

# IPF-RDA: An Information-Preserving Framework for Robust Data Augmentation

Suorong Yang , Hongchao Yang, Suhan Guo, Furao Shen , and Jian Zhao , *Senior Member, IEEE*

**Abstract**—Data augmentation is widely utilized as an effective technique to enhance the generalization performance of deep models. However, data augmentation may inevitably introduce distribution shifts and noises, which significantly constrain the potential and deteriorate the performance of deep networks. To this end, we propose a novel information-preserving framework, namely IPF-RDA, to enhance the robustness of data augmentations in this paper. IPF-RDA combines the proposal of (i) a new class-discriminative information estimation algorithm that identifies the points most vulnerable to data augmentation operations and corresponding importance scores; And (ii) a new information-preserving scheme that preserves the critical information in the augmented samples and ensures the diversity of augmented data adaptively. We divide data augmentation methods into three categories according to the operation types and integrate these approaches into our framework accordingly. After being integrated into our framework, the robustness of data augmentation methods can be enhanced and their full potential can be unleashed. Extensive experiments demonstrate that although being simple, IPF-RDA consistently improves the performance of numerous commonly used state-of-the-art data augmentation methods with popular deep models on a variety of datasets, including CIFAR-10, CIFAR-100, Tiny-ImageNet, CUHK03, Market1501, Oxford Flower, and MNIST, where its performance and scalability are stressed.

**Index Terms**—Robust data augmentation, deep learning, model generalization, interpretability.

## I. INTRODUCTION

**D**ATA augmentation (DA) has emerged as an effective technique for enhancing model optimization by providing diverse training data, which has been widely used in various

Received 18 September 2023; revised 19 August 2025; accepted 18 September 2025. Date of publication 22 September 2025; date of current version 9 January 2026. This work was supported in part by the STI 2030-Major Projects of China under Grant 2021ZD0201300, in part by the Fundamental Research Funds for the Central Universities under Grant 2024300394, and in part by the National Natural Science Foundation of China under Grant 62276127. Recommended for acceptance by K.-J. Yoon. (*Corresponding authors: Suorong Yang; Furao Shen.*)

Suorong Yang and Hongchao Yang are with the State Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China (e-mail: sryang@smail.nju.edu.cn; yanghc@smail.nju.edu.cn).

Suhan Guo and Furao Shen are with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China, and also with the School of Artificial Intelligence, Nanjing University, Nanjing 210023, China (e-mail: shguo@smail.nju.edu.cn; frshen@nju.edu.cn).

Jian Zhao is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: jianzhao@nju.edu.cn).

The implementation is available at <https://github.com/Jackbrocp/IPF-RDA>.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TPAMI.2025.3613005>, provided by the authors.

Digital Object Identifier 10.1109/TPAMI.2025.3613005

tasks [1], [2], [3]. Recently, many DA methods [4], [5], [6], [7], [8], [9], [10] have been proposed, achieving noteworthy results. Despite the effectiveness, recent research has underscored some challenges with DA: 1) Existing DA techniques face a dilemma between diversifying the data and maintaining its consistency. While DA methods can bolster model performance, they might simultaneously introduce distribution shifts, compromising performance on non-augmented data during inference [8], [9], [11]. Particularly concerning is the latent risk that DA may distort critical information for classification, resulting in mislabeled or ambiguously labeled augmented data. 2) DA methods primarily focus on globally transforming images using a unified strategy, with reduced attention given to the local discriminative context in individual images. This may result in a bottleneck for performance improvement [10], [12]. Consequently, these observations highlight the limitations of existing DA techniques, suggesting that inattentively designed augmentations could subtly but adversely affect the generalization capability of neural models. To avoid this, KeepAugment [9] has recently been proposed, utilizing the saliency map to assess the importance of pixels in images. At each training iteration, it performs data augmentation by keeping these salient points untouched. Although KeepAugment has achieved improvements over Cutout [4], CutMix [5], AutoAugment [6], and RandAugment [7], it is based on a basic assumption that the saliency map accurately identifies the points most sensitive to the classification decision; thus, these points are also the most vulnerable ones to the influence introduced by the DA operations. However, some publications have empirically shown that saliency methods do not provide insightful explanations to the classification decisions and cannot be used as the basis for classification [13], [14], [15], [16], rendering them non-class-discriminative. Meanwhile, the online importance scoring mechanism also introduces considerable computational overhead to the training process. To address this important but rarely studied challenge, an imminent problem arises: *How can we preserve the class-discriminative information without sacrificing the diversity of augmented data brought about by DA, hence enhancing its robustness?* By enhancing the methods' robustness, we can further improve the efficacy of DA methods and unleash their full potential.

In this work, we propose a novel Information-Preserving Framework for Robust Data Augmentation (IPF-RDA). This framework is motivated by the intriguing observation made by recent work showing that data augmentation may introduce distribution shift and consequently hurt the performance on unaugmented data during inference [8], [9]. The idea of our

method is straightforward yet effective: IPF-RDA identifies the most class-discriminative information in an offline manner and preserves this information while maintaining the diversity of augmented data during training. To achieve this, first, we theoretically prove that the vanilla saliency map is not sensitive to the classification results. Based on the theoretical proof, we introduce our offline class-discriminative information estimation algorithm (CDIEA), which identifies the points at which minor perturbations lead to misclassification. These points represent the most vulnerable features to data augmentation operations. Leveraging these critical points and their associated importance scores, IPF-RDA adaptively preserves a randomly selected rectangular region containing the class-discriminative information in each epoch during online training. The use of rectangular preservation ensures that the most critical features remain intact while maintaining the structural semantic characteristics of the object and its local context [4], [5], [9], [10], [17], combining both practical efficiency and effectiveness. Furthermore, by departing the CDIEA from the online training process, our framework imposes minimal additional training overhead during the online augmentation phase. Notably, our framework is highly scalable as it can significantly improve a number of state-of-the-art (SOTA) and widely adopted DA methods, including but not limited to Cutout [4], AdvMask [10], GridMask [17], Hide-and-Seek (HaS) [18], CutMix [5], RandomErasing [19], AutoAugment [6], Fast-AutoAugment [20], RandAugment [7], TrivialAugment [8], etc. Consequently, the main benefit of our framework lies in its ability to enhance robustness, improve the performance of data augmentation methods, and unleash their full potential, thereby facilitating optimized model training. Although the idea is simple, the proposed IPF-RDA is significantly effective and easy to implement. Extensive empirical evaluations, including supervised and semi-supervised image classification on CIFAR-10/100 [21], and Tiny-ImageNet [22], person re-identification (ReID) on CUHK03 [23] and Market1501 [24], and a series of qualitative and quantitative analyses are conducted. Experimental results demonstrate that IPF-RDA consistently improves the generalization performance of representative and widely used SOTA DA methods.

In summary, our contributions are:

- We introduce an information-preserving framework designed to integrate the most representative and widely used data augmentation methods, thereby further improving their performance by reducing the side effects caused by the augmentation operations, which are overlooked in prior works.
- Our study begins with a theoretical clarification highlighting that the vanilla saliency map is not sensitive to classification determinants. Next, we propose an offline algorithm for estimating class-discriminative information, which is designed to enhance data augmentation robustness.
- We develop a taxonomy that categorizes methods based on their operational types and subsequently integrates DA methods into our information-preserving framework according to their categories.
- Experimental results on various standard benchmark datasets demonstrate the efficacy of our framework. IPF-RDA can be employed to effectively enhance the

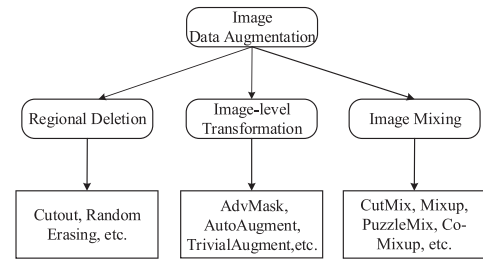


Fig. 1. A new taxonomy of image data augmentation approaches.

robustness of SOTA DA methods, thus significantly improving their performance on various neural architectures.

## II. RELATED WORK

Over recent years, many image data augmentation methods have been proposed based on different perspectives of image manipulations. As illustrated in Fig. 1, we introduce a new taxonomy that categorizes data augmentation approaches into three types: regional deletion, image-level transformation, and image mixing.

*Regional deletion* techniques, such as Cutout [4] and Random Erasing [19], have been commonly employed. Cutout [4] is a simple regularization technique that randomly masks out rectangular regions of input images. Random Erasing [19] aims to enhance the model's robustness to occlusion by randomly selecting a rectangular region in an image and replacing its pixels with random values. Despite the benefits of augmenting images with diverse positions and sizes of deleted regions, there exists a significant risk of eliminating critical areas during the augmentation process [9], [10].

To address this concern, our proposed framework integrates these methods but prevents them from removing highly class-discriminative regions, thereby enhancing the robustness of regional deletion-based DA approaches.

*Image-level transformation* includes image-level deletion-based methods such as Hide-and-Seek (HaS), GridMask [17], AdvMask [10], and many automatic augmentation methods. HaS involves randomly hiding multiple patches within training samples. GridMask [17] utilizes structured dropping regions, implicitly emphasizing the importance of balancing information reservation and deletion. AdvMask [10] first utilizes sparse adversarial attack points as key points and randomly masks some of them to force models to learn discriminative information in the images while some critical information is missing. Unlike AdvMask, our approach applies adaptive importance scores to evaluate the significance of critical points, indicating different levels of importance. Nevertheless, to foster data augmentation diversity, the masked regions of these methods are randomly generated. Although these methods can enhance the robustness of models to occlusions, such random deletion may inadvertently remove discriminative information crucial for classification or even delete objects of interest.

Automatic augmentation includes AutoAugment [6], Fast-AutoAugment [20], RandAugment [7], and TrivialAugment [8], etc. These methods primarily leverage reinforcement learning to explore optimal augmentation strategies and corresponding

parameters for each dataset in an offline manner. During the augmentation process, these methods randomly sample augmentation policies and strength settings in augmentation spaces. For instance, both AutoAugment [6] and Fast-AutoAugment [20] define the problem of finding the best augmentation policies as a discrete search problem and automatically search for the optimal combination of DA policies via reinforcement learning. RandAugment [7] adopts a significantly reduced search space, allowing it to be trained on the target task without a separate proxy task. To maintain the diversity of augmented data [25], these methods simultaneously apply different augmentation policies. However, the simultaneous application of multiple operations on a single image can potentially introduce distribution discrepancies [8], [9], [26]. In contrast, TrivialAugment [8] adopts an empirical approach, applying a single augmentation to each image. By employing a single augmentation policy, the rationale of TrivialAugment is to implicitly mitigate the impact of augmentation operations on images, effectively preserving the original image information.

We integrate image-level transformation-based DA methods into our proposed framework by explicitly recovering the critical information in images regardless of the extent to which image-level transformation methods alter images. Therefore, our framework is capable of enhancing the robustness of these DA methods while maintaining the diversity inherent in the original augmentation processes, thereby fully harnessing their potential.

*Image mixing*, such as Mixup [27] and CutMix [5], has gained attention in the field of data augmentation. PuzzleMix [28], a mixup method, explicitly utilizes the saliency information and the underlying statistics of the natural examples to enhance generalization. Meanwhile, Co-Mixup [29] introduces a new perspective on batch mixup by formulating an optimization objective that maximizes the data saliency of each individual mixup sample while encouraging supermodular diversity across the batch. These methods combine two images or sub-regions. Meanwhile, labels of the augmented data are fused based on the area proportions of the replaced regions. However, a challenge arises when critical regions essential for classification are affected or covered during the mixing process. This leads to ambiguous labels for the augmented data. For example, consider two images  $x_a$  and  $x_b$ , along with the corresponding labels  $y_a$  and  $y_b$ . CutMix randomly selects a region in  $x_a$  and replaces it with the corresponding region from  $x_b$ :

$$\begin{aligned}\tilde{x} &= \mathbf{M} \odot x_a + (\mathbf{1} - \mathbf{M}) \odot x_b \\ \tilde{y} &= \omega y_a + (1 - \omega) y_b\end{aligned}\quad (1)$$

where  $\mathbf{M} \in \{0, 1\}^{W \times H}$  denotes a binary mask indicating where to drop out or fill in from two images (i.e., filling with 0 within the selected region, otherwise with 1),  $\mathbf{1}$  is a binary mask filled with ones. Suppose that the selected region is of size  $w_M \times h_M$ ,  $\omega = \frac{w_M h_M}{WH}$  is determined as the proportion of the area of  $x_b$  in the composite image  $\tilde{x}$ . However, this kind of combination may inevitably introduce noisy samples when the area of the selected region is large, but there is no object from label  $y_b$  in that region. In this way, the augmented label is mainly  $y_b$  while the augmented image only contains objects of  $y_a$ .

