# PEARL: Preprocessing Enhanced Adversarial Robust Learning of Image Deraining for Semantic Segmentation

Xianghao Jiao
Dalian University of Technology
jxh@mail.dlut.edu.cn

Yaohua Liu
Dalian University of Technology
liuyaohua_918@163.com

Jiaxin Gao
Dalian University of Technology
jiaxinn.gao@outlook.com

Xinyuan Chu
Dalian University of Technology
chuxinyuan_dut@outlook.com

Xin Fan
Dalian University of Technology
xin.fan@dlut.edu.cn

Risheng Liu*
Dalian University of Technology
Peng Cheng Laboratory
rsliu@dlut.edu.cn

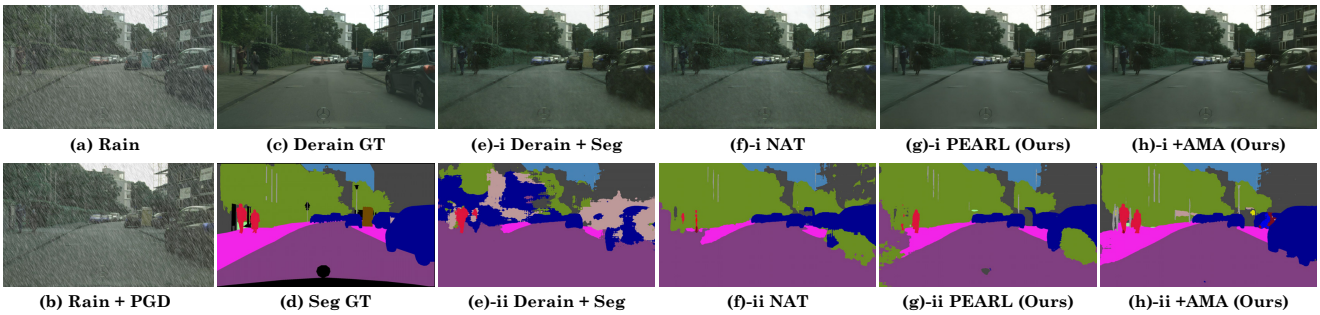



**Figure 1: The visualization results of image deraining and semantic segmentation tasks among the baseline (Derain + Seg) and our proposed NAT framework, PEARL framework and PEARL with AMA generator (denoted as +AMA) with the influence of both degradation factors, i.e., rain streaks and PGD attacks. It can be obviously seen that our proposed framework obtains derained images with higher quality which also leads to more accurate segmentation labels.**

## ABSTRACT

In light of the significant progress made in the development and application of semantic segmentation tasks, there has been increasing attention towards improving the robustness of segmentation models against natural degradation factors (e.g., rain streaks) or artificially attack factors (e.g., adversarial attack). Whereas, most existing methods are designed to address a single degradation factor and are tailored to specific application scenarios. In this work, we present the first attempt to improve the robustness of semantic segmentation tasks by simultaneously handling different types of degradation factors. Specifically, we introduce the Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework based on the analysis of our proposed Naive Adversarial Training (NAT) framework. Our approach effectively handles both rain streaks and adversarial perturbation by transferring the robustness of the segmentation model to the image derain model. Furthermore, as opposed to the commonly used Negative Adversarial Attack (NAA), we design the Auxiliary Mirror Attack (AMA) to introduce positive information prior to the training of the PEARL framework, which improves defense capability and segmentation performance. Our extensive experiments and ablation studies based on different derain methods and segmentation models have demonstrated the significant performance improvement of PEARL with AMA in defense against various adversarial attacks and rain streaks while maintaining high generalization performance across different datasets. The source codes are available at https://github.com/JiaoXianghao/PEARL.

*Corresponding author


*MM '23, October 29-November 3, 2023, Ottawa, ON, Canada*

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org/10.1145/3581783.3612164

## CCS CONCEPTS

• **Computing Methodologies** → **Computer vision**; **Semantic segmentation**; **Image deraining**; **Adversarial Defense**.

## KEYWORDS

adversarial attack, semantic segmentation, single image deraining, preprocessing enhanced learning, auxiliary mirror attack

**ACM Reference Format:**
Xianghao Jiao, Yaohua Liu, Jiaxin Gao, Xinyuan Chu, Xin Fan, and Risheng Liu. 2023. PEARL: Preprocessing Enhanced Adversarial Robust Learning of Image Deraining for Semantic Segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November*

*3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3581783.3612164

# 1 INTRODUCTION

Semantic segmentation [28, 53, 56] has achieved great advances with the development of deep learning networks and high-quality collected data. Meanwhile, the significant performance improvement of segmentation models also boost its application to satisfy various real-world demands [16, 33, 39, 62], e.g., self-driving systems, virtual reality, etc. Whereas, as various adversarial attacks and variations have been explored, the vulnerability of deep neural networks to these adversarial examples [25, 46, 58] also have attracted much attention. By introducing imperceptible adversarial perturbation to the input of semantic segmentation model, the segmentation results [2, 21, 48, 49] could be badly corrupted and results in serious safety issues. Moreover, since most existing methods [8, 54] developed their approaches relying on assumption of degradation-free scenarios [9], the performance of segmentation model has no guarantee under bad imaging conditions or adverse weather such as rain and fogs.

A series of attack and defense methods have been developed to improve adversarial robustness. Classic attack methods such as FGSM [17], PGD [34] and their variations [10, 47, 60] have shown degrading effect on segmentation tasks. Besides, another line of work [1, 18] has explored the difference between semantic segmentation and image classification to design task-specific attack methods. As one of most effective defense strategy, Adversarial Training (AT) [26, 40, 43] addresses the vulnerability of segmentation model by incorporating the adversarial example during the training process. In addition to few AT based methods [52], several works also apply teacher-student structure [4] and multitask learning [35] to improve the robustness of segmentation model.

