_s)) + \gamma \cdot \nabla_{\theta_{G_{adv}}}(\mathcal{L}_{mi}(G_{adv}, D; B_t))$;

$\theta_D \leftarrow \theta_D - \nabla_{\theta_D}(\mathcal{L}_y(G_{adv}, D; B_s)) - \gamma \cdot \nabla_{\theta_D}(\mathcal{L}_{mi}(G_{adv}, D; B_t))$;

$\theta_G \leftarrow$

$\theta_G - \nabla_{\theta_G}(\mathcal{L}_y(G, D; B_s)) + \beta \cdot \nabla_{\theta_G}(\mathcal{L}_{mi}(G, D; B_t)) + \alpha \cdot \nabla_{\theta_G}(\mathcal{L}_{FPMM}(G; B_s, B_t))$;

$\theta_D \leftarrow \theta_D - \nabla_{\theta_D}(\mathcal{L}_y(G, D; B_s)) + \beta \cdot \nabla_{\theta_D}(\mathcal{L}_{mi}(G, D; B_t))$;

iter \leftarrow iter + 1;

end

5. Experiments

In this section, we introduce six self-constructed datasets that include various poor visibility scenes. These datasets cover natural and commonly encountered perturbations, such as visual blur, fog, rain, snow, and low illuminance. To simulate low visibility conditions, we collected datasets from real-world scenarios or generated them using SOTA synthesis algorithms. We then evaluate the performance of our proposed approach against fifteen existing solutions on these low-quality datasets.

5.1. Experiment Settings

5.1.1. Training Strategies

In our experiments, we utilize three backbone networks, namely, AlexNet, VGG19, and ResNet-18 [51–53]. For training our proposed approach, we use the SGD optimizer with an initial learning rate of 0.1 and a Cosine Annealing learning rate schedule. The batch size is set to 128, and the training epochs are set to 90 for all experiments. The hyper-parameter α

is kept at 0.1 throughout the training process. β and γ are set to zero in the first 15 epochs and then increased to 0.2 and 0.1, respectively, for the remaining training process.

5.1.2. Comparison Methods

We compare our proposed approach with fifteen existing solutions, which can be divided into four categories: image restoration, statistical alignment, adversarial domain adaptation, and data augmentation. Image restoration aims to improve the visual quality of low-quality inputs before they are applied to high-level tasks [20–22,25,28,54–57]. Statistical feature alignment belongs to typical unsupervised domain adaptation techniques, which seek invariant representations between the clean and low-quality domains by aligning the statistics of deep features [39,40]. Adversarial domain adaptation integrates adversarial learning and domain adaptation in a two-layer game, in which the feature generator and domain discriminator are adversarially trained to learn invariant representations [41–44,58]. We also include two data augmentation strategies that apply multiple operations of image transformation and stylization [36,37].

5.1.3. Visual Blur

Visual blur is a common perturbation encountered in real-world scenarios that can be caused by various factors, such as a shaky or out-of-focus camera, low exposure time, and fast-moving objects. To account for this, we construct two blurry datasets: REDS-BLUR and ImageNet-BLUR, each serving a distinct purpose. The REDS-BLUR dataset consists of 6170 clean images and 2155 blurry images in 11 classes, where the blurry subset is manually collected from the REDS dataset [59], a video deblurring dataset that utilizes deep learning techniques to simulate real-world motion blur. This dataset is designed to be comparable in size to typical unsupervised domain adaptation benchmarks, such as Image-CLEF, OFFICE-31, and OFFICE-Home.

In addition to motion blur, we introduce the ImageNet-BLUR dataset, which includes three additional types of blur: defocus blur, zoom blur, and glass blur. The clean subset consists of 129,377 high-quality images from 200 classes of ImageNet [60]. We generate an unlabeled blurry subset of 129,381 images using the synthesis pipeline in [11] to simulate the different types of blur. The testing set of ImageNet-BLUR includes five severity levels for each blur type, resulting in a total of $4 \times 5 \times 5000$ testing samples. Figure 4a–c show examples of the REDS-BLUR and ImageNet-BLUR datasets.