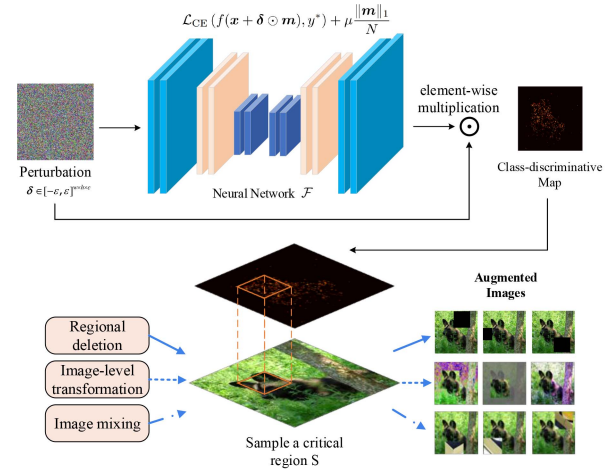


Fig. 2. Overview of the proposed framework.

To address this issue, our framework integrates CutMix by first replacing a less important region in  $x_a$  with a corresponding region from  $x_b$  to preserve the information from  $y_a$ . Meanwhile, we improve the calculation of  $\omega$  in our framework. Instead of relying on the area proportions, the weights of labels are calculated based on the importance scores of the selected regions.

*Information preserving* has been implicitly applied in more recent data augmentation methods [8], [10], [11], while KeepAugment [9] first explicitly preserves salient regions in the augmented data. Specifically, KeepAugment focuses on enhancing the fidelity of augmented images by leveraging the vanilla saliency map to identify informative regions in the original images. These informative regions are preserved during the augmentation process. However, it has been shown empirically that the vanilla saliency map is not sensitive to the classification decision [13], [14], [15], [16]. Meanwhile, the calculation of the saliency map in each iteration increases the training costs of the augmentation process. Instead, our framework utilizes an offline algorithm, CDIEA, to identify the most class-discriminative regions. Compared with KeepAugment, the proposed framework can improve upon a larger number of DA methods, including but not limited to Cutout, Random Erasing, HaS, GridMask, AdvMask, AutoAugment, RandAugment, TrivialAugment, etc., thus demonstrating higher generality.

Another line of related work is based on deep generative models [30], [31]. By training a generative adversarial network, one could utilize samples from the generator as the augmented data. Unfortunately, their adoption remains limited within the data augmentation research community due to significant computational requirements and the need for substantial training data. Specifically, the complexity associated with training generative networks and inferring augmented samples from them presents nontrivial challenges, limiting the widespread use of these methods [32].

### III. PROPOSED METHOD

As illustrated in Fig. 2, the proposed method consists of two components: the Class-Discriminative Information Estimation Algorithm (CDIEA) in the upper portion and the

Information-Preserving Framework (IPF) in the lower portion of the diagram. The former, CDIEA, operates offline and is designated to identify the most class-discriminative information (i.e., the positions and importance scores of critical pixels). Concurrently, IPF is responsible for preserving this critical information, which ensures robust augmentations throughout online training. In this section, we first provide a theoretical analysis in Section III-A to show that saliency methods are unsuitable for identifying class-discriminative information. Based on the theoretical analysis, we describe our proposed end-to-end CDIEA in Section III-B. Subsequently, in Section III-C, we describe how classification-sensitive information can be utilized to enhance the robustness of different types of DA techniques.

### A. Theoretical Analysis

Although KeepAugment shares a similar concept of preserving critical regions for improved performance, empirical studies from [16], [33], [34] have demonstrated that saliency maps exhibit a lack of sensitivity to classification decisions. In this section, we offer theoretical substantiation to corroborate this observation by establishing that conventional gradient-based saliency maps do not possess class-discriminative properties.

The saliency map involves computing the gradient of the loss function for the class of interest w.r.t. the input pixels  $x$ , denoted as  $g_x = \frac{\partial \mathcal{L}_c(x)}{\partial x}$ . To assess the sensitivity of the saliency map to classification decisions, we consider a successful adversarial attack on the model  $f$  by applying a perturbation  $\delta$  on the input image  $x$  of class  $y$ , that is,  $f(x + \delta) \neq y$ . Therefore, this perturbation  $\delta$  has already modified the classification decision. Subsequently, we proceed to examine the saliency map of an adversarial sample  $x + \delta$  using a ReLU network  $f$ . Before going into the main results, we need the following proposition in [35] for our proof.

**Proposition 1:** [35] Let  $\lambda_i$  and  $v_i$  be the  $i$ -th largest eigenvalue of the Hessian matrix and its corresponding eigenvector, respectively.  $\gamma_i \in \mathbb{R}$  is the projection length of  $g_x$  on  $v_i$ , and  $\beta$  is the update step used to obtain the adversarial sample. We have,

$$\delta = \sum_i \frac{(1 + \beta \lambda_i)^m - 1}{\lambda_i} \gamma_i v_i + \mathcal{R}_2(\beta) \quad (2)$$

**Lemma 1:** Let us assume the adversarial attack is accomplished in  $m$  steps. We establish the relationship between the saliency maps of sample  $x$  and the adversarial sample  $x + \delta$  in  $f$  as follows:

$$g_{x+\delta} = g_x + H_x \sum_i \frac{(1 + \beta \lambda_i)^m - 1}{\lambda_i} \gamma_i v_i + H_x \mathcal{R}_2(\beta) \quad (3)$$

where  $H_x$  is the Hessian matrix defined as  $H_x \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x \partial x^T} \mathcal{L}(f(x), y)$  and  $\mathcal{R}_2(\beta)$  represents the terms of  $\beta$  that are at least of second order.

Lemma (1) is proven in Appendix A, available online. Based on Lemma 1, we can see that the gradient of the adversarial sample can be approximately derived based on the model parameters regardless of specific  $\delta$ . Furthermore, the difference in the saliency map of the adversarial sample and the

initial sample can be approximated as  $\Delta_g = g_{x+\delta_m} - g_x$ . Considering that Hessian matrix  $H_x$  is a real-valued matrix and can be decomposed as  $H_x = V\Lambda V^{-1}$ , where  $\Lambda$  is a diagonal matrix:  $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_r]$ .  $V = [v_1, v_2, \dots, v_n] \in \mathbb{R}^{n \times n}$  contains  $n$  linearly independent eigenvectors, which means that  $v_i^T * v_j = 0$ , for  $i \neq j$ . Without loss of generality, we normalize the eigenvectors,  $V^T V = I$ . In this case,  $H_x$  can be decomposed as  $H_x = V\Lambda V^T$ . Therefore, it is possible to derive an easy-to-compute upper bound for  $\Delta_g$ , as given by the following proposition.

**Proposition 2:** Suppose that Hessian matrix  $H_x = V\Lambda V^T$ , where  $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_r]$  and  $V = [v_1, v_2, \dots, v_n]$ .  $\epsilon$  is the constraint of the  $\ell_p$  norm of the adversarial perturbation, i.e.,  $\|\delta\|_p \leq \epsilon$ . Then we have an upper bound of  $\|\Delta_g\|$ , given by:

$$\|\Delta_g\| \leq \sum_i |(e^{\epsilon \lambda_i} - 1)| \cdot |g_x^T v_i| \cdot \|v_i\| \quad (4)$$

Proposition 2 is proven in Appendix B, available online, which establishes an important insight that although  $\delta$  has already produced a significant change in classification results, it is noteworthy that the upper-bound constraint on gradient changes in the saliency map can constrain its ability to capture the most class-discriminative information within images. Consequently, the effectiveness of the gradient-based saliency map in revealing critical image information is limited under such constraints. To address this limitation, we propose our CDIEA to identify the most class-discriminative information in images.

### B. Class-Discriminative Information Estimation Algorithm (CDIEA)

As can be seen in Fig. 2, CDIEA is designed to automatically estimate the critical points within images, along with their corresponding importance scores. CDIEA is motivated by the realm of sparse adversarial attack research, which suggests that adding imperceptible perturbations to images can cause misclassification. Therefore, these pixels are also vulnerable to minor perturbations and similarly vulnerable to perturbations caused by data augmentation techniques. The importance scores indicate the extent to which these points influence classification decisions.

Specifically, we first formulate the optimization goal of CDIEA as follows:

$$\begin{aligned} \min_{\delta} \quad & \|\delta\|_0 \\ \text{s.t.} \quad & \delta \in [-\epsilon, \epsilon]^{w \times h \times c} \\ & \arg \max_{i=1, \dots, K} f_i(x + \delta) \neq y_{true} \end{aligned} \quad (5)$$

where  $\delta$  is the adversarial perturbation,  $\epsilon$  is the maximum perturbation magnitude,  $f_i$  is a classification model and classifies  $x + \delta$  as class  $i$ .  $x$  and  $y_{true}$  are the original image and its actual label, respectively. The classification-sensitive information is typically distributed throughout the whole image [36], [37], extending beyond single sub-regions to encompass both foreground and background regions. Therefore, we firmly limit the magnitude of perturbation of  $\epsilon$ , to identify more critical points instead of just a few highly sensitive ones subject to significant disturbance.

To precisely obtain the positions of critical points, we encode the perturbation  $\delta$  through a neural network  $\mathcal{F}: \mathbb{R}^{w \times h \times c} \rightarrow \mathbb{R}^{w \times h \times 1}$ .  $\mathcal{F}(\delta)$  is of the same size as the image, with each element indicating the presence of a critical pixel. Notably, the elements within the encoded tensor  $\mathcal{F}(\delta)$  are inherently real numbers. To ensure continuity and differentiability, we adopt the scaled sigmoid function [38] to generate an approximate binary matrix  $\mathbf{m} = \text{sigmoid}(\alpha \cdot \mathcal{F}(\delta))$ . The scaling factor  $\alpha$  controls the level of binarization. For sufficiently large  $\alpha$ ,  $\mathbf{m}$  tends to approximate a binary form, and the approximate ones in  $\mathbf{m}$  indicate the position of critical points. During training, following the suggestions in [38], we gradually increase  $\alpha$  from  $\alpha_{init}$  to  $\alpha_{end}$  to ensure accurate critical point estimation. The degree of binarization of  $\mathbf{m}$  is measured by:

$$\mathcal{B}(\mathbf{m}) = \frac{1}{N} \left( \sum_{i=1}^N \mathbb{I}(\mathbf{m}_i < 0.01) + \sum_{i=1}^N \mathbb{I}(\mathbf{m}_i > 0.99) \right) \quad (6)$$

Therefore, when  $\mathbf{m}_i$  falls below 0.01 or exceeds 0.99, it is deemed that  $\mathbf{m}_i$  has effectively converged to either 0 or 1, respectively. Meanwhile, the sparse perturbation  $\delta \odot \mathbf{m}$  ( $\odot$  denotes element-wise multiplication) encloses the class-discriminative information, including both positions and varying degrees of importance. A smaller value within the perturbation (denoted as  $\delta$ ) indicates that a lower level of perturbation is sufficient to achieve a significant impact. Therefore, the corresponding pixel is more vulnerable to influence and thus highlights its increased importance. Based on the aforementioned, we formulate our loss function as follows:

$$\mathcal{L} = \mathcal{L}_{\text{CE}}(f(\mathbf{x} + \delta \odot \mathbf{m}), y^*) + \mu \frac{\|\mathbf{m}\|_1}{N} \quad (7)$$

where  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm of the vector,  $\mathcal{L}_{\text{CE}}$  is the cross-entropy loss, and  $y^* \neq y_{\text{true}}$ . Since  $\mathbf{m}$  is approximately binarized, the second term approximately represents the sparsity of  $\mathbf{m}$ .  $\mu$  is a dynamic parameter to balance the two terms, which is calculated by:

$$\mu = \frac{\nu}{N} \sum_{i=1}^N \mathbb{I}(\mathbf{m}_i > 0.5) \quad (8)$$

where  $\mathbf{m}_i$  is the  $i$ -th element of the mask  $\mathbf{m}$ ,  $\nu$  is a parameter to adjust the numeric disparity between loss items. Considering the constrain of  $\delta$  in (5), we utilize dense attack strategy [39], [40] to quickly update  $\delta$  under the  $\ell_\infty$ -norm constrain. In this way, we can constrain both the  $\ell_0$ -norm and  $\ell_\infty$ -norm of the perturbation. Specifically, the update formula of  $\delta$  is as follows:

$$\begin{aligned} \mathbf{g}_{t+1} &= \sigma \cdot \mathbf{g}_t + \frac{\nabla_{\mathbf{x}} \mathcal{L}}{\|\nabla_{\mathbf{x}} \mathcal{L}\|_1} \\ \delta_{t+1} &= \text{Clip}_\epsilon \{ \delta_t - \beta \cdot \text{sign}(\mathbf{g}_{t+1}) \} \end{aligned} \quad (9)$$

where  $\mathcal{L}$  is the loss defined by (7),  $\sigma$  denotes the momentum decay factor,  $\beta$  represents the update step, and  $\text{Clip}_\epsilon\{\cdot\}$  is used to project adversarial perturbation into the  $\ell_\infty$ -ball of radius  $\epsilon$ . Both  $\mathbf{g}_0$  and  $\delta_0$  are initialized to zero. Upon completing the iterative process,  $\mathbf{m}$  will be strictly binarized with a threshold of 0.5 to indicate the positions of the critical points explicitly. The general workflow of CDIEA is outlined in Algorithm 1.