To improve the limited performance of semantic segmentation under extreme weather, recently proposed methods have explored various techniques for low-level tasks. Take the rainy weather as an example, Single Image Deraining (SID) [11, 15, 45] aims to remove the degradation noise from the input rainy images and retains as much context details as possible, which could be embedded as the low-level preprocessing procedure to benefit the downstream segmentation tasks. In comparison with the optimization based methods [24, 31, 51, 57], varieties of deep learning based methods [14, 22, 29] explore different network structures to obtain significant performance based on massive training data. Besides, several methods [31, 59] have also incorporated the high-level semantic knowledge as efficient feedback to facilitate the deraining process.

Whereas, the above methods essentially focus on eliminating a specific influence factor to enhance the adaptability or robustness of segmentation model in real-world applications. To be general, the environmental degradation phenomenon and artificially introduced adversarial perturbation share similar principles for segmentation tasks, and could be regarded as some specific form of degradation factors added to the input. From this new perspective, we make our attempt to design a novel framework to jointly handle both types of degradation factors without introducing additional network parameters or task-specific loss functions.

Firstly, we introduce the Naive Adversarial Training (NAT) framework, which improves the robustness of segmentation model based on AT while handling the rain streaks by embedding extra image deraining module. Whereas, separately removing the rain steaks and defending adversarial perturbation will deteriorate the derain model and introduce residual perturbation to the output of the derain model, which finally affects the downstream tasks. Inspired by the idea which designs specialized transformation module concatenated to the original classification model to defend adversarial examples, we here propose to transfer the robustness of segmentation model to the derain model, and design the Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework to simultaneously deal with both adversarial perturbation and rain streaks. Moreover, as opposed to the Negative Adversarial Attack (NAA), we propose the Auxiliary Mirror Attack (AMA) to introduce "positive" information prior of the adversarial attack to the supervised training of derain model, which enhances the defense capability of derain model and improves the segmentation results eventually. Experimentally, we conducted extensive experiments and ablation studies to demonstrate the performance improvement of both derain and segmentation results with quantitative and visualization results. Moreover, we also verify the generalization performance of our framework across different datasets.

The main contributions of this paper are summarized as follows.

- To the best of our knowledge, we make the first attempt delving into downstream semantic segmentation tasks influenced by both natural degradation factor (e.g., rain streaks) and artificially generated degradation factors (e.g., adversarial attacks), and significantly improve the downstream task performance under bad weather while retaining the robustness against adversarial attacks.
- In contrast with the proposed Naive Adversarial Training (NAT) framework, we introduce our Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework to transfer the robustness of segmentation model, aiming to obtain high-performance robust derain model, which can effectively eliminate the influence of both degradation factors on the segmentation attacks.
- We design another Auxiliary Mirror Attack (AMA) as opposed to the Negative Adversarial Attack (NAA) to embed positive perturbation to the proposed PEARL framework, which facilitates the robust learning to improve the defense capability and leads to better segmentation results.
- Extensive experimental results and ablation studies on different derain and segmentation models have demonstrated the effectiveness of PEARL framework and AMA module to enhance the robustness against various degradation attacks and improve the deraining performance. Besides, we also verify the generalization performance of our proposed framework based on different datasets.

# 2 RELATED WORKS

## 2.1 Image Deraining for Semantic Segmentation

**High-level Segmentation Task.** As a specific form of pixel-level dense classification tasks, semantic segmentation have been well developed to explore the contextual dependencies and capture the

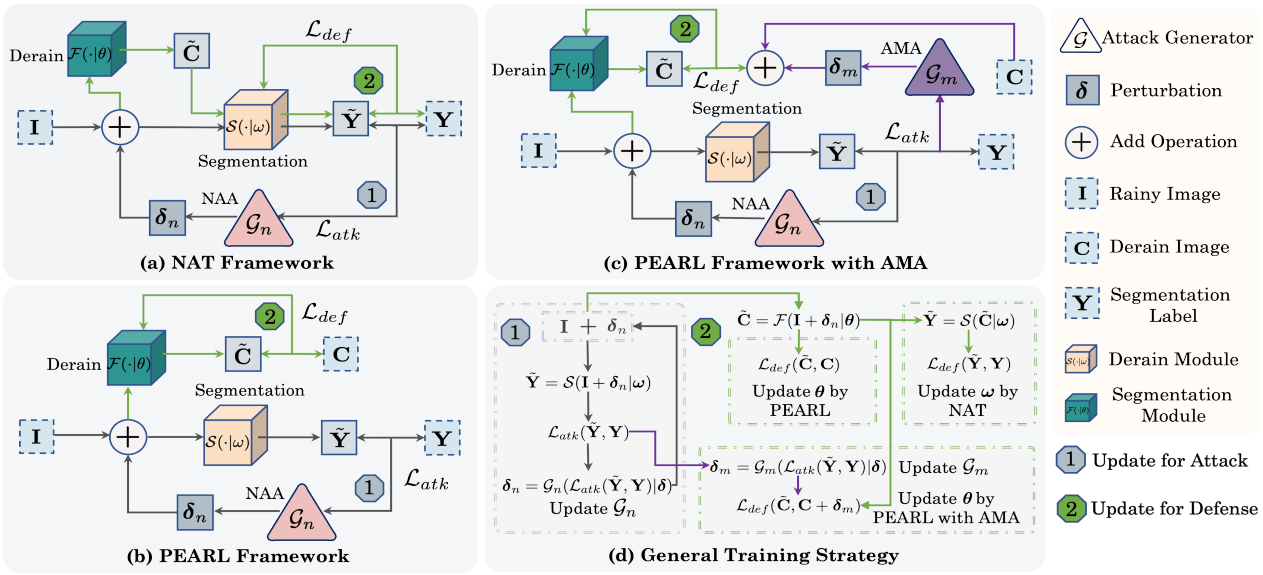



**Figure 2: The first three subfigures illustrate the Naive Adversarial Training (NAT) training framework for handling rain streaks and adversarial attacks for image segmentation model, our Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework and its whole pipeline with proposed Auxiliary Mirror Attack (AMA) technique. The last subfigure describes the training strategy for NAT, PEARL and PEARL with AMA. We use gray, green and purple lines to denote the optimization cycle of attack, defense and additional flow introduced by the AMA module.**