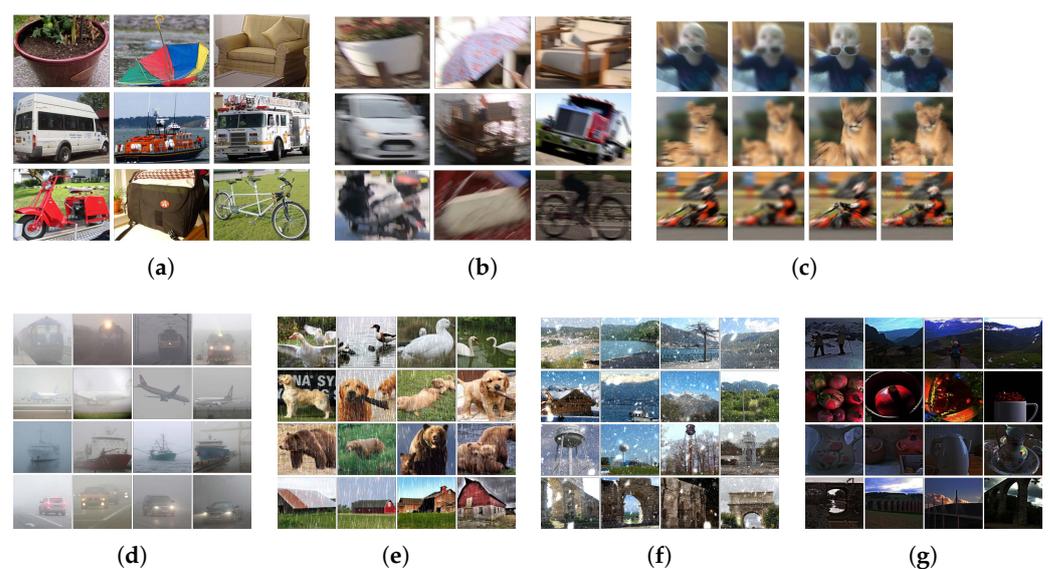


Figure 4. Some examples in the (a) REDS-BLUR clean set, (b) REDS-BLUR blurry set, (c) ImageNet-BLUR, (d) Web-FOG, (e) ImageNet-RAIN, (f) ImageNet-SNOW, and (g) ImageNet-DARK dataset.

5.1.4. Fog

Fog weather is a long-standing challenge for practical computer vision, leading to different amounts of partial occlusion in images. To simulate real-world fog, we manually collected fog images from the web and named the dataset Web-FOG. The dataset contains 4797 clean images and 4724 real-world fog images in 12 classes. The number of fog images in each class are: Bird (364), Boat (400), Bridge (426), Building (404), Bus (370), Car (410), People (388), Plane (462), Streetlamp (342), Train (412), Tree (404), and Truck (342). Some examples of the fog images from Web-FOG are shown in Figure 4d.

5.1.5. Rain

Rain is a common outdoor weather condition that degrades the visibility due to its scattering and blurry effects. To create a synthetic rain dataset, we generated rain streaks using a classic rain rendering algorithm [61] and selected background scenes from ImageNet. The rain rendering model can simulate real-world rain scenes by capturing the interactions between the lighting direction, viewing direction, and the oscillating shape of the rain streaks. We name this dataset ImageNet-RAIN, which contains 129,377 clean images and 129,381 unlabeled rain images in 200 classes. Examples of the rainy images from ImageNet-RAIN are displayed in Figure 4e.

5.1.6. Snow

Snow particles, such as snowflakes, often cause severe occlusion in images. To address this, we created the ImageNet-SNOW dataset, which synthesizes snow images from clean backgrounds from ImageNet and 2000 snow templates varying in transparency, size, and location from the CSD snow scene dataset [9]. Each snow image is generated by combining a clean image with a randomly selected snow template. The ImageNet-SNOW dataset is of a similar magnitude to ImageNet-RAIN. We present some samples of ImageNet-SNOW in Figure 4f.

5.1.7. Low Illuminance

Low-light conditions often occur due to inadequate light or under-exposed cameras, resulting in images with object shapes but discarded local details, such as image texture. Texture information plays a key role in semantic-level tasks, such as object recognition [34]. To evaluate our method in low-light conditions, we designed the ImageNet-DARK dataset. To mimic realistic low-light conditions, we use a two-stage synthesis strategy to adjust low-light distribution from the perspectives of local and global regions. Following the generation of low-light data [55], we retrain the ZeroDCE enhancement model [55] with the low-exposure parameter for local low-light adjustment. The original ZeroDCE with the normal exposure parameter restores low-light images, but the revised ZeroDCE generates low-light images through a reverse process. After local adjustment, we globally manipulate the exposure intensity to form the final visually similar version. The clean backgrounds of ImageNet-DARK are selected from ImageNet. We present some low-light examples in Figure 4g.

5.2. Experiment Results

5.2.1. Evaluation on Visual Blur

We evaluate the performance of our proposed approach and comparison methods on the REDS-BLUR dataset, and the results are presented in Table 1. Our approach outperforms the baselines, achieving gains of 13.9%, 11.4%, and 12.6% in AlexNet, ResNet-18, and VGG19, respectively. Among the deblurring algorithms, most of them have a positive effect on blurry object recognition, except for RL. Specifically, SRN achieves the largest increase of 14.7% in ResNet-18. For the UDA methods, their performance varies across different network structures. For example, CDAN obtains a gain of 7.1% in AlexNet but a drop of 6.5% in VGG19. In contrast, W-DANN and DIRT-T are the most stable approaches among them, improving 4.6% and 2.9% on average, respectively. Table 2 shows

the comparison of our approach and existing solutions on the ImageNet-BLUR dataset. Our PKAL approach demonstrates great efficacy in all blur severity levels and types, with an average improvement of 16%. The best performance of 42.4% for defocus blur and 49.0% for zoom blur show the great transferability of PKAL.

Table 1. Comparison of accuracy (%) on the REDS-BLUR dataset for the proposed method and potential solutions in different backbone networks.