---

**Algorithm 1:** The General Workflow of CDIEA.
 

---

**Input:** an image  $\mathbf{x}$ , maximum number of iterations  $T$ , classification model  $f$ , maximum perturbation magnitude  $\epsilon$ , hyperparameter  $\nu$ , scaling factor  $\alpha_{init}$  and  $\alpha_{end}$ , update step  $\beta$

**Output:** adversarial perturbation  $\delta$

- 1:  $\delta_0 \leftarrow 0$
  - 2:  $\mathbf{g}_0 \leftarrow 0$
  - 3:  $\alpha_0 \leftarrow \alpha_{init}$
  - 4:  $\alpha_{step} \leftarrow (\alpha_{end} - \alpha_{init})/T$
  - 5: Randomly initialize the encoder  $\mathcal{F}_0$
  - 6: **for**  $t = 0:T - 1$  **do**
  - 7:    $\mathbf{m}_t \leftarrow \text{sigmoid}(\alpha_t \cdot \mathcal{F}_t(\delta_t))$
  - 8:   Calculate the dynamic parameter  $\mu_t$  based on  $\mathbf{m}_t$  and (8)
  - 9:   Calculate the loss  $\mathcal{L}$  according to  $\mathbf{m}_t$ ,  $\delta_t$ ,  $\mu_t$  and (7)
  - 10:   Update  $\mathbf{g}_{t+1}$  and  $\delta_{t+1}$  according to (9)
  - 11:   Update the encoder  $\mathcal{F}_{t+1}$  based on the loss  $\mathcal{L}$  and SGD with momentum
  - 12:   **if**  $t/T > 0.9$  and  $\mathcal{B}(\mathbf{m}) < 0.99$  **then**
  - 13:      $\alpha_{t+1} \leftarrow \alpha_t + 10 \cdot \alpha_{step}$
  - 14:   **else**
  - 15:      $\alpha_{t+1} \leftarrow \alpha_t + \alpha_{step}$
  - 16:   **end if**
  - 17: **end for**
  - 18:  $\mathbf{m}_T \leftarrow \text{sigmoid}(\alpha_T \cdot \mathcal{F}_T(\delta_T))$
  - 19:  $\delta \leftarrow \delta_T \odot \mathbf{m}_T$
  - 20: **return**  $\delta$
- 

In particular, to minimize the computational cost induced by CDIEA and enhance the practicability of our framework, the generation of class-discriminative information is conducted offline. Once this information is obtained for each dataset, the downstream model training utilizing IPF-RDA will no longer incorporate any training or inference from CDIEA. This preserves the effectiveness of the framework with a streamlined and efficient implementation.

### C. Information-Preserving Framework (IPF)

Within this section, we describe the information-preserving framework and its integration with diverse data augmentation methods, as classified within the taxonomy depicted in Fig. 1. It is known that the efficacy of augmentation techniques benefits from providing augmented samples exhibiting greater diversity [41]. Guided by this notion, the core idea of our framework is to strike a balance between diversity enhancement and the preservation of class-discriminative information, which is a slight tradeoff. Based on (5), smaller perturbation values within  $\delta$  align with heightened sensitivity, where even minor perturbations significantly influence the pixels. Meanwhile, as suggested in previous works [9], [10], [17], which underscore the effectiveness of retaining structural regions, our framework focuses on preserving a structurally informative sub-region during each iteration. Consequently, the importance map  $\eta$  of the critical pixels can be calculated as the reciprocal of the

perturbation, i.e.,

$$\eta_{ij} = \begin{cases} 1/(\delta_{ij} * m_{ij}), & \text{if } m_{ij} \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where  $i = 1, 2, \dots, w$  and  $j = 1, 2, \dots, h$ . The sensitivity of individual pixels is encapsulated in the non-zero entries of  $\eta$ , serving as a direct indicator of their significance. Therefore, our framework introduces an importance score  $\mathcal{IS}$  for rectangular regions  $S$ , which is defined as:

$$\mathcal{IS}(\eta, S) = \sum_{(i,j) \in S} \eta_{ij}. \quad (11)$$

If  $\mathcal{IS}(\eta, S)$  attains a larger value,  $S$  comprises more critical points with high importance scores. These regions, thus, are deliberately safeguarded against any data augmentation-induced alterations. Accurate and efficient assessments of the importance of arbitrary regions, represented by  $S$ , are crucial. Addressing this, we undertake a statistical aggregation of  $\mathcal{IS}$  values pertaining to all potential regions  $S$  of a certain length  $l$  offline. The  $\tau$ -th percentile, denoted as  $Q_{\tau|\mathcal{IS}}$ , serves as the discerning threshold. This threshold signifies that a region  $S$  exhibiting an  $\mathcal{IS}$  value higher than  $Q_{\tau|\mathcal{IS}}$  is more likely to contain critical information and should be carefully preserved. As a result, the additional computational overhead introduced by our framework during training is solely to sample a region  $S$  with an appropriate importance score. This additional computational overhead is trivial when compared to the predominant costs intrinsic to model training efforts, indicating the high efficiency of our framework.

Based on the taxonomy in Fig. 1, our proposed framework carefully preserves critical information to enhance the robustness of different data augmentation types.

*a) Regional Deletion:* We prevent regional deletion-based data augmentation methods from inadvertently removing critical regions with high  $\mathcal{IS}$ . To achieve this, we enforce a constraint on the importance score  $\mathcal{IS}$  of randomly removed regions; the  $\mathcal{IS}$  is rigorously below the threshold, denoted as  $Q_{\tau|\mathcal{IS}}$ . For an image and the corresponding class-discriminative information  $\eta$ , our framework ensures that the randomly removed region  $S$  satisfies:  $\mathcal{IS}(\eta, S) \leq Q_{\tau|\mathcal{IS}}$ . In this way, critical information, such as objects of interest, remains almost unaffected during the augmentation process. Our framework achieves enhanced robustness for regional deletion-based DA methods and concurrently maintains a diversified set of augmented data. Striking this balance is of paramount importance, as it facilitates the protection of critical information while preserving the essential variability necessary for comprehensive and effective data augmentation strategies.

*b) Image-level Transformation:* In the context of image-level augmentation, our framework aims to recover a random critical sub-region  $S$  with  $\mathcal{IS}$  higher than  $Q_{\tau|\mathcal{IS}}$  after the augmentation process. Meanwhile, the remainder of the image regions undergoes comprehensive augmentation via the employed data augmentation techniques.

The adoption of the proposed framework achieves dual objectives. Primarily, it guarantees the diversity of the augmented

---

**Algorithm 2:** The Pseudocode of IPF for Robust Data Augmentation.

---

**Input:** A batch of images  $\mathbf{x}$ , importance map  $\eta$ ,  $\tau$ -th percentile  $Q_{\tau|\eta}$ , batch size  $N$ , length  $l$ , DA method  $\mathcal{A}$

- 1: **for**  $i = 0:N$  **do**
- 2:    **(a) Regional deletion**
- 3:    **repeat**
- 4:      Sample a rectangle region  $S$  of length  $l$  in  $\eta_i$
- 5:      Compute  $\mathcal{IS}$  of  $S$  according to (11)
- 6:    **until**  $\mathcal{IS}(\eta_i, S) \leq Q_{\tau|\mathcal{IS}}$
- 7:    Erase region  $S$  w.r.t. DA methods
- 8:    **(b) Image-level transformation**
- 9:    **repeat**
- 10:      Sample a rectangle region  $S$  of length  $l$  in  $\eta_i$
- 11:      Compute  $\mathcal{IS}$  of  $S$  according to (11)
- 12:    **until**  $\mathcal{IS}(\eta_i, S) \geq Q_{\tau|\mathcal{IS}}$
- 13:    Recover region  $S$  in the augmented image  $\mathcal{A}(x_i)$
- 14:    **(c) Image mixing**
- 15:    Permute the current batch of images as  $\mathbf{x}'$
- 16:    **repeat**
- 17:      Sample a rectangle region  $S$  of length  $l$  in  $\eta_i$
- 18:      Compute  $\mathcal{IS}$  of  $S$  according to (11)
- 19:    **until**  $\mathcal{IS}(\eta_i, S) \leq Q_{\tau|\mathcal{IS}}$
- 20:    Replace the region  $S$  with a random region  $S'$  in  $\mathbf{x}'$  according to (1)
- 21:    Compute weight  $\lambda$  to fuse labels according to (12)
- 21: **end for**

**Output:** A batch of augmented images

---

data, which is crucial for practical training. Simultaneously, the framework preserves classification-sensitive insights from perturbation during augmentation, thus bolstering the inherent robustness of the data augmentation approaches.

*c) Image Mixing:* Our framework integrates region-level mixing data augmentation approaches through a two-fold approach to maintain the simplicity and efficiency of the original method pipeline. To this end, we first prioritize the preservation of label information by substituting a less critical region in the first image  $x_a$ , thus preserving the information of label  $y_a$  to the greatest extent possible. Concurrently, we improve the calculation of label coefficients in (1) based on the ratio of the importance scores of the replaced regions  $\mathbf{M}_b$  from the second image  $x_b$  within the composite image  $\tilde{x}$  as follows:

$$\omega = \frac{\mathcal{IS}(\mathbf{M}_b)}{\mathcal{IS}(\tilde{x})}. \quad (12)$$

This modified strategy yields a logical advantage. Specifically, when the chosen region from  $x_b$  does not encompass any object associated with  $y_b$ , the  $\mathcal{IS}(\mathbf{M}_b)$  becomes very small and negligible. Consequently, the corresponding weight  $\omega$  is substantially reduced, leading the fused label  $\tilde{y}$  to be predominantly  $y_a$ . Therefore, this strategy yields a more coherent and rational way to minimize potential noise stemming from the data augmentation operations. Finally, we present the pseudocode of IPF in Algorithm 2.

#### IV. EXPERIMENTS

This section empirically validates the proposed IPF-RDA across various tasks and investigates its behaviors through ablation studies conducted on seven widely used benchmark datasets. We incorporate 10 commonly used SOTA DA methods into our framework to further enhance their robustness and practical performance. Specifically, our evaluation encompasses a diverse range of tasks, commencing with supervised image classification on widely used benchmark datasets, including CIFAR-10/100 [42] and Tiny-ImageNet [22]. Subsequently, the focus shifts to assessing the effectiveness of IPF-RDA within the context of person re-identification tasks to demonstrate the scalability of IPF-RDA. This experiment is conducted on the CUHK03 [23] and Market1501 [24] datasets. Furthermore, we extend the application of IPF-RDA to semi-supervised learning algorithms [43], thereby underscoring its versatility across multiple learning paradigms. Additionally, a thorough comparison of the effectiveness of IPF-RDA and KeepAugment [9] is conducted.