long-range relationship. Chen et.al. [6] proposed DeepLab, which introduces the atrous convolution for explicit resolution control and uses spatial pyramid pooling for multi-scale objective segmentation. Then they developed the DeepLabv3 [7] for further improvements to atrous convolution and atrous spatial pyramid pooling modules and also incorporates image-level features for global context. Zhao et. al. [61] presents PSPNet which exploits global context information and pyramid pooling module to improve segmentation performance on various scene parsing datasets. Xie et.al. [50] proposed SegFormer to unifiy Transformers with lightweight MLP decoders and achieves state-of-the-art performance with simple and efficient design. Practically speaking, the above methods have spared efforts to work on degradation-free scenes, which may faces serious performance decrease under adverse weather.

**Low-level Deraining Task.** Single Image Deraining (SID) [11, 15] has been well developed to deal with different rain streaks and improve the downstream tasks for practical applications. Typically, Li et.al. [31] proposed RESCAN to incorporate dilated convolutional neural networks and recurrent neural networks to remove rain streaks in multiple stages. Ren et.al. [38] constructs a better and simpler baseline deraining network, called PReNet, which provides consistent improvements of the architecture and loss functions. Zamir et.al. [55] introduces a multi-stage architecture called MPRNet to progressively learn restoration functions for degraded inputs and balances the competing goals of spatial details and high-level contextualized information in image restoration tasks. Recently, Valanarasu et. al. [44] proposed transformer-based model with a single encoder and a decoder that can restore an image degraded by any weather condition. Besides, a line of works [23, 29] also explore the high-level semantic information, such as the detection and segmentation results, to guide the optimization of deraining process. Note that our propose training framework of image deraining implies no explicit requirements of the network structure or design of loss functions, which makes it capable to incorporate recent-proposed methods to obtain higher performance.

## 2.2 Adversarial Attacks and Defenses

**Adversarial Attacks.** It has been investigated [2, 49] that the segmentation model also shows vulnerability to these artificially introduced adversarial examples. Generally speaking, the adversarial attacks include two categories, i.e., black-box attacks [36] and white-box [3] attacks. Here we focus on the gradient-based white-box attack which is capable to access full knowledge of the model under attack (known as target model), and generated imperceptible perturbations by computing the gradient of target model. Several commonly used adversarial attack methods include FGSM [17] and PGD [34], which generated single-step and multi-step perturbation for the input image. Based on two basic attacks, different attack methods have been explored by introducing practical techniques [12, 60]. kurakin et.al. [27] proposed BIM attack and demonstrates that machine learning systems are vulnerable to adversarial examples even in physical world scenarios. Carlini et.al. [5] challenges the effectiveness of defensive distillation and introduces the optimization based attack method denoted as CW. In addition to the above general-purpose attacks, several works [1, 18] also conduct impressive investigation on the robustness of segmentation and introduce effective improvements of the PGD attack, which also shows its necessity of training robust segmentation model for better defense against the adversarial degradation factors.

**Adversarial Defense.** Generally speaking, Adversarial Training (AT) [26, 43] trains the model to defend the adversarial example by minimizing the attack objective, which also make it more time-consuming due to generation of adversarial example and tasks more epochs to converge. Whereas, few works have explored the effectiveness of AT on the segmentation model tasks. Practically, by setting additional branches in the target model during training and dividing pixels into multiple branches, Xu et.al. [52] proposed DDC-AT for improving the robustness of deep neural networks on semantic segmentation tasks. In addition, another branch of defense methods have investigated different transformations, such as image compression and pixel deflection [37, 41], embedded to preprocess the input, thus remove the adversarial perturbation. There also has been a lack of research in recent years that have continued to investigate this direction. Instead of directly using AT to handle different degradation factors, we here employ the preprocessing based idea to construct the robust learning process with embedded derain model, which is supposed to jointly handle both rain streaks and adversarial attacks. Besides, our framework retains more flexibility to be further improved with task-specific model design and additional loss functions.

## 3 PROPOSED METHOD

In this section, we first provide simplified problem definition of different degradation attacks factors and AT to derive the Naive Adversarial Training (NAT) framework for improving robustness of image segmentation model. With analysis of the limitation of NAT, we further propose our Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework with designed Auxiliary Mirror Attack (AMA) generator, by which simultaneously remove the rain streaks and improve the robustness to defend downstream adversarial attacks.

### 3.1 Naive Adversarial Training Framework Against Degradation Attacks

In this work, we consider an image segmentation model $\mathcal{S}(\cdot|\boldsymbol{\omega})$ parameterized by $\boldsymbol{\omega}$. Given a training dataset $\mathcal{D}_{\text{tr}}$ with labeled data pairs, the segmentation output can be represented as $\tilde{\mathbf{Y}} = \mathcal{S}(\mathbf{C}|\boldsymbol{\omega})$, where $\mathbf{C}$ denotes the input image, and $\tilde{\mathbf{Y}}$ denotes the output label of segmentation. Therefore, this downstream task aims to optimize the following objective: $\min_{\boldsymbol{\omega}} \mathcal{L}_{\text{seg}}(\tilde{\mathbf{Y}}, \mathbf{Y})$, where $\mathbf{Y}$ denotes the groundtruth label of segmentation.