	Methods	AlexNet	ResNet-18	VGG19	Avg.
–	Baseline	61.9	67.7	72.0	67.2
Image Restoration	RL [54]	53.9	54.8	48.5	52.4
	ECP [28]	64.4	69.3	65.1	66.3
	SRN [20]	73.1	82.4	81.3	78.9
	DeblurGAN [21]	66.1	72.2	70.4	69.6
	DeblurGANv2 [22]	67.2	77.3	74.0	72.8
Statistical Features Alignment	DAN [39]	65.9	68.4	68.4	67.6
	Deep CORAL [40]	59.6	66.5	63.3	63.1
Adversarial Domain Adaptation	DANN [41]	61.5	62.5	64.4	62.8
	CDAN [43]	69.0	68.3	65.5	67.6
	MCD [42]	45.3	69.5	65.0	59.9
	VADA [44]	61.4	66.6	73.1	67.0
	DIRT-T [44]	65.7	67.7	76.9	70.1
	W-DANN [58]	64.0	73.0	78.3	71.8
Proposed method	FPMM	64.7	69.3	70.8	68.3
	PKAL (FPMM+MIRT)	75.8	79.1	84.6	79.8

Table 2. The performance (%) of the proposed approach and comparison methods on the ImageNet-BLUR dataset.

Methods	Clean	Motion Blur					Other Kinds of Blur			Avg.
		1	2	3	4	5	Defocus Blur	Zoom Blur	Glass Blur	
Baseline	74.1	63.3	52.8	36.3	21.7	15.3	26.7	31.0	29.0	38.9
DAN [39]	70.7	68.6	61.4	49.0	32.9	24.7	33.3	44.9	31.9	46.4
Deep CORAL [40]	71.6	66.7	59.0	45.9	30.9	22.4	32.8	43.1	29.6	44.7
DANN [41]	73.7	72.3	67.2	55.6	39.0	29.3	38.5	47.3	38.0	51.2
W-DANN [58]	74.2	72.8	68.1	56.5	39.3	28.6	37.6	48.5	36.0	51.3
CDAN [43]	71.6	73.7	67.7	61.2	45.6	34.3	37.6	47.6	38.9	53.1
VADA [44]	70.0	71.1	67.1	56.2	38.7	29.1	41.0	48.4	45.4	51.9
RandAug [36]	71.8	64.9	55.0	40.7	25.7	18.1	28.7	39.1	30.9	41.7
AugMix [37]	71.4	67.2	62.8	53.4	39.5	30.2	35.9	45.6	35.9	49.1
FPMM	74.9	72.2	66.1	54.7	37.7	27.6	38.9	43.3	31.8	49.7
PKAL (FPMM+MIRT)	73.4	72.9	69.7	61.9	48.1	37.9	42.4	49.0	38.8	54.9

5.2.2. Evaluation on Fog

We evaluate the performance of our approach and comparison methods on the Web-FOG dataset, and the results are presented in Table 3. The proposed PKAL achieves the largest gain of 31.4% compared with the baseline, and even using the single FPMM achieves the leading increase of 19.1%, which is comparable to the performance of RandAug and AugMix. The UDA methods improve by above 18% on the foggy data, especially with a large positive margin of 26.0% for DIRT-T and 24.4% for DANN. However, FFA-Net, which has a dehazing effect, only shows a slight improvement of 2.3% compared with the baseline model.

Table 3. Comparison of accuracy (%) on the self-collected datasets in other poor visibility conditions including fog, rain, snow, and low illuminance. * denotes the different image restoration algorithm for each image perturbation: FFA-Net [25] on Web-FOG, MPRNet [57] on ImageNet-RAIN, DesnowNet [56] on ImageNet-SNOW, and Zero-DCE [55] on ImageNet-DARK.

Methods	Avg. Clean	Web-FOG	ImageNet-RAIN	ImageNet-SNOW	ImageNet-DARK
Baseline	79.3	57.4	34.9	32.8	25.5
Image Restoration Module *	79.3	59.7	64.4	44.5	40.8
DAN [39]	79.3	77.1	60.7	57.5	55.3
Deep CORAL [40]	79.6	75.7	59.4	54.5	54.0
DANN [41]	79.3	81.8	65.8	63.0	53.3
W-DANN [58]	79.4	81.0	62.5	60.6	55.1
CDAN [43]	80.0	81.7	65.6	63.2	55.9
VADA [44]	80.5	77.6	62.8	60.7	47.8
DIRT-T [44]	81.2	83.4	67.6	65.2	58.6
RandAug [36]	78.0	73.4	39.4	36.8	29.5
AugMix [37]	77.5	76.8	43.3	41.1	32.0
FPMM	80.3	76.5	54.3	49.1	49.0
PKAL (FPMM+MIRT)	80.7	88.7	68.3	67.3	58.1

5.2.3. Evaluation on Snow

Table 3 presents the results on the ImageNet-SNOW dataset. Both our proposed PKAL and single FPMM approaches demonstrate their efficacy in handling snow occlusion, achieving remarkable gains of 34.5% and 16.3%, respectively. RandAug and AugMix also exhibit improvements of 4.0% and 8.3%, respectively, consistent with their performance on the ImageNet-RAIN dataset. For domain adaptation, DIRT-T and CDAN outperform other domain adaptation approaches, with improvements of 32.4% and 30.4%, respectively. With its de-snowing effect, DesnowNet also improves model robustness under snow occlusion, with a gain of 11.7% compared to the baseline.