Besides, to gain deeper insights into our framework, we have meticulously orchestrated a series of analytical experiments, aiming at elucidating the functioning and efficacy of the proposed methodology. First, to demonstrate the effectiveness of our approach, we present visualizations showcasing the results of augmented data by various methods. Meanwhile, to demonstrate the performance improvements achieved by our framework (e.g., models with enhanced feature extraction capabilities), we offer a set of t-SNE visualizations [44] on the MNIST dataset [45]. These visualizations serve as a demonstrative testament to the performance enhancements instilled by our framework. Second, to validate the capability of IPF-RDA in promoting models to learn better representations, we visualize the class activation maps (CAM) [46]. The CAM is obtained by models trained with IPF-RDA on the Oxford Flower dataset [47]. We then employ transfer learning by employing the models trained by IPF-RDA as backbones on CIFAR-100, followed by fine-tuning for CIFAR-10. Third, empirical results are provided to evaluate the magnitude of the gradient change in (22) relative to the original gradient. Next, we evaluate the effectiveness of CDIEA in capturing the class-discriminative information. Moreover, we conduct experiments related to additional training costs to verify the efficiency of our approach. Lastly, we perform ablation studies, a systematic exploration to investigate the effect of different hyperparameters on the performance of IPF-RDA. We report the average results over three independent runs with different random seeds.

##### A. Experimental Setups for Image Classification

We incorporate ten commonly used DA methods into our framework, including Cutout [4], CutMix [5], RandomErasing [19], HaS [18], GridMask [17], AdvMask [10], AutoAugment [6], Fast-AutoAugment [20], RandAugment [7], and TrivialAugment [8]. In the context of supervised learning, we closely follow previous works [4], [6], [9], [10] with our setup. Specifically, all images are preprocessed by dividing each pixel value by 255 and normalizing by the dataset-specific statistics. For the CIFAR dataset, we pad images to  $36 \times 36$ , randomly crop

them into  $32 \times 32$ , and horizontally flip them with a probability of 50%. We train 1800 epochs with cosine learning rate decay for Shake-Shake [50] and Pyramid-Shakedrop [51] using SGD with Nesterov Momentum and a learning rate of 0.1, a batch size of 256,  $1e^{-3}$  weight decay, and cosine learning rate decay. We conduct training for 300 epochs for all other deep networks, utilizing SGD with Nesterov Momentum and a learning rate of 0.1, a batch size of 128, a  $5e^{-4}$  weight decay, and cosine learning rate decay. Regarding the Tiny-ImageNet dataset, we resize the images into  $64 \times 64$ , initialize models with ImageNet pre-trained weights, and then fine-tune models using various augmentations. The baseline approach involves solely padding and horizontal flipping. To facilitate fair comparisons, all methods are implemented with the same training configurations.

*Parameter Settings:* Following the suggestions in [38], we gradually increase  $\alpha$  from  $\alpha_{init}$  to  $\alpha_{end}$  during training. This adjustment serves a dual purpose: it facilitates the convergence of elements within  $m$  toward binary values while circumventing the algorithm's degeneration into random pixel selection.  $\alpha_{end}$  is set to 100 on CIFAR and 10 on Tiny-ImageNet and others. The change of  $\alpha_{init}$  has little effect on the results, so it is set to 0.1.  $\gamma$  in (12) is typically set as 10 on CIFAR and 100 on other datasets. For IPF, we adopt the common settings from [4], [9] with parameters  $\tau$  and  $l$  settings to minimize parameter tuning overhead. Specifically,  $l$  is set to 16 on CIFAR-10 and Tiny-ImageNet, and 8 on CIFAR-100;  $\tau$  is consistently fixed as 0.6 across datasets. The design of our framework ensures that these parameters are agnostic to the choice of network architectures and augmentation techniques. Therefore, the parameters of our framework can be easily set, thus facilitating the seamless adoption and implementation of our approach.

##### B. Supervised Image Classification

*1) Results on CIFAR-10 and CIFAR-100:* We conduct a comprehensive evaluation of our framework on the CIFAR-10 and CIFAR-100 datasets. For evaluation, various popular deep networks are utilized, including ResNet-18/50/110 [48], Wide-ResNet-28-10 (WRN-28-10) [49], ShakeShake2-32/96 [50] and Pyramid-Shakedrop [51] whose hyper-parameters are precisely the same as those used in [4], [8], [20]. The error rates on CIFAR-10/100 test sets are shown in Table I.

It can be observed that the original data augmentation techniques contribute to the improvements in model performance. However, our proposed framework further enhances the efficacy of these SOTA data augmentation approaches. Specifically, IPF-RDA consistently yields substantial improvements across nearly all data augmentation methods. Particularly, on both CIFAR-10 and CIFAR-100, our method can achieve 0.1-1.0% and 0.1-2.7% error rate reductions, respectively. For instance, after integrating Cutout into our framework, IPF-Cutout reduces the test error rates of the original Cutout using ResNet-50 on CIFAR-10 by 0.92 percent. On CIFAR-100, for relatively small models like ResNet-18 and ResNet-50, IPF-RDA improves data augmentation approaches by about 0.8 percent. In contrast, for larger models like ResNet-110, Wide-ResNet-28-10, and Shakeshake models, our framework can improve most methods by over 1 percent.

TABLE I  
ERROR RATES (%) ON CIFAR-10 AND CIFAR-100 ARE SUMMARIZED IN THIS TABLE ACROSS DEEP MODELS. \* MEANS THE PUBLISHED RESULTS IN PREVIOUS PAPERS. THE BETTER RESULTS ARE *BOLD-FACED*. FAA: FAST-AUTOAUGMENT.

Dataset	Method	ResNet18 [48]	ResNet50 [48]	ResNet-110 [48]	WRN-28-10 [49]	Shake-2-32 [50]	PyramidNet [51]
CIFAR-10 [21]	baseline	4.72*	4.34	4.55	4.48	5.10	2.80
	Cutout [4]	3.99*	4.19*	3.56	3.08*	3.04*	1.80
	IPF-Cutout	<b>3.22</b> ↓0.77	<b>3.27</b> ↓0.92	<b>3.36</b> ↓0.20	<b>2.70</b> ↓0.38	<b>2.73</b> ↓0.31	<b>1.68</b> ↓0.12
	CutMix [5]	3.36*	3.19*	3.51	3.07*	3.53	1.72
	IPF-CutMix	<b>3.18</b> ↓0.18	<b>2.84</b> ↓0.35	<b>3.24</b> ↓0.27	<b>2.45</b> ↓0.63	<b>2.93</b> ↓0.60	<b>1.53</b> ↓0.19
	AutoAugment [6]	3.49*	3.41	3.34	2.99	2.70	1.70
	IPF-AutoAugment	<b>2.98</b> ↓0.51	<b>3.13</b> ↓0.28	<b>3.03</b> ↓0.31	<b>2.57</b> ↓0.42	<b>2.37</b> ↓0.33	<b>1.31</b> ↓0.39
	RandAugment [7]	3.53	3.75	4.03	3.06*	2.95	1.50
	IPF-RandAugment	<b>3.21</b> ↓0.32	<b>2.82</b> ↓0.93	<b>3.29</b> ↓0.74	<b>2.48</b> ↓0.58	<b>2.57</b> ↓0.38	<b>1.38</b> ↓0.12
	Fast-AutoAugment [20]	4.01	3.31	4.04	2.70*	3.58	1.58
	IPF-FAA	<b>3.01</b> ↓1.00	<b>3.05</b> ↓0.26	<b>3.31</b> ↓0.73	<b>2.51</b> ↓0.19	<b>2.74</b> ↓0.84	<b>1.42</b> ↓0.16
	GridMask [17]	3.62	3.85	4.03	2.77	3.09	1.69
	IPF-GridMask	<b>3.53</b> ↓0.09	<b>3.45</b> ↓0.40	<b>3.56</b> ↓0.47	<b>2.70</b> ↓0.07	<b>2.87</b> ↓0.22	<b>1.59</b> ↓0.10
	Has [18]	3.90	4.40	4.68	3.06	3.11*	2.05
	IPF-Has	<b>3.60</b> ↓0.30	<b>3.28</b> ↓1.12	<b>3.78</b> ↓0.90	<b>2.91</b> ↓0.15	<b>2.68</b> ↓0.43	<b>1.74</b> ↓0.31
	RandomErasing [19]	4.31	4.18	4.61	3.08	3.54*	2.21
	IPF-RandomErasing	<b>3.63</b> ↓0.68	<b>3.33</b> ↓0.85	<b>3.71</b> ↓0.90	<b>2.72</b> ↓0.36	<b>2.81</b> ↓0.73	<b>1.77</b> ↓0.44
	AdvMask [10]	3.56*	3.31*	3.60	2.98*	2.97*	1.73
IPF-AdvMask	<b>3.50</b> ↓0.06	<b>3.11</b> ↓0.20	<b>3.28</b> ↓0.32	<b>2.64</b> ↓0.34	<b>2.79</b> ↓0.18	<b>1.52</b> ↓0.21	
TrivialAugment [8]	3.23	2.87	3.13	2.29	2.70	1.42	
IPF-TrivialAugment	<b>3.18</b> ↓0.05	<b>2.85</b> ↓0.02	<b>3.04</b> ↓0.09	<b>2.23</b> ↓0.06	<b>2.63</b> ↓0.07	<b>1.34</b> ↓0.08	
CIFAR-100 [21]		ResNet-18 [48]	ResNet-50 [48]	ResNet-110 [48]	WRN-28-10 [49]	Shake-2-32 [50]	Shake-2-96 [50]
	baseline	22.46*	22.59	23.83	21.04	23.35	18.34
	Cutout [4]	21.96*	21.38	21.62	18.40*	20.63	16.54
	IPF-Cutout	<b>21.47</b> ↓0.49	<b>18.85</b> ↓2.53	<b>21.40</b> ↓0.22	<b>17.36</b> ↓1.04	<b>18.71</b> ↓1.92	<b>16.00</b> ↓0.54
	CutMix [5]	19.55	18.66	19.32	17.33	20.43	15.29
	IPF-CutMix	<b>19.08</b> ↓0.47	<b>17.37</b> ↓1.29	<b>19.20</b> ↓0.12	<b>16.10</b> ↓1.23	<b>19.80</b> ↓0.63	<b>15.03</b> ↓0.26
	AutoAugment [6]	20.22	17.92	20.37	17.10*	17.71	14.28
	IPF-AutoAugment	<b>19.62</b> ↓0.60	<b>17.50</b> ↓0.42	<b>18.11</b> ↓2.26	<b>16.24</b> ↓0.86	<b>17.38</b> ↓0.33	<b>14.08</b> ↓0.20
	RandAugment [7]	21.70	19.05	22.32	17.10*	20.00	17.02
	IPF-RandAugment	<b>20.94</b> ↓0.76	<b>18.10</b> ↓0.95	<b>20.95</b> ↓1.37	<b>16.38</b> ↓0.72	<b>19.26</b> ↓0.74	<b>16.60</b> ↓0.42
	Fast-AutoAugment [20]	20.89	20.92	23.87	17.30*	18.61	15.41
	IPF-FAA	<b>20.00</b> ↓0.89	<b>17.94</b> ↓2.98	<b>18.97</b> ↓4.90	<b>16.55</b> ↓0.75	<b>17.12</b> ↓1.49	<b>14.60</b> ↓0.81
	GridMask [17]	24.77	21.62	21.27	19.60	20.57	17.45
	IPF-GridMask	<b>23.64</b> ↓1.13	<b>20.13</b> ↓1.49	<b>21.20</b> ↓0.07	<b>17.53</b> ↓2.07	<b>19.20</b> ↓1.37	<b>17.30</b> ↓0.15
	Has [18]	21.81	21.24	22.11	19.78	21.03	15.92
	IPF-Has	<b>21.63</b> ↓0.18	<b>18.63</b> ↓2.61	<b>20.85</b> ↓1.26	<b>17.57</b> ↓2.21	<b>19.76</b> ↓1.27	<b>15.84</b> ↓0.08
	RandomErasing [19]	24.03	22.21	22.85	19.43	21.19	18.36
	IPF-RandomErasing	<b>21.61</b> ↓2.42	<b>20.89</b> ↓1.32	<b>21.34</b> ↓1.51	<b>17.52</b> ↓1.91	<b>20.03</b> ↓1.16	<b>17.10</b> ↓1.26
AdvMask [10]	21.57*	21.01*	21.20	19.30*	20.04*	16.94	
IPF-AdvMask	<b>21.42</b> ↓0.15	<b>19.19</b> ↓1.82	<b>20.93</b> ↓0.27	<b>18.60</b> ↓0.70	<b>19.01</b> ↓1.03	<b>16.05</b> ↓0.89	
TrivialAugment [8]	21.33	18.67	20.41	16.31	17.86	15.70	
IPF-TrivialAugment	<b>20.67</b> ↓0.66	<b>18.35</b> ↓0.32	<b>18.99</b> ↓1.42	<b>16.20</b> ↓0.11	<b>17.52</b> ↓0.34	<b>15.29</b> ↓0.41	

Meanwhile, our framework's effectiveness is particularly pronounced for information deletion-based data augmentation methods, such as Cutout, as well as for automatic augmentations, such as AutoAugment. These methodologies, prone to introducing perturbations in augmented images, significantly benefit from the mitigating influence of IPF-RDA, manifesting substantial gains in overall performance. A plausible explanation for the superior performance of our method is that our method suppresses noises caused by data augmentation and helps models to learn better generalization abilities.