Typically speaking, the adversarial attack is supposed to deteriorate the output label of segmentation model by introducing visually imperceptible perturbation, i.e., $\boldsymbol{\delta}$ to the input image, which can be reformulated as

$$\boldsymbol{\delta} = \underset{\boldsymbol{\delta}, \|\boldsymbol{\delta}\|_p \leq \epsilon}{\arg\max}\, \mathcal{L}_{atk}(\mathcal{S}(\mathbf{C} + \boldsymbol{\delta}|\boldsymbol{\omega}), \mathbf{Y}), \tag{1}$$

where $\boldsymbol{\delta}$ is usually bounded with $\epsilon$-toleration $\ell_p$-norm, $\boldsymbol{\delta} \in [0, 1]$, and $\mathcal{L}_{atk}$ is the adversarial loss to measure the distance between generated degraded example and ground truth. Typically, we could consider the same form of $\mathcal{L}_{\text{seg}}$ to define the adversarial loss function. Based on the above formulation, when we apply $K$-step PGD method to generate the adversarial example, the perturbation at $k$-th step can be denoted as

$$\boldsymbol{\delta}^{k+1} \leftarrow \Pi_\epsilon(\boldsymbol{\delta}^k + \alpha \cdot sgn(\nabla_{\boldsymbol{\delta}} \mathcal{L}_{atk}(\mathcal{S}(\mathbf{C} + \boldsymbol{\delta}^k|\boldsymbol{\omega})), \mathbf{Y})), \tag{2}$$

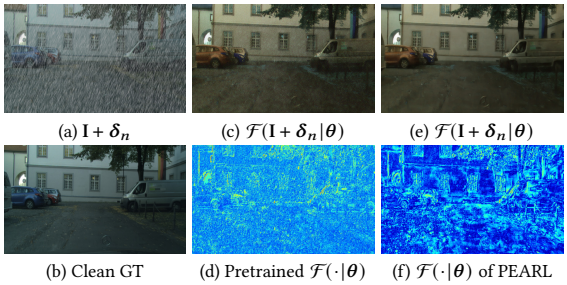
where $k = 0, \ldots, K-1$, $\alpha$ is the step size for perturbation generation, $\Pi_\epsilon(\cdot)$ and $sgn(\cdot)$ denotes the projection operation and element-wise *sign* operation, respectively. The initial perturbation $\boldsymbol{\delta}^0$ is sampled from uniform distribution $U(-\epsilon, \epsilon)$. In the following, we use $\boldsymbol{\delta}_n$ to represent the adversarial attack $\boldsymbol{\delta}^K$ generated by a specific Negative Adversarial Attack (NAA) generatior denoted as $\boldsymbol{\delta}_n = \mathcal{G}_n(\mathcal{L}_{atk}(\mathcal{S}(\mathbf{C} + \boldsymbol{\delta}|\boldsymbol{\omega}), \mathbf{Y})|\boldsymbol{\delta})$, (e.g., PGD), in order to distinguish them from the auxiliary mirror attacks we introduced later.

As it is mentioned above, AT have been extensively investigated to defend the adversarial attacks by solving the following minimax optimization problem

$$\underset{\boldsymbol{\omega}}{\text{minimize}}\, \mathbb{E}_{(\mathbf{C},\mathbf{Y}) \in \mathcal{D}_{\text{tr}}} \Big[ \underset{\boldsymbol{\delta}, \|\boldsymbol{\delta}\|_p \leq \epsilon}{\text{maximize}}\, \mathcal{L}_{atk}(\mathcal{S}(\mathbf{C} + \boldsymbol{\delta}_n|\boldsymbol{\omega}), \mathbf{Y}) \Big]. \tag{3}$$

By alternatively optimizing $\boldsymbol{\omega}$ and generating new perturbation $\boldsymbol{\delta}_n$ with $\mathcal{G}_n(\cdot|\boldsymbol{\delta})$, the robustness of segmentation against adversarial samples generated by different types of NAAs can be consistently improved. The objective of adversarial defense for the segmentation model is denoted as $\mathcal{L}_{def}$, which is usually defined as the same form of $\mathcal{L}_{atk}$.

Whereas, we consider more general setting where the manually designed adversarial attack is essentially regarded as one of the specific form of degradation attacks factors. In this case, we are allowed to consider various degradation factors such as inevitable noises caused by extreme weather, e.g., rain and fog, which are prior existing parts of the original input $\mathbf{C}$. Here we consider the degraded factors as rainstreaks, and denote the degraded rainy image as $\mathbf{I}$ when the rain streaks exist in the input.



(a) $\mathbf{I} + \boldsymbol{\delta}_n$ (c) $\mathcal{F}(\mathbf{I} + \boldsymbol{\delta}_n|\boldsymbol{\theta})$ (e) $\mathcal{F}(\mathbf{I} + \boldsymbol{\delta}_n|\boldsymbol{\theta})$

(b) Clean GT (d) Pretrained $\mathcal{F}(\cdot|\boldsymbol{\theta})$ (f) $\mathcal{F}(\cdot|\boldsymbol{\theta})$ of PEARL

**Figure 3: We compare the processed heat maps of pretrained derain model and our proposed framework to show the difference between derain results and groundtruth with both rain streaks and BIM attack ($\epsilon = 4/255$).**

To alleviate the negative impact of both degradation attacks, i.e., the rain streaks and adversarial perturbation, we first propose the Naive AT (NAT) framework, which can be illustrated in Fig. 2(a). It first embed pretrained derain model (denoted as $\mathcal{F}(\cdot|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameters of derain model) to remove the rain streaks, and retain the robust segmentation model to handle the adversarial attacks for downstream tasks. Whereas, since the derain model encompasses little prior of the adversarial distribution, the perturbations added in the rainy image may deteriorate the deraining

results seriously. In Fig. 3, we analyze the heatmap of deraining results processed with pretrained derain model and the one trained with our proposed framework. As it can be observed, when the perturbation generated to attack the segmentation model exists in the rainy image $\mathbf{I}$, it will also severely degrade the derain result and leave imperceptible disturbance in the output, which is a mix of multiple degradation factors. Consequently, the residual perturbation and rain streaks left in $\tilde{\mathbf{C}} = \mathcal{F}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\theta})$ also results in the performance decrease of robust segmentation model to defend the adversarial attacks, which increases the difficulty of AT based framework. As one of the most significant contributions, we fully explore the potential capability of embedded derain model, and design the following Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework to effectively defend both manually designed attack and natural degradation attacks.