5.2.4. Evaluation on Low Illuminance

We present all results on the ImageNet-DARK dataset in Table 3. Consistent with the previous results, our proposed PKAL and single FPMM approaches exhibit leading performance in low illuminance conditions, achieving gains of 32.6% and 23.5%, respectively, compared to the baseline. UDA methods also improve model robustness, with positive gains ranging from 22.3% to 33.1%. DIRT-T exhibits the largest improvement among these approaches. RandAug and AugMix improve by 4.0% and 6.5%, respectively, in low illuminance situations, similar to their performance in rain and snow occlusion. In addition to improving the visual effect, the positive margin of 15.3% demonstrates the effectiveness of Zero-DCE in low illuminance situations.

5.2.5. Evaluation on Rain

Table 3 summarizes the results on the ImageNet-RAIN dataset. Our proposed PKAL approach achieves the best performance, improving by 33.4% compared to the baseline. The single FPMM approach also significantly improves model robustness against rain occlusion, with a gain of 19.4%. RandAug and AugMix both achieve moderate gains of 4.5% and 8.4%, respectively. For UDA methods, adversarial-based adaptation outperforms statistical feature alignment. CDAN and DAN both exhibit improvements of 30.7% and 25.8%, respectively. MPRNet achieves a remarkable gain of 29.5% through its de-raining effect.

6. Discussion

6.1. The Benefit of PKAL

We validate the discriminative ability of the deep representations learned from our approach using t-Distributed Stochastic Neighbor Embedding (t-SNE) [62], a non-linear visualization for high-dimensional features. In Figure 5, we show the t-SNE visualizations of deep features from different methods for Web-FOG. While FPMM and UDA methods show a similar effect by separating easily identified categories such as people, tree, and bird, they still struggle to distinguish between several classes with similar semantic content such as car, bus, and truck, which remain close to each other. In contrast, the t-SNE result of PKAL demonstrates its ability to further increase the intra-class compactness and inter-class discrepancy of hard-to-classify examples in the deep feature space.

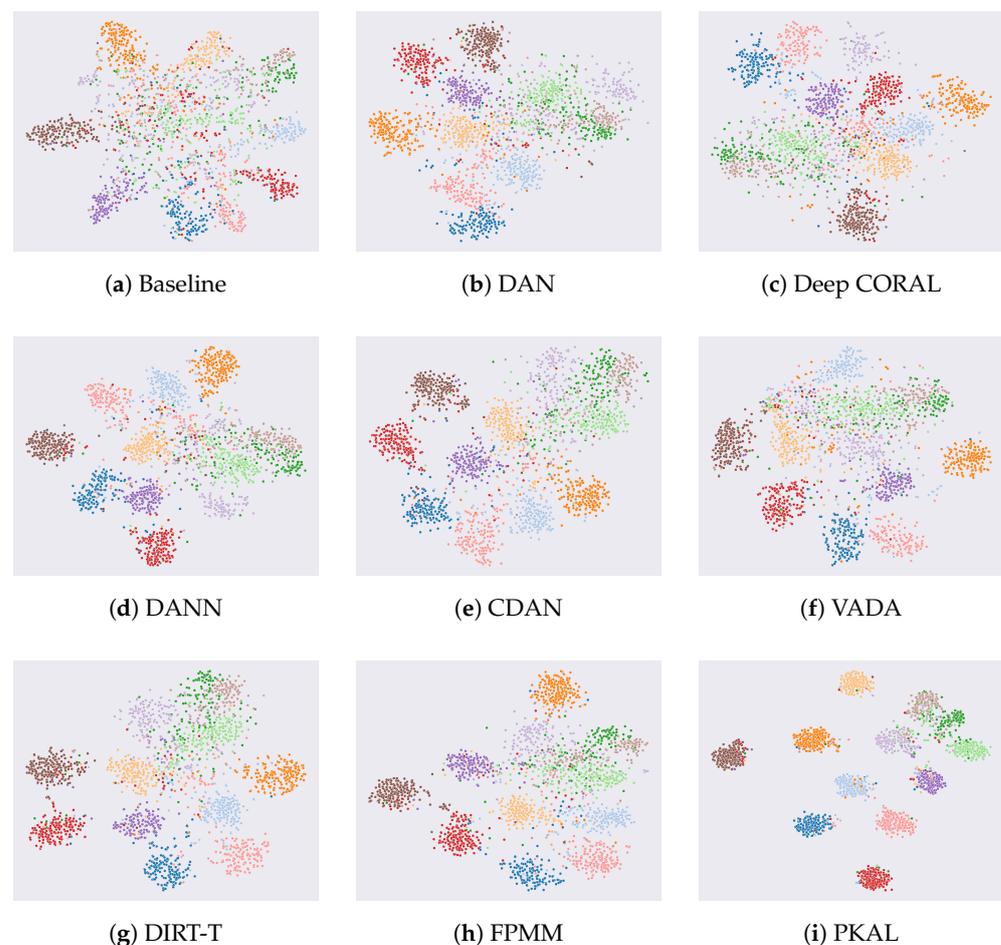


Figure 5. The t-SNE visualizations of deep semantic features for different approaches. Each color represents a different class label.