2) *Results on Tiny-ImageNet*: The comparative evaluation of augmentation methods, before and after their integration into our proposed framework, is presented in Table II using the Tiny-ImageNet dataset. For a comprehensive assessment, we employ a spectrum of popular deep network architectures, encompassing ResNet-18/50/110 [48], Wide-ResNet-50-2 [49], DenseNet121/201 [52], and ResNext50/101 [53].

It can be observed that the original data augmentation approaches contribute to performance improvements for deep

models compared to the baseline model. More importantly, the performance of IPF-RDA becomes unequivocally apparent, as it substantially elevates the overall generalization capabilities of these data augmentation strategies, effectively enhancing their robustness and unleashing their full potential during the model training process. For instance, the error rates of ResNet-50, as witnessed in conjunction with the original augmentation methods, experience a reduction exceeding 1% following the application of IPF-RDA. Meanwhile, the extent of improvement can be discernibly influenced by the choice of architecture. Specifically, compared to ResNets and Wide-ResNets, DenseNets generally suffer less from overfitting due to the architecture design and thus appear to benefit less from data augmentation techniques.

### C. Person Re-Identification

Data augmentation is also widely used in person re-identification tasks. To demonstrate the scalability of IPF-RDA, we apply the proposed algorithm in the context of the ReID task,

TABLE II  
ERROR RATES (%) OF DIFFERENT DEEP NETWORKS ON THE VALIDATION SET OF TINY-IMAGE NET. \* MEANS THE PUBLISHED RESULTS IN PREVIOUS PAPERS. THE BETTER RESULTS ARE **BOLD-FACED**. FAA: FAST-AUTOAUGMENT.

Method	ResNet-18 [48]	ResNet-50 [48]	WRN-50-2 [49]	DenseNet121 [52]	DenseNet201 [52]	ResNext50 [53]	ResNext101 [53]
baseline	38.62	26.39	18.45	37.99	31.08	20.24	14.94
Cutout [4]	31.73	22.55	17.73	33.54	30.68	17.84	14.47
IPF-Cutout	<b>28.94</b> ↓2.79	<b>20.97</b> ↓1.58	<b>16.75</b> ↓0.98	<b>33.24</b> ↓0.30	<b>30.60</b> ↓0.08	<b>17.45</b> ↓0.39	<b>14.25</b> ↓0.22
CutMix [5]	35.91	23.59	17.68	37.40	33.70	17.79	14.36
IPF-CutMix	<b>30.51</b> ↓5.40	<b>20.88</b> ↓2.71	<b>16.96</b> ↓0.72	<b>33.46</b> ↓3.94	<b>31.15</b> ↓2.55	<b>17.38</b> ↓0.41	<b>14.05</b> ↓0.31
RandomErasing [19]	36.00	24.67	18.11	34.38	31.68	18.48	14.27
IPF-RandomErasing	<b>28.58</b> ↓7.42	<b>20.89</b> ↓3.78	<b>16.73</b> ↓1.38	<b>33.38</b> ↓1.00	<b>31.01</b> ↓0.67	<b>17.43</b> ↓1.05	<b>14.07</b> ↓0.20
TrivialAugment [8]	29.98	24.40	16.54	38.86	31.10	19.09	14.23
IPF-TrivialAugment	<b>27.66</b> ↓2.32	<b>19.86</b> ↓4.54	<b>16.10</b> ↓0.44	<b>33.91</b> ↓4.95	<b>30.43</b> ↓0.67	<b>18.79</b> ↓0.30	<b>13.61</b> ↓0.62
AutoAugment [6]	31.72	29.75	17.48	34.76	32.83	18.72	15.06
IPF-AutoAugment	<b>30.54</b> ↓1.18	<b>23.42</b> ↓6.33	<b>17.01</b> ↓0.47	<b>34.43</b> ↓0.33	<b>32.29</b> ↓0.54	<b>18.41</b> ↓0.31	<b>14.40</b> ↓0.66
Fast-AutoAugment [20]	32.26	21.06	17.16	33.65	31.70	18.96	14.78
IPF-FAA	<b>29.83</b> ↓2.43	<b>20.31</b> ↓0.75	<b>16.98</b> ↓0.18	<b>33.26</b> ↓0.39	<b>31.57</b> ↓0.13	<b>17.11</b> ↓1.85	<b>14.34</b> ↓0.44
GridMask [17]	37.28*	22.12	17.75	33.99	31.36	18.95	14.12
IPF-GridMask	<b>30.78</b> ↓6.50	<b>21.52</b> ↓0.60	<b>16.49</b> ↓1.26	<b>33.51</b> ↓0.48	<b>31.24</b> ↓0.12	<b>17.45</b> ↓1.50	<b>13.64</b> ↓0.48
HaS [18]	36.49*	24.68	18.23	34.26	31.74	19.48	13.74
IPF-HaS	<b>36.27</b> ↓0.22	<b>23.89</b> ↓0.79	<b>17.81</b> ↓0.42	<b>33.90</b> ↓0.36	<b>31.60</b> ↓0.14	<b>18.56</b> ↓0.92	<b>13.26</b> ↓0.48
AdvMask [10]	34.71*	20.16	16.61*	33.84	30.92	17.91	14.17
IPF-AdvMask	<b>33.09</b> ↓1.62	<b>19.93</b> ↓0.23	<b>15.99</b> ↓0.62	<b>33.69</b> ↓0.15	<b>30.57</b> ↓0.35	<b>17.60</b> ↓0.31	<b>13.74</b> ↓0.43
RandAugment [7]	34.33	24.13	17.75	36.05	32.25	19.64	14.80
IPF-RandAugment	<b>34.23</b> ↓0.10	<b>23.82</b> ↓0.31	<b>17.63</b> ↓0.12	<b>35.59</b> ↓0.46	<b>31.99</b> ↓0.26	<b>19.57</b> ↓0.07	<b>14.73</b> ↓0.07

TABLE III  
DETECTION MEAN AVERAGE PRECISION (mAP) RESULTS ON CUHK03 [23] AND MARKET1501 [24] DATASETS. MAP (%) IS REPORTED FOR COMPARISON

Method	CUHK03 [23]	Market1501 [24]
Cutout [4]	81.9	89.8
IPF-Cutout	<b>82.1</b> ↑0.2	<b>90.1</b> ↑0.3
HaS [18]	77.1	86.9
IPF-HaS	<b>77.5</b> ↑0.4	<b>88.4</b> ↑1.5
RE [19]	81.6	90.6
IPF-RE	<b>82.3</b> ↑0.7	<b>90.8</b> ↑0.2

specifically employing the CUHK03 [23] and Market1501 [24] datasets.

Since Cutout, HaS, and Random Erasing are the most commonly used data augmentation methods in ReID, we incorporate these methods into our framework. We adopt the feature pyramid branch model [54], using the standard hyperparameters from the codebase.<sup>1</sup> We fix nearly all the training setups but utilize various data augmentation approaches to assess their performance during training. The mean average precision (mAP) scores across three independent trials are shown in Table III. Evidently, it can be observed that IPF-RDA further improves the generalization performance of these methods significantly. Specifically, when IPF-RDA is applied in conjunction with HaS, our framework outperforms the original HaS by 1.5% and 0.4% on Market1501 and CUHK03, respectively. The reason behind the superior performance of our method lies in its ability to effectively suppress the complete or substantial removal of critical objects in images when applied alongside these methods, all while preserving data diversity. Therefore, models can be trained more effectively. This emphasizes the efficacy and scalability of IPF-RDA in bolstering the performance of data augmentation strategies within the realm of ReID tasks.

<sup>1</sup><https://github.com/anocodetest1/FPB/>

TABLE IV  
SEMI-SUPERVISED LEARNING TEST ACCURACY ON CIFAR-10

Method	Acc (%)	Prec (%)
Cutout [4]	91.54	91.01
IPF-Cutout	<b>91.71</b> ↑0.17	<b>91.70</b> ↑0.69
HaS [18]	86.96	86.18
IPF-HaS	<b>91.32</b> ↑4.36	<b>89.59</b> ↑3.41
RandomErasing [19]	91.10	90.79
IPF-RE	<b>92.22</b> ↑1.12	<b>92.17</b> ↑1.38
AutoAugment [6]	92.97	92.91
IPF-AA	<b>93.66</b> ↑0.69	<b>93.44</b> ↑0.53
RandAugment [7]	91.03	90.89
IPF-RA	<b>93.26</b> ↑2.23	<b>93.03</b> ↑2.14
FAA [20]	91.01	91.11
IPF-FAA	<b>91.07</b> ↑0.06	<b>91.30</b> ↑0.19
TrivialAugment [8]	93.34	93.10
IPF-TA	<b>93.47</b> ↑0.13	<b>93.20</b> ↑0.10

#### D. Semi-Supervised Learning

Data augmentation has been employed as a powerful tool for developing SOTA semi-supervised learning (SSL) algorithms. In light of this, we extend the application of the proposed framework to the domain of unsupervised data augmentation [43] (UDA) on CIFAR-10, aiming to demonstrate the wide-ranging applicability of our method. To facilitate our evaluation, we divide the training set of CIFAR-10 into two subsets: a labeled set comprising 4000 samples and an unlabeled set with others by randomly removing parts of the labels. All training setups closely follow previous work [43], [55] but employ different data augmentations.

The results of both accuracy and precision metrics using Wide-ResNet-28-2 [49] are shown in Table IV. It can be observed that IPF-RDA consistently elevates the performance of diverse augmentations in terms of both accuracy and precision. Particularly noteworthy are the discernible improvements achieved by IPF-RDA, with substantial 4.36%, 1.12%, and

TABLE V  
COMPARISON BETWEEN IPF-RDA AND KEEP AUGMENT BASED ON CUTOUT [4] WITH VARYING CUTOUT LENGTH. WE PRESENT THE ERROR RATES OF THE RESNET-18 MODEL ON THE CIFAR-10 TEST SET. \* MEANS THE PUBLISHED RESULTS IN [9].

Cutout Length	Cutout [4]	KeepCutout [9]	IPF-Cutout
8	4.7*	4.9*	<b>4.33</b>
12	4.6*	4.5*	<b>3.85</b>
16	4.4*	3.9*	<b>3.58</b>
20	4.5*	4.0*	<b>3.53</b>
24	5.1*	4.4*	<b>3.88</b>

TABLE VI  
ADDITIONAL COMPARISON WITH KEEP AUGMENT [9] ON CIFAR-10 AND CIFAR-100, INCLUDING BOTH KEEP CUTOUT AND KEEP AUTO AUGMENT (KEEPAA). \* MEANS THE PUBLISHED RESULTS IN [9].

	ResNet-18	ResNet-110	WRN28-10
CIFAR-10 [42]			
Cutout [4]	4.40*	5.20*	3.10*
KeepCutout [9]	3.80*	4.50*	2.70*
IPF-Cutout	<b>3.14</b>	<b>3.37</b>	<b>2.57</b>
CIFAR-100 [21]			
AutoAugment [6]	20.22	17.10	17.71
KeepAA [9]	19.70	17.10	17.80
IPF-AA	<b>19.62</b>	<b>16.24</b>	<b>17.38</b>

0.69% improvements in test accuracy for HaS, RandomErasing, and AutoAugment, respectively. These findings conclusively underscore the efficacy of IPF-RDA within the realm of SSL tasks.