## 3.2 Preprocessing Enhanced Adversarial Robust Learning (PEARL) Framework

To be general, the decomposition mapping function of derain model could be rationally reformulated as: $\mathbf{I} = \mathbf{C} + \mathbf{R}$, where $\mathbf{C}$ and $\mathbf{R}$ represent the clean background and rain layers extracted from the degraded input. According to the above formulation of adversarial attacks, the degraded input with the adversarial perturbation is denoted as $\mathbf{I}+\boldsymbol{\delta}_n$. In this case, the introduced adversarial perturbation may be regarded as certain noises added to the clean image to some extent. Typically, the denoising task aims to learn the following denoising mapping function: $\tilde{\mathbf{C}} \rightarrow \mathbf{C}$, where $\tilde{\mathbf{C}}$ and $\mathbf{C}$ denote the input image with noises and clean image. From this perspective, the rain streaks and adversarial perturbation can be all treated as noises, which could be further removed by embedding denoisers.

Inspired by the preprocessing based methods [19, 32] which remove the adversarial noise by designing specific transformation modules, we here make our attempt to transfer the robustness of segmentation model against adversarial attacks to the embedded derain model. In this case, we no longer follow the AT based formulation in Eq. (3) to implement NAT framework in Fig. 2(a), and introduce the following Preprocessing Enhanced Adversarial Robust Learning (PEARL) framework:

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \mathcal{L}_{def}(\mathcal{F}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\theta}), \mathbf{C}) \\ &s.t. \boldsymbol{\delta} \in \underset{\boldsymbol{\delta}, \|\boldsymbol{\delta}\|_p \leq \epsilon}{\operatorname{argmax}} \mathcal{L}_{atk}(\mathcal{S}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\omega}), \mathrm{Y}), \end{aligned} \tag{4}$$

where $\mathcal{L}_{def}$ could be specified as the objective function of derain model. Besides, we can introduce additional regularization terms to $\mathcal{L}_{def}$ as task priors of downstream segmentation tasks based on the output of derain model. As described in Fig. 2(b) and Eq. (4), the degraded example is still generated by adding adversarial perturbation to the rainy image, while we replace the outer minimization optimization of $\mathcal{S}(\cdot|\boldsymbol{\omega})$ in Eq. (3) with training derain model to learn the following decomposition mapping function:

$$\mathbf{I}+\boldsymbol{\delta}_n \rightarrow \tilde{\mathbf{C}} + (\mathbf{R}+\boldsymbol{\delta}_n), \tag{5}$$

where $\tilde{\mathbf{C}}$ is approximated by minimizing $\mathcal{L}_{def}(\tilde{\mathbf{C}}, \mathbf{C})$.

Practically, we simply restore the segmentation weights pretrianed based on the clean images, and optimize the negative attack generator and derain model in an alternative manner. The derain model trained with PEARL framework is supposed to jointly remove the rain streaks and adversarial noise, thus make the derain results, i.e., $\mathcal{F}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\theta})$, closer to the clean image. Consequently, the preprocessed results have weakened the negative influence of both degradation attack factors, which also enhance the downstream segmentation tasks to a great extent. In the next subsection, to make the utmost of generated adversarial noise based on the inner maximization, we introduce another auxiliary mirror attack to mimic the deterioration process of adversarial attack, and incorporate the generated positive perturbation to facilitate the noise decomposition of derain model.

## 3.3 Auxiliary Mirror Attack (AMA)

In this subsection, we propose another enhancement technique to assist the optimization of derain model and further improve the performance of downstream segmentation tasks. Based on previous definition of $\mathcal{G}_n$, the generated perturbation $\boldsymbol{\delta}_n = \mathcal{G}_n(\mathcal{L}_{atk}(\tilde{\mathrm{Y}}, \mathrm{Y})|\boldsymbol{\delta})$ is added to the input of derain model to involve the degraded attack, as illustrated in Fig. 2(b). By minimizing the outer objective $\mathcal{L}_{def}$ in Eq. (4), we have injected the noise distribution of adversarial attack into the derain model such that $\mathcal{F}(\cdot|\boldsymbol{\theta})$ generalize to this decomposition mapping task and minimize the distance between $\tilde{\mathbf{C}}$ and $\mathbf{C}$. Whereas, it has been investigated [30] that due to limited hardware support and influences of inevitable adverse shooting conditions, the given ground truth may also contain unpredictable biases, which misguide the derain tasks even the downstream tasks. The above phenomenon enlightens us to rethink the supervised clean data and refine them with the proposed auxiliary mirror attack.