We apply Grad-CAM to visualize the informative regions of our proposed approach by projecting learned deep representations back onto the raw pixels. Figure 6 shows that the baseline model learns salient features mainly on the background rather than on the object, resulting in inaccurate predictions due to the fog occlusion. On the other hand, the attention of FPMM successfully focuses on the object to be recognized, highlighting its informative features. Furthermore, our results suggest that deep features learned from PKAL can better reflect the informative content of the image.

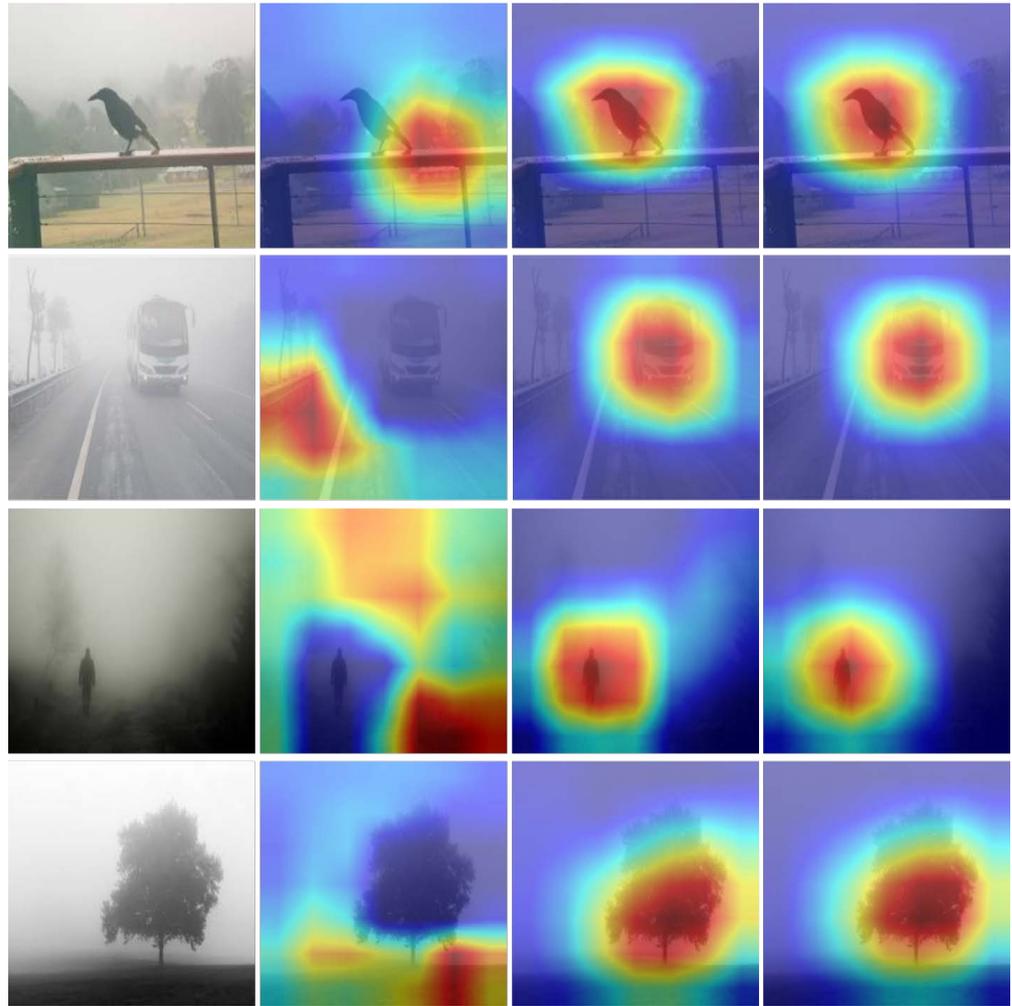


Figure 6. From left to right, we show raw foggy images and their Grad-CAM heat maps from the baseline model and the proposed FPMM and PKAL. The red regions represent the key information used by the model for decision-making.