### E. Comparison With KeepAugment [9]

We compare our proposed framework and KeepAugment. Our primary focus is on their application in conjunction with Cutout [4] and AutoAugment [6] techniques, as KeepAugment primarily brings improvements upon these two methods. To facilitate comparison, we rigorously follow the experimental settings suggested in [9]. As KeepAugment’s source code for reproduction remains unavailable, we undertake its implementation based on an unofficial reproduction,<sup>2</sup> which yields competitive results compared to the published results in the original paper. For CIFAR-10, we directly use the reported results in [9] and follow its training settings for fairness. For CIFAR-100, which was not reported in the original paper, we reproduce the results under our experimental settings.

Table V presents the test accuracy with different  $l$  values. It can be seen that although KeepAugment has brought improvements to the performance of Cutout, our proposed framework consistently outperforms KeepAugment, yielding more significant improvements in the performance of Cutout. Furthermore, Table VI presents the test accuracy on CIFAR-10 and CIFAR-100, employing various deep networks. While both KeepAugment and IPF-RDA exhibit performance improvements concerning Cutout and AutoAugment, IPF-RDA manifests notably higher test accuracy, e.g., achieving 1.13% accuracy improvement on CIFAR-10 using ResNet-110.

<sup>2</sup>[https://github.com/cjf8899/KeepAugment\\_Pytorch](https://github.com/cjf8899/KeepAugment_Pytorch)

TABLE VII  
COMPARISON WITH PUZZLEMIX [28] AND CO-MIXUP [29] ON CIFAR-100 AND TINY-IMAGENET ACROSS MULTIPLE ARCHITECTURES. WE USE THE REPORTED RESULTS FROM [28] (\*). PER-EPOCH TRAINING TIME (S) IS MEASURED ON CIFAR-100 USING PREACTRESNET-18 WITH A 1-NVIDIA-2080-TI-GPU SERVER.

	Tiny-Imagenet	CIFAR-100			Per-epoch costs
	PreActR18	PreActR18	WRN16-8	ResNeXt29-4-24	
PuzzleMix	36.52*	20.62*	19.24*	21.12*	72.61s
Co-Mixup	35.85*	19.87*	19.15*	19.78*	192.37s
Ours	<b>34.33</b>	<b>18.91</b>	<b>18.67</b>	<b>19.23</b>	<b>42.31s</b>

A plausible explanation for the robust performance of IPF-RDA resides in its preservation of the most sensitive augmentation-relevant information within images. This information is the most critical for classification decisions, thereby enhancing the overall robustness of DA methods.

### F. Comparison With Saliency-Guided DA

We compare our proposed framework with PuzzleMix and Co-Mixup on CIFAR-100 and Tiny-ImageNet, closely following the experimental settings and architectures used in [28], [29]. As shown in Table VII, our method consistently achieves better performance across datasets and architectures. In addition to its performance advantage, our method offers better training efficiency due to its lightweight design, incurring minimal extra cost during online training. We also account for the offline cost of the CDIEA module, which requires only about 2.5 GPU hours (Table XI). Notably, this offline cost is a one-time preprocessing step per dataset. As a result, when training PreActResNet-18 for 300 epochs, the total GPU time for our method is approximately 6.03 hours, which is comparable to PuzzleMix (6.05 hours) and markedly lower than Co-Mixup (16.03 hours). By decoupling a lightweight offline stage from efficient online training, our framework achieves a better balance between practicality and scalability.

### G. Analytical Results

1) *Visualization of Augmented Data*: To facilitate a comprehensive comparison of the augmented samples by different methods, we offer a visual analysis as depicted in Fig. 3. It can be observed that the original augmentation methods may impact the critical information crucial for the accurate classification of images, resulting in the introduction of noisy augmented samples during the augmentation process. Conversely, IPF-RDA is capable of effectively preventing critical information from being removed or recovering the damaged critical information while concurrently upholding the diversity intrinsic to the augmented data. Therefore, IPF-RDA can enhance the robustness of data augmentation methods and further unleash their full potential for improved performance and generalization in various applications.

2) *t-SNE Visualization of Deep Features*: To perform a comparative analysis between models employing and not employing IPF-RDA, we visualize the learned deep features derived from the CIFAR-10 dataset. To this end, we leverage the t-SNE algorithm [44], which provides a visual representation to

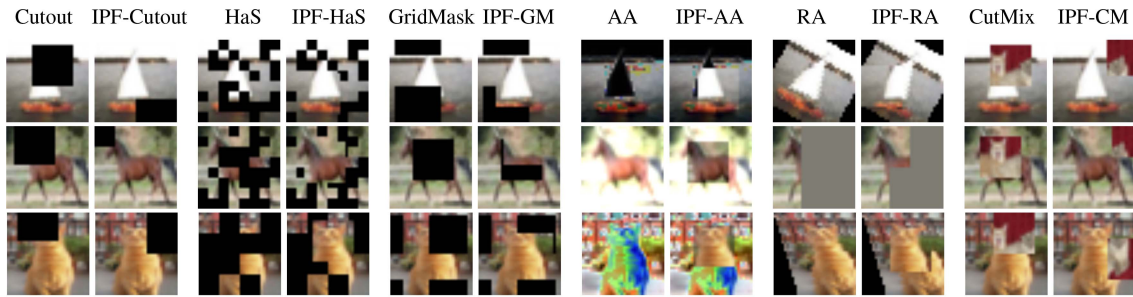
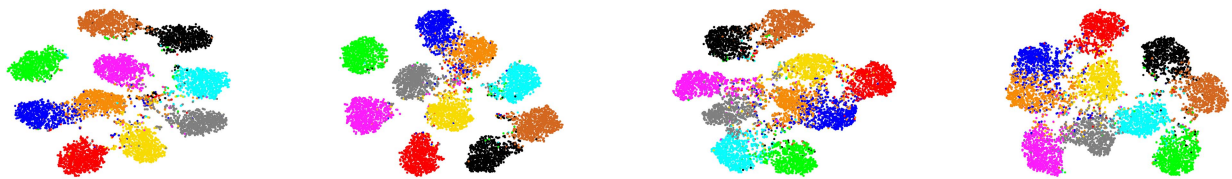


Fig. 3. Visualization of augmented data on CIFAR-10 using different DA approaches.



(a) ResNet-50, FAA,  
Test Acc: 96.06%  
DI:  $2.04 \times 10^{-6}$

(b) ResNet-50, IPF-FAA,  
Test Acc: 96.95%  
DI:  $3.28 \times 10^{-5}$

(c) Transferred ResNet-50,  
FAA, Test Acc: 92.26%  
DI:  $4.43 \times 10^{-5}$

(d) Transferred ResNet-50,  
IPF-FAA, Test Acc: 93.45%  
DI:  $7.58 \times 10^{-5}$

Fig. 4. Visualization of deep features on CIFAR-10 using t-SNE algorithm [44]. Each color denotes a class. DI: Dunn index.

compare the performance of different models. Since automatic augmentation methods typically obtain stronger feature extraction capabilities [7], [8], we utilize a representative method, Fast-AutoAugment (FAA), to train the ResNet-50 models. The results of the R-50 models trained with FAA and IPF-FAA are presented in Fig. 4(a) and (b), respectively.

It can be observed intuitively that, using IPF-RDA, the distribution of different clusters presents superior inter-cluster separation and intra-cluster compactness, effectively enhancing the discriminative capability of learned features. In order to accurately quantify these clustering results, the Dunn index (DI) is utilized, which is a quantitative metric considering both inter-cluster and intra-cluster distance. DI is defined as follows:

$$DI = \frac{\min_{1 \leq i \neq j \leq m} \delta(C_i, C_j)}{\max_{1 \leq j \leq m} \Delta_j} \quad (13)$$

where separation  $\delta(C_i, C_j)$  is the inter-cluster distance metric between clusters  $C_i$  and  $C_j$ , and compactness  $\Delta_j$  calculates the mean distance between all pairs in each cluster. Therefore, a higher DI means better clustering. As is shown in Fig. 4 using IPF-RDA, the clustering results achieve much higher DI values. Specifically, models trained with the original FAA exhibit a DI value of  $2.04 \times 10^{-6}$ , while the DI value with IPF-FAA escalates to  $3.28 \times 10^{-5}$ . Therefore, with the integration of our framework, the DI value increases rapidly, indicating enhanced clustering performance. Meanwhile, the enhanced clustering performance demonstrates a notable enhancement in the performance of FAA when integrated into our framework.

3) *Transfer Learning*: In the data augmentation context, DA methods' effectiveness is also assessed with transfer learning [56], [57]. Therefore, we pre-train models on the CIFAR-100

TABLE VIII  
TRANSFERRED TEST ACCURACY ON CIFAR-10. IPF-\* MEANS THE CORRESPONDING METHODS ARE INTEGRATED INTO OUR FRAMEWORK.

	Original	IPF-*
Cutout [4]	92.41	<b>92.80</b>
RandomErasing [19]	92.58	<b>92.94</b>
CutMix [5]	92.78	<b>92.99</b>
GridMask [17]	91.50	<b>92.11</b>
AdvMask [10]	91.51	<b>91.73</b>
AutoAugment [6]	92.80	<b>93.19</b>
RandAugment [7]	92.69	<b>92.84</b>
FAA [20]	92.26	<b>93.45</b>
TrivialAugment [8]	92.75	<b>93.39</b>

dataset using various augmentations with and without IPF-RDA, followed by fine-tuning them on the CIFAR-10 dataset. Theoretically, better augmentation methods enhance model performance and transferability to a greater extent. Table VIII illustrates the transferred test accuracy of various data augmentation methods with and without the utilization of IPF-RDA. It can be observed that the incorporation of IPF-RDA yields consistent improvements in the transferred test accuracy. Notably, the integration of IPF-RDA showcases its potential in improving the performance of regional deletion-based data augmentation methods by approximately 0.4%, image mixing-based methods by 0.2%, and image-level methods by around 0.5%. Especially for Fast-AutoAugment, the transferred test accuracy experiences an impressive boost of 1.19% with the integration of IPF-RDA. Therefore, our framework significantly enhances the performance of the original data augmentation methods.

Additionally, the ResNet-50 models trained on the CIFAR-100 and transferred to CIFAR-10 using FAA and IPF-FAA are

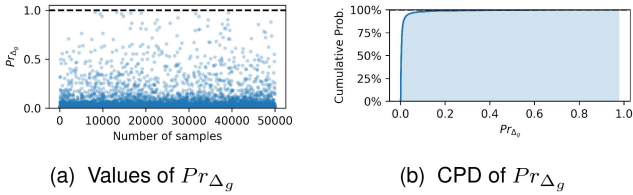


Fig. 5. The proportion of the gradient change  $\Delta_g$  w.r.t.  $g_x$ .

also presented in Fig. 4(c) and (d), respectively. In comparison between Fig. 4(c) and (d), the DI value of original FAA is  $4.43 \times 10^{-5}$ , while that of IPF-FAA is as high as  $7.58 \times 10^{-5}$ . A higher DI value proves that our framework is highly effective in enhancing the transfer performance of data augmentation methods, resulting in models with stronger feature extraction capabilities.

4) *Values of the Proportion  $Pr_{\Delta_g}$* : As aforementioned, the extent of alteration in the saliency map of the adversarial sample is bound by an upper limit, as established by (4). This upper limit’s relative smallness compared to the original gradient signifies a critical insight - the conventional saliency map lacks sensitivity to classification decisions and fails to encapsulate the most discriminative information inherent in images. To assess the practical significance of this upper limit in relation to the original gradient, we undertake an empirical assessment by calculating the ratio of  $\Delta_g$  to  $g_x$ , denoted as  $Pr_{\Delta_g} = \|\Delta_g\|/\|g_x\|$ . Fig. 5(a) and (b) provide the visualizations of the  $Pr_{\Delta_g}$  values and the cumulative probability distribution (CPD) of  $Pr_{\Delta_g}$  utilizing the CIFAR-10 training set, respectively. Evidently, it can be observed that the majority of samples exhibit  $Pr_{\Delta_g}$  values below 1% of the original gradient, with more than 98% of instances falling within this threshold. The gradient changes for the vast majority of samples are very small magnitudes, which are difficult to identify and capture. These findings demonstrate that despite the significant changes in image classification results, the gradient changes observed in saliency maps are still notably small in magnitude. Consequently, the saliency methods cannot be used as a basis for classification decisions, which also confirms the validity of our proposed approach.