Specifically, inspired by the idea [42] which establishes the correlation between restoration and objective detection tasks by generating pseudo ground truth for upstream restoration tasks, we here design an Auxiliary Mirror Attack (AMA) generator denoted as $\mathcal{G}_m(\cdot|\boldsymbol{\delta})$ to generate the mirror attack of NAA aiming to minimize the attack objective $\mathcal{L}_{atk}$. In comparison with the negative impact of $\boldsymbol{\delta}_n$ generated by $\mathcal{G}_n(\cdot|\boldsymbol{\delta})$, $\mathcal{G}_m(\cdot|\boldsymbol{\delta})$ is supposed to dynamically adjust the derain results with attack prior of the inner maximization objective, and add the "positive" perturbation to the clean image. Then the objective of PEARL framework with AMA can be further reformulated as:

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \mathcal{L}_{def}(\mathcal{F}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\theta}), \mathbf{C}+\boldsymbol{\delta}_m) \\ &s.t. \boldsymbol{\delta} \in \underset{\boldsymbol{\delta}, \|\boldsymbol{\delta}\|_p \leq \epsilon}{\operatorname{argmax}} \mathcal{L}_{atk}(\mathcal{S}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\omega}), \mathrm{Y}), \end{aligned} \tag{6}$$

where $\boldsymbol{\delta}_m = \mathcal{G}_m(\mathcal{L}_{atk}(\mathcal{S}(\mathbf{I}+\boldsymbol{\delta}_n|\boldsymbol{\omega}), \mathrm{Y})|\boldsymbol{\delta})$. We describe the whole pipeline of our PEARL framework with AMA in Fig. 2(b). In some degree, PEARL intends to train the derain model to decompose the clean image, thus defend the adversarial attack generated by $\mathcal{G}_n(\cdot|\boldsymbol{\delta})$, while AMA moves one more step to interpolate the mirror attack of $\mathcal{G}_n(\cdot|\boldsymbol{\delta})$ to the ground truth. Consequently, by minimizing $\mathcal{L}_{def}(\tilde{\mathbf{C}}, \mathbf{C}+\boldsymbol{\delta}_m)$, our proposed framework with AMA turns the decomposition mapping in Eq. (6) to the following one:

$$\begin{aligned} \mathbf{I}+\boldsymbol{\delta}_n \rightarrow \tilde{\mathbf{C}} + (\mathbf{R}+\boldsymbol{\delta}_n) &\Rightarrow \mathbf{I}+\boldsymbol{\delta}_n \rightarrow (\mathbf{C}+\boldsymbol{\delta}_m) + (\mathbf{R}+\boldsymbol{\delta}_n) \\ &\Rightarrow \mathbf{I}+\boldsymbol{\delta}_n \rightarrow (\mathbf{C}+\mathbf{R}) + (\boldsymbol{\delta}_n+\boldsymbol{\delta}_m). \end{aligned} \tag{7}$$

It can be observed that the generated $\boldsymbol{\delta}_m$ added in $\mathbf{C}$ serves as distribution prior of $\boldsymbol{\delta}_n$, which facilitates the robust learning of

**Table 1: Evaluation results with both natural and artificial degradation factors on synthesized Cityscapes dataset. Adversarial attack is generated by BIM ($K = 3, 5, 10$), PGD10 and CW, respectively. We report the defense results with perturbation value $\epsilon = 8/255$, and more results for the perturbation $\epsilon = 4/255$ can be found in the supplementary materials.**

| Methods | Rain+BIM3 | | | Rain+BIM5 | | | Rain+BIM10 | | | Rain+PGD10 | | | Rain+CW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mIoU | allAcc | PSNR | mIoU | allAcc | PSNR | mIoU | allAcc | PSNR | mIoU | allAcc | PSNR | mIoU | allAcc | PSNR |
| Seg | 2.81 | 31.36 | 17.44 | 2.46 | 27.61 | 17.39 | 2.41 | 27.40 | 17.39 | 2.42 | 27.87 | 17.39 | 1.80 | 21.74 | 17.41 |
| Robust Seg | 2.16 | 38.32 | 17.37 | 2.08 | 38.02 | 17.28 | 2.06 | 37.93 | 17.23 | 2.05 | 37.91 | 17.22 | 2.09 | 38.04 | 17.29 |
| Derain + Seg | 9.31 | 38.28 | 29.46 | 3.79 | 20.02 | 28.83 | 1.90 | 13.56 | 28.24 | 1.92 | 12.84 | 28.25 | 3.12 | 15.13 | 28.83 |
| NAT | 38.39 | 85.03 | 29.78 | 34.31 | 82.00 | 28.80 | 31.37 | 79.24 | 28.08 | 31.31 | 79.10 | 28.08 | 34.57 | 82.12 | 28.68 |
| PEARL(Ours) | 47.81 | 88.80 | **32.62** | 44.70 | 87.10 | **32.31** | 41.03 | 83.86 | **31.86** | 41.69 | 84.44 | **31.88** | 46.16 | 87.12 | **32.30** |
| +AMA(Ours) | **48.55** | **88.81** | 32.56 | **46.14** | **87.55** | 32.21 | **43.75** | **85.95** | 31.74 | **44.60** | **86.34** | 31.77 | **47.73** | **87.84** | 32.20 |

derain model in Eq. (6) based on its original decomposition function. Meanwhile, since AMA introduces the information of adversarial attack on segmentation model to the ground truth of derain model, the output results will consistently benefit the segmentation tasks to some extent.

In comparison with the NAT framework in Fig. 2 (a), which forms two cycles by alternatively optimizing the attack generator and segmentation model with $\mathcal{L}_{atk}$ and $\mathcal{L}_{def}$, our complete pipeline with both PEARL framework and AMA generator creates a new cycle by introducing AMA to the optimization of deraining model in Fig. 2 (c). Besides, we also analyze the difference of training strategies between NAT frameworks and our PEARL with AMA to help understand how to update the attack and defense module in Fig. 2 (d). In the next section, we will demonstrate the significant performance improvement and generalization performance of this new framework with different quantitative and qualitative metrics on derain and segmentation tasks.

# 4 EXPERIMENTS

## 4.1 Experimental Settings

***Degradation factors and Metrics***. For natural degradation factor (i.e. rain streaks), we manually synthesize rain streaks based on original Cityscapes and VOC dataset( the initial PSNR and SSIM are 17.45 / 0.5566). For artificially generated degradation factor (i.e. adversarial attack), we use BIM attack for training, while BIM, PGD, CW are used for testing the defense performance ($\epsilon = 4/255$, $8/255$). As for the metrics, two type of pixel-wise Accuracy (overall accuracy allAcc and mean of class-wise accuracy mAcc) and mean of class-wise Intersection over Union (mIoU) are used to evaluate the performance of segmentation, which also reflects the robustness against different degradation factors. In addition, PSNR and SSIM are used for the low-level restoration tasks. More implementation details could be found in the supplementary materials.