6.2. The Effect of FPMM

In order to assess the effectiveness of FPMM in our approach, we utilize Corresponding Angle (CA) [63] to verify the alignment effect on the shallow features between the clean and corrupted data. We obtain the eigenvectors U_s^k of the k -th channel low-level feature from the clean image and U_t^k from the low-quality image through Singular Value Decomposition (SVD). The cosine value of CA can be computed as follows.

$$\cos(\psi_i^k) = \frac{\langle u_{s,i}^k, u_{t,i}^k \rangle}{\|u_{s,i}^k\| \|u_{t,i}^k\|} \quad (15)$$

where $u_{s,i}^k$ and $u_{t,i}^k$ indicate the i -th eigenvector in U_s^k and U_t^k , respectively, with the i -th largest singular value. We improve the visualization of our results by computing $1 + \cos(\psi_i^k)$. This value is indicative of the feature similarity between the clean and low-quality domains, with a larger cosine similarity implying a better feature similarity. Our experiments involve computing the channel-wise cosine similarities for the top 10 CAs in the baseline model with and without the FPMM. Figure 7a demonstrates that the deep model with FPMM exhibits larger CA values compared to the baseline model without FPMM. These results indicate that the FPMM can help deep models learn domain-invariant shallow features from low-quality data.

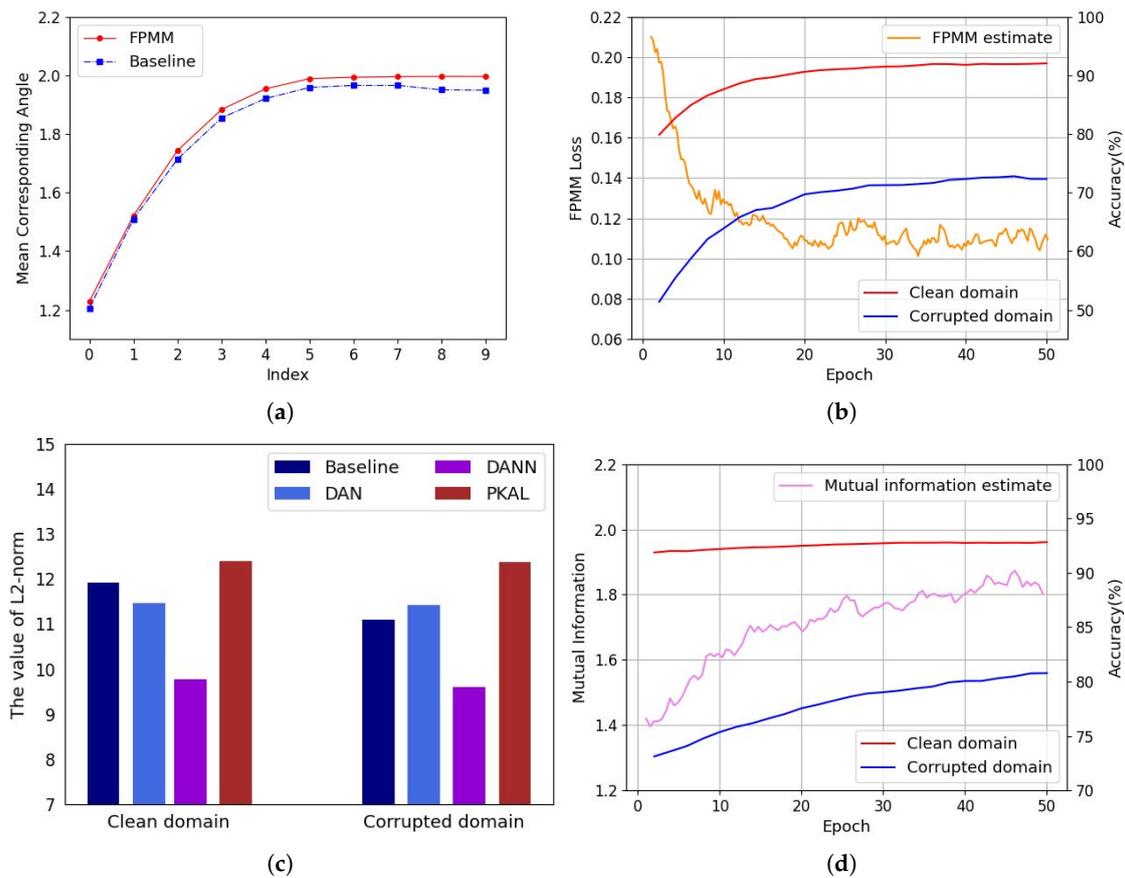


Figure 7. (a) The comparison of the top 10 corresponding angles in the shallow features between the baseline and FPMM. (b) The dynamic record of the FPMM loss in the training process. (c) The L_2 norm of deep features for PKAL and other approaches. (d) The variation of the mutual information estimate during the training time.

Tables 1–3 demonstrate the effectiveness of FPMM, which improves model performance on both clean and low-quality data. Figure 7b depicts the change in FPMM loss and testing accuracy during the training process. As the FPMM loss decreases, the testing accuracy on both clean and corrupted data increases, demonstrating the benefits of reducing prior knowledge-based feature discrepancy. Moreover, the FPMM method requires little training time since it updates only a few parameters of the shallow layers in the back-propagation process.