5) *Class Activation Maps on Oxford Flower Dataset*: Since class activation maps (CAM) visualization can provide visual explanations for understanding model performance, we incorporate CAM visualization as a tool to analyze the discriminative regions used by models. Since regional deletion-based methods can enhance the robustness of occlusion and increase the receptive field, we compare the performance before and after integrating these methods into our framework. Fig. 6 illustrates the CAM generated by ResNet-50 trained with commonly used Cutout and Random Erasing on the Oxford Flower dataset [47]. After being integrated into our framework, these methods are more inclined to locate and highlight the most relevant parts of the main objects, while the background is almost ignored. For instance, the original Cutout and RandomErasing’s regions of interest cover only part of the main object (i.e., flower) and contain some irrelevant background information, indicating an overfitting problem. In contrast, utilizing our framework, these

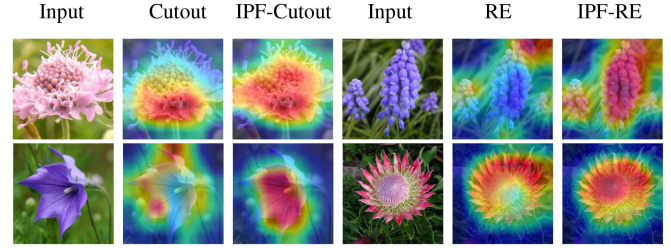


Fig. 6. Class activation mapping (CAM) for Oxford Flower Dataset [47].

TABLE IX  
EFFICACY OF CDIEA ON VARIOUS BENCHMARK DATASETS WITH  $\epsilon = 8/255$

Dataset	Success Rate(%)	$\ell_0(\delta)$	% of pixels
CIFAR-10 [21]	100	164.1	16.0
CIFAR-100 [21]	100	172.5	16.8
Tiny-ImageNet [22]	98.9	248.75	6.1
Market1501 [24]	100	1731.04	3.5
CUHK03 [23]	99.9	1889.9	3.8

methods’ regions of interest cover almost the entire flower while ignoring much of the background, indicating better generalization ability. To summarize, these visualizations underscore how our framework surpasses the efficacy of the original regional deletion-based data augmentation methodologies.

6) *Efficacy of CDIEA*: The results of CDIEA are crucial for balancing the diversity of augmented data and identifying the critical information. To evaluate the efficacy of CDIEA in identifying the most class-discriminative information in images, we present the success rate of  $x + \delta$  in influencing classification decisions, i.e., attack success rate (ASR). Table IX presents results on several benchmark datasets. It can be observed that CDIEA achieves a nearly 100% success rate on these datasets while maintaining a low  $\ell_0$  norm, which means that only a marginal subset of candidate pixels are pinpointed as critical pixels. As aforementioned, this preservation of limited but highly informative pixels significantly contributes to the diversity of augmented images, because only a small portion of the data will be considered as key information to preserve. Consequently, by adaptively identifying and preserving the most class-discriminative regions within images, CDIEA ensures that augmented data not only retains its diversity but also provides valuable information for model training.

7) *Efficacy of Rectangular Regions Preservation in IPF-RDA*: In IPF-RDA, following [4], [8], [9], [10], [17], we preserve a rectangular region containing the class-discriminative information. This design helps maintain the structured semantics in images while incurring negligible computational overhead.

To assess the effectiveness of this choice, we compare IPF-RDA with a point-level preservation variant based on image-level transformations (IPF-TA), in which the most class-discriminative points are retained after augmentation. As shown in Table X, rectangular region preservation consistently outperforms the point-level preservation across different datasets and deep models by a large margin. These results highlight that maintaining contiguous rectangular areas better preserves the

TABLE X  
EFFICACY OF RECTANGLE-LEVEL PRESERVATION ON CIFAR-10/100 USING RESNET-18/50. WE REPORT ERROR RATE (%).

Method	CIFAR-10		CIFAR-100	
	ResNet-18	ResNet-50	ResNet-18	ResNet-50
Point-level	5.60	5.20	24.10	21.32
Ours	<b>3.18</b>	<b>2.85</b>	<b>20.67</b>	<b>18.35</b>

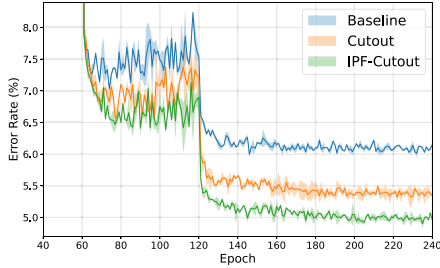


Fig. 7. Curves of test errors on CIFAR-10 with ResNet-110.

structural and semantic characteristics of images. In contrast, point-level preservation may suffer from instability, as isolated points can be overly sensitive to adversarial attack pixels, thereby introducing misleading signals during training. Moreover, we further discuss potential extensions to more advanced preservation strategies, such as superpixel-based regions, in Section V.

8) *Convergence Analysis of IPF-RDA*: While our method significantly contributes to model training, in this section, we demonstrate its ability to enhance the convergence rate of data augmentation methods, such as Cutout. To more clearly present the dynamic evolution of test errors throughout the training process, following the experimental settings suggested in [32], we train ResNet-110 models on the CIFAR-10 dataset, utilizing a multi-step learning rate decay schedule. The learning rate is initialized as 0.1 and multiplied by 0.2 at epochs 60, 120, 160, and 200. It is worth noting that all other experimental settings remain unchanged.

Fig. 7 illustrates the test error curves corresponding to the baseline, Cutout, and IPF-Cutout, respectively. It can be observed that IPF-Cutout consistently showcases substantial improvements compared to both the baseline and the original Cutout technique. This improved performance becomes particularly pronounced after the second learning rate drop, subsequently stabilizing the test errors of IPF-Cutout at notably lower levels. By incorporating the Cutout augmentation within the proposed framework, we not only expedite the convergence rate but also enhance the overall performance beyond what can be achieved by the original Cutout approach. This empirical evidence substantiates the remarkable efficacy of our framework, underscored by its capacity to facilitate more efficient training processes and substantially elevate overall performance.

H. Additional Offline Training Cost

We analyze the additional offline training costs introduced by our CDIEA module. For a better understanding, we compare it with the full training costs of widely-used deep models on CIFAR-100, including ResNet-18, ResNet-50, WRN-28-10,

TABLE XI  
COMPARISON OF TRAINING COSTS BETWEEN CDIEA AND POPULARLY USED DEEP MODELS ON CIFAR-100. MODEL COSTS ARE MEASURED USING A 2-NVIDIA-2080TI-GPU SERVER, WHILE CDIEA COST IS MEASURED ON A SINGLE NVIDIA-2080TI GPU. WE REPORT TOTAL GPU HOURS (H).

	R-18	R-50	WRN-28-10	Shake-2-32	CDIEA
Training Cost (h)	4.3	9.1	21.4	44.5	2.5

TABLE XII  
ADDITIONAL TIME CONSUMPTION OF IPF-RDA ON CIFAR-10/100 DATASETS ACROSS DIFFERENT DEEP ARCHITECTURES

Networks	FLOPs	Additional Training Costs			
		Cutout [4]		AutoAugment [6]	
		CIFAR-10	CIFAR-100	CIFAR-10	CIFAR-100
ResNet-18 [48]	1.82G	+1.1%	+1.4%	+2.4%	+3.1%
ResNet-50 [48]	4.12G	+0.8%	+1.4%	+0.8%	+0.1%
WRN-28-10 [49]	5.25G	+1.0%	+1.0%	+1.0%	+1.2%
Shake-26-2x32d [50]	426.89M	+0.1%	+0.1%	+9.6%	+2.4%

and Shake-Shake-2x 32 d, using a batch size of 256 and 8 parallel workers. As shown in Table XI, CDIEA requires only 2.5 GPU hours for training, significantly lower than the cost of training any of the standard models. More importantly, this is a one-time cost per dataset: once trained, CDIEA generates class-discriminative guidance that can be reused across different architectures and experiments without retraining. This design minimizes computational overhead and enhances the training efficiency of our method.

I. Additional Online Training Cost

Time costs are also a crucial metric for evaluating data augmentation methods. In this section, we provide the additional training costs incurred by the IPF-RDA framework during the online training phase. All models are trained on 2 NVIDIA RTX2080TI GPUs with a batch size of 256 and 8 parallel workers across five independent random trials. Specifically, in Table XII, we integrate the two most representative methods, Cutout and AutoAugment, into our framework on the CIFAR-10/100 datasets employing various deep networks. It can be observed that IPF-RDA increases the computational costs of the original data augmentation methods by no more than 3% across these most widely used deep networks on both CIFAR-10 and CIFAR-100 datasets. Because CDIEA is decoupled from the online training procedure, it does not introduce additional training costs. Therefore, the additional training costs are engendered by the random sampling of structural regions for information preservation. Consequently, IPF-RDA does not significantly affect the training efficiency of these data augmentation approaches. For a detailed discussion on the offline training costs of CDIEA, please refer to Section V.

J. Ablation Study

1) *The Effect of CDIEA*: To further evaluate the effectiveness of CDIEA, we perform an ablation study in which CDIEA is replaced with alternative forms of key information: (1) corner points extracted using the Harris detector, and (2) saliency

TABLE XIII  
THE EFFECT OF CDIEA ON CIFAR-10 USING VARIOUS ARCHITECTURES. WE  
REPORT ERROR RATES (%)

	Resnet-18	Resnet-50	WRN-28-10
baseline	4.72	4.34	4.48
Corner Point	3.66	3.44	2.59
Salient Point	3.52	3.30	2.55
Ours	<b>3.18</b>	<b>2.85</b>	<b>2.23</b>

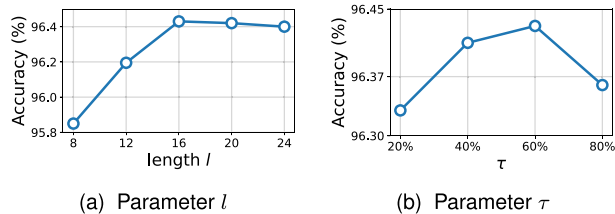


Fig. 8. The ablation study of parameters for IPF-RDA.

maps generated by a pretrained ResNet-50. The saliency-based information can be seamlessly incorporated into our framework.

As shown in Table XIII, all variants of our information-preserving framework outperform the baseline, highlighting the overall benefits of the IPF in data augmentation. Among them, CDIEA consistently achieves the lowest error rates across various architectures, validating its superior capability in identifying class-discriminative information. These results demonstrate that, while pretrained saliency-based methods offer certain advantages, CDIEA provides more effective guidance when integrated into the IPF framework.

2) *The Effect of Parameters*: To better understand the efficacy of two parameters in IPF-RDA, specifically denoted as  $l$  and  $\tau$ , we embark on a series of ablation studies conducted on the CIFAR-10 dataset. As demonstrated in Fig. 8(c), a controlled examination is carried out wherein  $\tau$  is fixed as 60%, while varying  $l$  values spanning the range from 8 to 24. It can be seen that the highest accuracy is obtained when  $l = 16$ , which is consistent with the suggestions in [4]. In Fig. 8(d), we fix  $l$  as 16 and employ various  $\tau$  values from 20% to 80%. It can be observed that IPF-RDA with  $\tau = 60\%$  performs best. In fact, a larger value of  $\tau$  indicates that the preserved regions contain information with higher importance scores. However, larger  $\tau$  values lead to a reduced sample space size, limiting the diversity of augmented data and increasing the sampling consumption. Meanwhile, a small value of  $\tau$  leads to our method being insufficiently sensitive in capturing key information. For a more comprehensive exploration encompassing diverse combinations of  $\tau$  and  $l$ , please refer to Appendix F, available online.

*Parameter Settings*: The selection of the parameter  $l$  is consistent with the suggestion in [4]. Expressly, we set  $l$  to 8 for CIFAR-100 and 16 for others. Moreover, adhering to the recommendations articulated in [9],  $\tau$  is set as a constant value of 60% across all datasets. Consequently, the parameters within the IPF-RDA are easy to configure, enhancing the practical applicability and ease of utilization.

## V. DISCUSSION AND FUTURE WORK

In this work, we propose a novel information-preserving framework to preserve critical information for robust data augmentation, namely IPF-RDA. The architecture of IPF-RDA consists of a two-step process. It first identifies the most class-discriminative information, concurrently assigning corresponding importance scores. Subsequently, this information forms the foundation for the seamless integration of various types of data augmentation methods, carefully categorized based on their operational characteristics. Our experimental results demonstrate that IPF-RDA effectively enhances the robustness of data augmentation methods and substantially further improves the generalization performance of deep neural networks. Thus, while data augmentation techniques have already proven their effectiveness in enhancing the performance of deep models, our framework strengthens their potential capabilities to a considerable extent.