***Dataset and Model***. We implement our experiments based on two popular semantic segmentation datasets, including Cityscapes [9] and PASCAL VOC 2012 [13]. In the following, we train the model based on the training dataset of Cityscapes, while both datasets are used for testing to verify the performance improvement and generalization ability of the proposed framework. Here we employ two widely used models, i.e., PSPNet [61] and DeepLabv3 [7] for the downstream segmentation task. ResNet50 [20] is adopted as backbone feature extractor, and we follow the default setting of model configuration for training and testing. As for the derain models, we implement four mainstream deraining models, TransWeather [44], MPRNet [55], PReNet [38] and RESCAN [31] to verify the consistent performance of PEARL framework and its insensitivity to the architecture of derain model.

## 4.2 Experimental Results

We first evaluate the performance of different strategies when both rain streaks and adversarial perturbation exists in the segmentation input. Generally speaking, we consider several basic strategies and our proposed framework to address this challenging task. We use Seg, Robust Seg and Derain + Seg to represent the basic model trained with clean image and two models for only handling rain streaks or adversarial examples. Meanwhile, we test the performance of our proposed NAT framework, PEARL framework and PEARL with AMA generator (denoted as +AMA).

In Tab. 1, we consider BIM ($K = 3, 5, 10$), PGD and C&W attack constrained by $\ell_{\text{inf}}$ norm together with the rain streaks to attack the segmentaion task on synthesized Cityscapes dataset. As it can be observed, both degradation factors could incur a sharp decline in the performance of downstream segmentation task. When the attack intensity is weak ($\epsilon = 4/255$, the results can be found in the supplementary material), embedding the pretrained derain model may help protect the segmentation tasks to some extent. Whereas, once the attack intensity increases to $8/255$, the deraining model with little attack prior will also be affected by the perturbation, which causes serious performance decrease. Besides, as the adversarial robustness of the segmentation model improves, the perturbation generated by the same attack method also becomes stronger, which can be reflected in the decline of PSNR.

In contrast, the three proposed solutions, which take into account both factors, significantly promote the defense capability of segmentation tasks. Among these three solutions, PEARL framework (with AMA) gains much more improvement on both derain and segmentation metrics. Under a relatively weak attack ($\epsilon = 4/255$), the effectiveness of AMA can not be clearly verified. As the intensity of adversarial attack increases ($\epsilon = 8/255$), with a slight trade-off on deraining performance (0.1 decrease of PSNR), AMA enables better

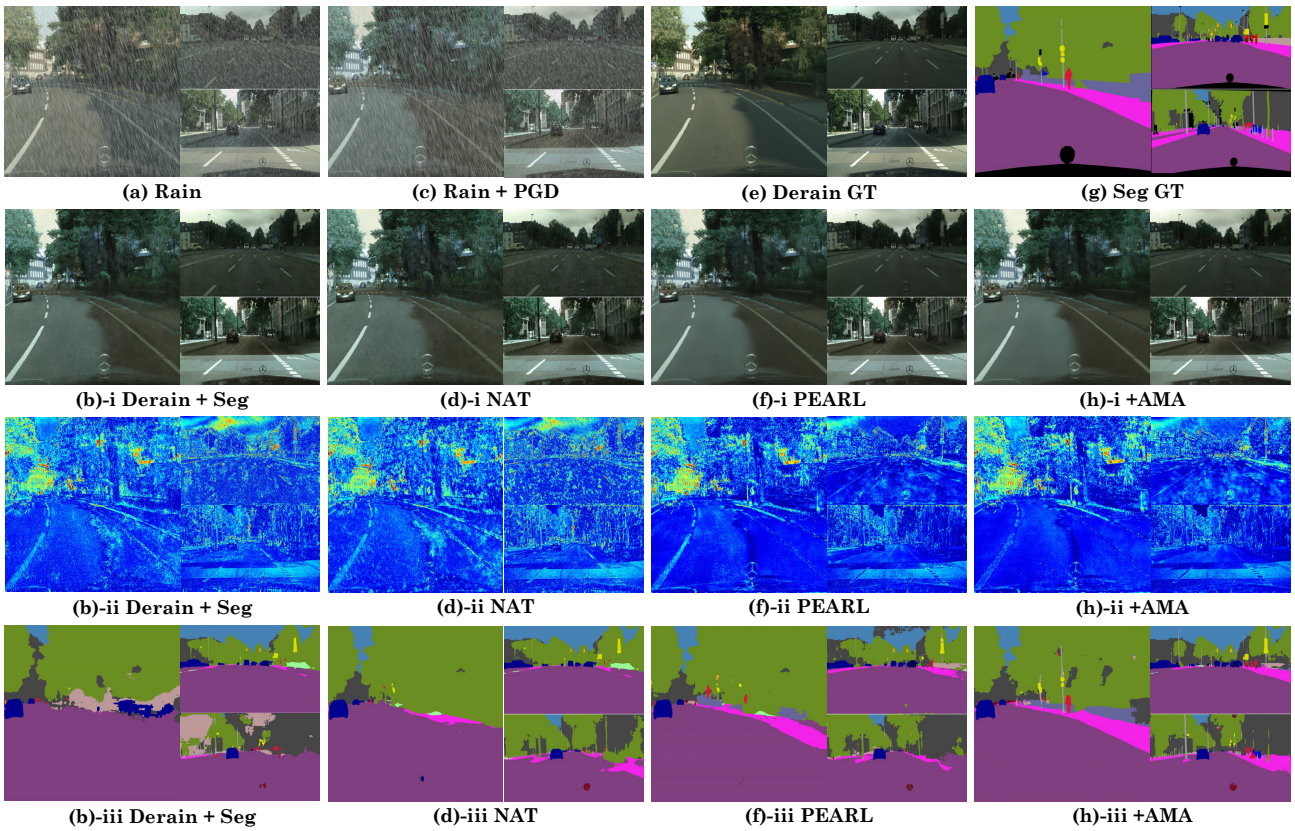



**Figure 4: Comparison of the deraining and segmentation results among different methods on synthesized Cityscapes dataset. The second to fourth rows of images represent the deraining results, heat map of the difference between the derained image and clean image, and the segmentation labels.**

performance of downstream segmentation task on both mIoU and allAcc metrics. It is also worth noting that for unseen attacks (PGD and CW attack), PEARL framework together with AMA assistance still maintains a stable defense effect.