6.3. The Effect of MIRT

In this section, we explore the effect of MIRT on the norm of deep features, which has been shown to represent discriminability and transferability across domains [63,64]. We compare the L_2 feature norm of the final output of the feature generator G across different approaches. Figure 7c illustrates the L_2 feature norm in the clean domain (left) and low-quality domain (right). Notably, MIRT achieves a larger L_2 feature norm compared to other approaches, indicating the better adaptability of deep features learned from MIRT. Furthermore, Figure 7d demonstrates the change in mutual information estimate and classification accuracy during the training period. We observe a similar trend of performance improvement on the low-quality data as the mutual information estimate increases, verifying the benefits of the mutual information-based adversarial learning in MIRT.

7. Conclusions

In this study, we propose PKAL, a novel approach for visual recognition in poor visibility conditions. The PKAL method integrates FPM, a feature matching module that reduces feature discrepancy between clean and low-quality domains, and MIRT, a robust learning strategy that refines discriminative semantic features and task-specific decision boundaries for low-quality data through adversarial learning based on mutual information. Our proposed approach is evaluated on five typical low visibility conditions, including visual blur, fog, rain, snow, and low illuminance. The experimental results demonstrate consistent performance gains across various low visibility conditions, underscoring the effectiveness of our approach.

Author Contributions: Conceptualization, J.Y. (Jiangang Yang) and J.Y. (Jianfei Yang); Methodology, J.Y. (Jianfei Yang), J.L. and S.W.; Data curation, L.L. and J.Y. (Jiangang Yang); Writing, J.Y. (Jiangang Yang) and L.L.; Project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by SunwayAI computing platform (SXHZ202103) and the National Key Research and Development Program (2021YFB2501403).

Data Availability Statement: The data and code that support the findings and experiments of this study are openly available in CodeOcean at <https://codeocean.com/capsule/3516791/tree> on 8 August 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
2. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
3. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef] [PubMed]
4. Pei, Y.; Huang, Y.; Zou, Q.; Lu, Y.; Wang, S. Does haze removal help cnn-based image classification? In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 682–697.
5. VidalMata, R.G.; Banerjee, S.; RichardWebster, B.; Albright, M.; Davalos, P.; McCloskey, S.; Miller, B.; Tambo, A.; Ghosh, S.; Nagesh, S.; et al. Bridging the gap between computational photography and visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4272–4290. [CrossRef]
6. Cai, J.; Zuo, W.; Zhang, L. Dark and bright channel prior embedded network for dynamic scene deblurring. *IEEE Trans. Image Process.* **2020**, *29*, 6885–6897. [CrossRef]
7. Li, S.; Araujo, I.B.; Ren, W.; Wang, Z.; Tokuda, E.K.; Junior, R.H.; Cesar, R.; Zhang, J.; Guo, X.; Cao, X. Single image deraining: A comprehensive benchmark analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3838–3847.
8. Zhu, H.; Peng, X.; Chandrasekhar, V.; Li, L.; Lim, J.H. Dehazegan: When image dehazing meets differential programming. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2019; pp. 1234–1240.
9. Chen, W.T.; Fang, H.Y.; Hsieh, C.L.; Tsai, C.C.; Chen, I.; Ding, J.J.; Kuo, S.Y. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4196–4205.
10. Arruda, V.F.; Paixao, T.M.; Berriel, R.F.; De Souza, A.F.; Badue, C.; Sebe, N.; Oliveira-Santos, T. Cross-domain car detection using unsupervised image-to-image translation: From day to night. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
11. Hendrycks, D.; Dietterich, T. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv* **2019**, arXiv:1903.12261.
12. Hendrycks, D.; Basart, S.; Mu, N.; Kadavath, S.; Wang, F.; Dorundo, E.; Desai, R.; Zhu, T.; Parajuli, S.; Guo, M.; et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 8340–8349.
13. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. Autoaugment: Learning augmentation strategies from data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 113–123.
14. Zhang, S.; Zhen, A.; Stevenson, R.L. GAN based image deblurring using dark channel prior. *arXiv* **2019**, arXiv:1903.00107.
15. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353.