While our extensive validation demonstrates that the proposed framework integrates well with a wide range of DA techniques, yielding consistent performance gains, incorporating image-level mixing methods, such as Mixup [27], presents certain challenges. These methods generate convex combinations of image and labels, making it difficult to isolate and preserve class-discriminative information from the mixed content. Beyond image-level augmentation, feature augmentation methods [58], [59] face similar issues, as pixel-level discriminative regions may not directly map to latent features. Extending our framework to address these challenges remains an interesting direction for future exploration.

Regarding spatial transformations-based DA, geometric deformation potentially misaligns the augmented images with the original class-discriminative features. Our framework can be readily adapted to these cases by applying the same geometric transformation to the preserved regions, thereby maintaining spatial alignment. However, when the transformation is overly aggressive, the key information may be compromised. For instance, when critical objects are displaced or their structural integrity is warped, the preserved regions may no longer contain meaningful semantics. To mitigate such adverse effects, additional strategies, such as semantic-aware augmentation or dynamic region re-selection, may be necessary to ensure effective information preservation.

In addition, while CDIEA has proven effective in capturing critical information within images, it inevitably incurs additional computational costs. Similar to automatic DA methods that use reinforcement learning to pre-search the augmentation space, our framework is designed to strike an optimal balance between efficiency and performance. By decoupling CDIEA from the online training process, we ensure that its impact on overall training time remains minimal. Once the one-time procedure is completed, no further CDIEA training or inference is needed. In the future, we envision a more lightweight variant that can be optionally integrated into the online training loop, offering an extra level of efficiency without altering the core benefits of the framework.

During augmentation, we leverage the rectangle-level regions for information preservation in IPF-RDA, which achieves a

favorable balance between efficacy and efficiency [9], [10], [17]. While this design effectively maintains structured spatial information with minimal computational cost, more advanced alternatives, such as superpixel-based or semantic-segmentation-guided region selection, could align the preserved areas more closely with object boundaries and semantic parts, thereby enhancing the retention of fine-grained class-discriminative features based on our point-level importance scores. As such methods may inevitably introduce extra overhead, an interesting direction for future work is to explore lightweight, semantically aligned preservation strategies that maintain these benefits without compromising efficiency.

While our evaluation demonstrates that IPF-RDA consistently improves performance, its benefits may be compromised in scenarios with severe label noise. For example, we conduct experiments on CIFAR-100 with 50% label noise. In this setting, the baseline, TrivialAugment, and IPF-TA achieve error rates of 36.35%, 36.80%, and 36.40%, respectively. These results indicate that although IPF-RDA can alleviate the performance drop, it cannot deliver substantial improvements in the presence of noisy data. This is because preserving class-discriminative regions may inadvertently reinforce misleading features and learning signals tied to incorrect labels, thereby impairing training. Developing noise-robust preservation strategies, such as label uncertainty estimation, remains an interesting direction for future work.

## VI. CONCLUSION

In this paper, we propose an information-preserving framework for robust data augmentation (IPF-RDA) with the aim of enhancing the robustness of data augmentation methods and further improving their performance. The primary motivation behind IPF-RDA is the need to identify the class-discriminative pixels that are most vulnerable to perturbations; thus, these sensitive points are also very vulnerable to data augmentation operations. Subsequently, our framework protects this sensitive information from being influenced after augmentations. By preserving this critical information while not sacrificing the diversity of augmented data, our framework can boost the full potential of data augmentation approaches. Experiment results on various computer vision benchmarks demonstrate the effectiveness and efficiency of the proposed method.

## ACKNOWLEDGMENT

The authors thank Jinqiao Li, Geng Zhang, Yun Hao, and Jiajun Hu from Nanjing University for their helpful discussions and comments.

## REFERENCES

- [1] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognit.*, vol. 137, 2023, Art. no. 109347.
- [2] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Jul. 2019.
- [3] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," 2022, *arXiv:2204.08610*.
- [4] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.
- [5] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6022–6031.
- [6] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation strategies from data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 113–123.
- [7] E. D. Cubuk, B. Zoph, J. Shlens, and Q. Le, "RandAugment: Practical automated data augmentation with a reduced search space," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 18613–18624.
- [8] S. G. Müller and F. Hutter, "TrivialAugment: Tuning-free yet state-of-the-art data augmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 774–782.
- [9] C. Gong, D. Wang, M. Li, V. Chandra, and Q. Liu, "KeepAugment: A simple information-preserving data augmentation approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1055–1064.
- [10] S. Yang, J. Li, T. Zhang, J. Zhao, and F. Shen, "AdvMask: A sparse adversarial attack-based data augmentation method for image classification," *Pattern Recognit.*, vol. 144, 2023, Art. no. 109847.
- [11] S. Lin, Z. Zhang, X. Li, and Z. Chen, "SelectAugment: Hierarchical deterministic sample selection for data augmentation," in *Proc. 37th AAAI Conf. Artif. Intell. 35th Conf. Innov. Appl. Artif. Intell. 13th Symp. Educ. Adv. Artif. Intell.*, 2023, pp. 1604–1612.
- [12] A. Ma, N. Dvornik, R. Zhang, L. Pishdad, K. G. Derpanis, and A. Fazly, "SAGE: Saliency-guided mixup with optimal rearrangements," 2022, *arXiv:2211.00113*.
- [13] J. Gu and V. Tresp, "Saliency methods for explaining adversarial attacks," 2019, *arXiv:1908.08413*.
- [14] P.-J. Kindermans et al., "The (Un)reliability of saliency methods," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Berlin, Germany: Springer, 2019, pp. 267–280.
- [15] A. Mahendran and A. Vedaldi, "Salient deconvolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 120–135.
- [16] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 9525–9536.
- [17] P. Chen, S. Liu, H. Zhao, and J. Jia, "GridMask data augmentation," 2020, *arXiv:2001.04086*.
- [18] K. K. Singh and Y. J. Lee, "Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 3544–3553.
- [19] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. 34th AAAI Conf. Artif. Intell. 32nd Innov. Appl. Artif. Intell. Conf. 10th AAAI Symp. Educ. Adv. Artif. Intell.*, 2020, pp. 13001–13008.
- [20] S. Lim, I. Kim, T. Kim, C. Kim, and S. Kim, "Fast AutoAugment," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, Art. no. 598.
- [21] A. Krizhevsky et al., "Learning multiple layers of features from tiny images," 2009.
- [22] P. Chrabaszcz, I. Loshchilov, and F. Hutter, "A downsampled variant of ImageNet as an alternative to the CIFAR datasets," 2017, *arXiv:1707.08819*.
- [23] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 152–159.
- [24] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2015, pp. 1116–1124.
- [25] S. Yang, S. Guo, J. Zhao, and F. Shen, "Investigating the effectiveness of data augmentation from similarity and diversity: An empirical study," *Pattern Recognit.*, vol. 148, 2024, Art. no. 110204.
- [26] S. Yang, F. Shen, and J. Zhao, "EntAugment: Entropy-driven adaptive data augmentation framework for image classification," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2024, pp. 197–214.
- [27] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 2866–2878.
- [28] J.-H. Kim, W. Choo, and H. O. Song, "Puzzle mix: Exploiting saliency and local statistics for optimal mixup," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5275–5285.
- [29] J.-H. Kim, W. Choo, H. Jeong, and H. O. Song, "Co-Mixup: Saliency guided joint mixup with supermodular diversity," 2021, *arXiv:2102.03065*.
- [30] K. Islam, M. Zaheer, A. Mahmood, K. Nandakumar, and N. Akhtar, "GenMix: Effective data augmentation with generative diffusion model image editing," 2024, *arXiv:2412.02366*.

- [31] N.-T. Tran, V.-H. Tran, N.-B. Nguyen, T.-K. Nguyen, and N.-M. Cheung, "On data augmentation for GAN training," *IEEE Trans. Image Process.*, vol. 30, pp. 1882–1897, 2021.
- [32] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, and C. Wu, "Regularizing deep networks with semantic data augmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3733–3748, Jul. 2022.
- [33] A. Mahendran and A. Vedaldi, "Salient deconvolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2016, pp. 120–135.
- [34] J. Gu and V. Tresp, "Saliency methods for explaining adversarial attacks," 2019, *arXiv:1908.08413*.
- [35] X. Cheng, H. Zhang, Y. Xin, W. Shen, and Q. Zhang, "Why adversarial training of reLU networks is difficult?," 2022, *arXiv:2205.15130*.
- [36] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [37] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 818–833.
- [38] J.-H. Luo and J. Wu, "AutoPruner: An end-to-end trainable filter pruning method for efficient deep model inference," *Pattern Recognit.*, vol. 107, 2020, Art. no. 107461.
- [39] F. Croce and M. Hein, "Sparse and imperceivable adversarial attacks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 4724–4732.
- [40] Y. Dong et al., "Boosting adversarial attacks with momentum," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9185–9193.
- [41] R. Gontijo-Lopes, S. Smullin, E. D. Cubuk, and E. Dyer, "Tradeoffs in data augmentation: An empirical study," in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 13560–13586.
- [42] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on CIFAR-10," Univ. Toronto, Toronto, Canada, Tech. Rep., 2010.
- [43] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 6256–6268.
- [44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
- [45] Y. LeCun, C. Cortes, and C. J. Burges, "The MNIST database of handwritten digits," 1998. [Online]. Available: <http://yann.lecun.com/exdb/mnist>
- [46] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.
- [47] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Proc. 6th Indian Conf. Comput. Vis. Graph. Image Process.*, 2008, pp. 722–729.
- [48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [49] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proc. Brit. Mach. Vis. Conf.*, 2016, pp. 87.1–87.12.
- [50] X. Gastaldi, "Shake-shake regularization," 2017, *arXiv:1705.07485*.
- [51] Y. Yamada, M. Iwamura, T. Akiba, and K. Kise, "Shakedrop regularization for deep residual learning," *IEEE Access*, vol. 7, pp. 186126–186136, 2019.
- [52] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [53] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1492–1500.
- [54] S. Zhang, Z. Yin, X. Wu, K. Wang, Q. Zhou, and B. Kang, "FPB: Feature pyramid branch for person re-identification," 2021, *arXiv:2108.01901*.
- [55] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 3239–3250.
- [56] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [57] F. Zhuang et al., "A comprehensive survey on transfer learning," in *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.
- [58] P. Li, D. Li, W. Li, S. Gong, Y. Fu, and T. M. Hospedales, "A simple feature augmentation for domain generalization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 8866–8875.
- [59] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," 2017, *arXiv:1702.05538*.
- [60] K. Sohn et al., "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 596–608.



**Suorong Yang** received the BSc degree from Nanjing University, Nanjing, China, in 2019. He is currently working toward the PhD degree in computer science with Nanjing University, Nanjing, China. His current research interests include data augmentation and deep learning.



**Hongchao Yang** received the BSc degree from Wuhan University, Wuhan, China, in 2022. He is currently working toward the MS degree in computer science with Nanjing University, Nanjing, China. His current research interests include deep learning and neural networks.



**Suhan Guo** received the BA degree from Johns Hopkins University, Baltimore, Maryland, USA, in 2016, and the MPH degree from New York University, New York, New York, USA, in 2019. She is currently working toward the PhD degree in artificial intelligence with Nanjing University, Nanjing, China. Her current research interests include data augmentation and neural networks.



**Furao Shen** received the BSc and MSc degrees in mathematics from Nanjing University, Nanjing, China, in 1995 and 1998, respectively, and the PhD degree from the Tokyo Institute of Technology, Tokyo, Japan, in 2006. He is currently a full professor of computer science and technology with Nanjing University. His current research interests include neural computing and robotic intelligence.



**Jian Zhao** (Senior Member, IEEE) received the BS degree from Nanjing University, Nanjing, China, in 2001, the MSc degree from the Hamburg University of Technology, Hamburg, Germany, in 2004, and the PhD degree in electrical engineering from the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, in 2010. Since December 2010, he has been with the Institute for Infocomm Research, Singapore. His research interests include optimization techniques in wireless communications, multiuser MIMO communications, and cooperative communications. He has received a number of awards, including the DAAD-Siemens Asia 21st Century Scholarship, IEEE Globecom 2008 Best Paper Award, and Chinese Government Award for Outstanding Self-Financed Students Abroad.