16. Flores, M.; Valiente, D.; Gil, A.; Reinoso, O.; Paya, L. Efficient probability-oriented feature matching using wide field-of-view imaging. *Eng. Appl. Artif. Intell.* **2022**, *107*, 104539. [[CrossRef](#)]
17. Bei, W.; Fan, X.; Jian, H.; Du, X.; Yan, D. GeoGlue: Feature matching with self-supervised geometric priors for high-resolution UAV images. *Int. J. Digit. Earth* **2023**, *16*, 1246–1275. [[CrossRef](#)]
18. Son, C.H.; Ye, P.H. New Encoder Learning for Captioning Heavy Rain Images via Semantic Visual Feature Matching. *arXiv* **2021**, arXiv:2105.13753.
19. Hjelm, R.D.; Fedorov, A.; Lavoie-Marchildon, S.; Grewal, K.; Bachman, P.; Trischler, A.; Bengio, Y. Learning deep representations by mutual information estimation and maximization. *arXiv* **2018**, arXiv:1808.06670.
20. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182.
21. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8183–8192.
22. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8878–8887.
23. Guo, Q.; Sun, J.; Juefei-Xu, F.; Ma, L.; Xie, X.; Feng, W.; Liu, Y.; Zhao, J. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 1487–1495.
24. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3937–3946.
25. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
26. Pan, J.; Sun, D.; Pfister, H.; Yang, M.H. Blind image deblurring using dark channel prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1628–1636.
27. Chen, L.; Fang, F.; Wang, T.; Zhang, G. Blind image deblurring with local maximum gradient prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1742–1750.
28. Yan, Y.; Ren, W.; Guo, Y.; Wang, R.; Cao, X. Image deblurring via extreme channels prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4003–4011.
29. Zhang, J.; Pan, J.; Ren, J.; Song, Y.; Bao, L.; Lau, R.W.; Yang, M.H. Dynamic scene deblurring using spatially variant recurrent neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2521–2529.
30. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1222–1230.
31. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [[CrossRef](#)]
32. DeVries, T.; Taylor, G.W. Improved regularization of convolutional neural networks with cutout. *arXiv* **2017**, arXiv:1708.04552.
33. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6023–6032.
34. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv* **2018**, arXiv:1811.12231.
35. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
36. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2019; pp. 702–703.
37. Hendrycks, D.; Mu, N.; Cubuk, E.D.; Zoph, B.; Gilmer, J.; Lakshminarayanan, B. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv* **2019**, arXiv:1912.02781.
38. Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv* **2014**, arXiv:1412.3474.
39. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. In Proceedings of the International Conference on Machine Learning (PMLR), Lille, France, 6–11 July 2015; pp. 97–105.
40. Sun, B.; Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. In Proceedings of the Computer Vision–ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10, 15–16 October 2016; pp. 443–450.
41. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning (PMLR), Lille, France, 6–11 July 2015; pp. 1180–1189.

42. Saito, K.; Watanabe, K.; Ushiku, Y.; Harada, T. Maximum classifier discrepancy for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3723–3732.
43. Long, M.; Cao, Z.; Wang, J.; Jordan, M.I. Conditional adversarial domain adaptation. In Proceedings of the Advances in Neural Information Processing Systems 31, Montreal, QC, Canada, 3–8 December 2018.
44. Shu, R.; Bui, H.H.; Narui, H.; Ermon, S. A dirt-t approach to unsupervised domain adaptation. *arXiv* **2018**, arXiv:1802.08735.
45. Krause, A.; Perona, P.; Gomes, R. Discriminative clustering by regularized information maximization. In Proceedings of the Advances in Neural Information Processing Systems 23, Vancouver, BC, Canada, 6–9 December 2010.
46. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [[CrossRef](#)]
47. Springenberg, J.T. Unsupervised and semi-supervised learning with categorical generative adversarial networks. *arXiv* **2015**, arXiv:1511.06390.
48. Poole, B.; Ozair, S.; Van Den Oord, A.; Alemi, A.; Tucker, G. On variational bounds of mutual information. In Proceedings of the International Conference on Machine Learning (PMLR), Long Beach, CA, USA, 9–15 June 2019; pp. 5171–5180.
49. Cheng, P.; Hao, W.; Dai, S.; Liu, J.; Gan, Z.; Carin, L. Club: A contrastive log-ratio upper bound of mutual information. In Proceedings of the International Conference on Machine Learning (PMLR), Virtual, 13–18 July 2020; pp. 1779–1788.
50. Wang, F.; Kong, T.; Zhang, R.; Liu, H.; Li, H. Self-Supervised Learning by Estimating Twin Class Distribution. *IEEE Trans. Image Process.* **2023**, *32*, 2228–2236. [[CrossRef](#)] [[PubMed](#)]
51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25, Lake Tahoe, NV, USA, 3–6 December 2012.
52. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 17–30 June 2016; pp. 770–778.
54. Richardson, W.H. Bayesian-based iterative method of image restoration. *JoSA* **1972**, *62*, 55–59. [[CrossRef](#)]
55. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
56. Liu, Y.; Jaw, D.W.; Huang, S.C.; Hwang, J.N. DesnowNet: Context-aware deep network for snow removal. *IEEE Trans. Image Process.* **2018**, *27*, 3064–3073. [[CrossRef](#)] [[PubMed](#)]
57. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.
58. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning (PMLR), Sydney, NSW, Australia, 6–11 August 2017; pp. 214–223.
59. Nah, S.; Baik, S.; Hong, S.; Moon, G.; Son, S.; Timofte, R.; Mu Lee, K. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Nashville, TN, USA, 19–25 June 2019.
60. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
61. Garg, K.; Nayar, S.K. Photorealistic rendering of rain streaks. *ACM Trans. Graph. (TOG)* **2006**, *25*, 996–1002. [[CrossRef](#)]
62. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
63. Chen, X.; Wang, S.; Long, M.; Wang, J. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In Proceedings of the International Conference on Machine Learning (PMLR), Long Beach, CA, USA, 9–15 June 2019; pp. 1081–1090.
64. Cui, S.; Wang, S.; Zhuo, J.; Li, L.; Huang, Q.; Tian, Q. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3941–3950.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.