Furthermore, we also show the visualization results in Fig. 4 to demonstrate that our PEARL framework helps obtain higher quality derained images and effectively facilitates the downstream segmentation tasks to defend two degradation factors, which leads to better segmentation results. From the processed heat maps in the third row, it can be clearly seen that the output of deraining model trained by PEARL left much less noise than other solutions, which also demonstrates the effectiveness of PEARL framework to obtain derain images with better visual effects.

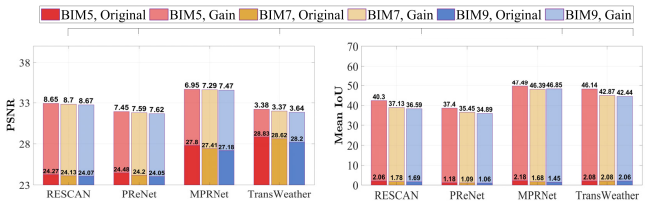



**Figure 5: We illustrate the performance improvement of our PEARL Framework on PSNR and SSIM based on different attack intensities ($K = 5, 7, 9$) and derain models, including RESCAN, PReNet, MPRNet and Transweather.**

Then we adopt four state-of-the-art deraining methods to verify the insensitivity of PEARL framework to the architecture of derain models, and the results are shown in Fig. 5. We train these four models with the same strategy of PEARL with AMA in Fig. 2 (d). It can be seen that our framework can not only improve the PSNR of these methods under different intensities of attack factors, but also significantly improve the downstream segmentation tasks to a large margin.

Finally, we fix the trained deraining model and replace the PSPNet model with DeepLabv3 to show the generalization performance of derain model trained with PEARL to handle the adversarial attacks for the segmentation model. The results under different attacks and the segmentation results of different classes are shown in Table 2 and Figure 6 respectively. It can be seen that in the face of new downstream architecture, except for the unknown attack (CW), the deraining model trained by the PEARL framework and the AMA assistant can achieve a defense effect so close to the original PSPNet model.

## 4.3 Ablation Study

In essence, the motivation of PEARL framework together with AMA is to protect the downstream segmentation tasks from the impact of both natural degradation factor and artificially generated degradation factors. Here we conduct ablation experiments to analyze

the practical effect of our framework to defend these degradation factor separately.

**Table 2: Reporting the defense performance of NAT, PEARL, and PEARL with AMA on the synthesized Cityscapes dataset.**

| Methods | Rain | | Rain + BIM | | Rain + PGD | | Rain + CW | |
|---|---|---|---|---|---|---|---|---|
| | mIoU | PSNR | mIoU | PSNR | mIoU | PSNR | mIoU | PSNR |
| NAT | 51.94 | 31.38 | 33.58 | 28.03 | 43.92 | 30.15 | 30.89 | 27.59 |
| PEARL | **58.39** | 33.08 | 43.70 | 30.16 | 52.35 | **32.77** | 38.02 | **31.90** |
| +AMA | 57.86 | **33.12** | **52.51** | **32.77** | **53.68** | 32.74 | **41.24** | 31.76 |

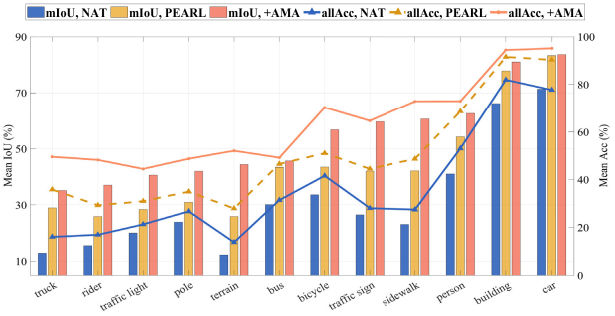



**Figure 6: Illustrating the mIoU and allAcc of different classes for NAT, PEARL and Pearl with AMA based on DeepLabv3.**

Specifically, we first validate the deraining model trained by our framework on images with only rain streaks in Tab. 3. It can be observed that the trade-off between accuracy on clean data and the robustness to defend adversarial attacks also influences the performance of derain model trained by PEARL and PEARL with AMA. When only the rain streaks exist in the input, the model trained by our framework also gains worse performance on these images without adversarial perturbation. But we are also surprised to find that the derain performance are further improved as extra bonus to obtain better visualization results.

**Table 3: Results of different metrics with single degradation factor, i.e. rain streak.**

| Methods | mIoU | mAcc | allAcc | PSNR | SSIM |
|---|---|---|---|---|---|
| Derain+Seg | **37.52** | **39.64** | **97.00** | 31.41 | 90.87 |
| PEARL | 31.96 | 36.37 | 96.01 | **33.13** | 92.69 |
| +AMA | 31.79 | 36.10 | 96.53 | 33.06 | **93.05** |

As for Fig. 7, it comes to a conclusion that our framework indeed enables the deraining model to obtain the ability to eliminate adversarial perturbation under different attack intensity to a great extent, and AMA also further improved the segmentation results and quality of restored images when the rain no longer exists.

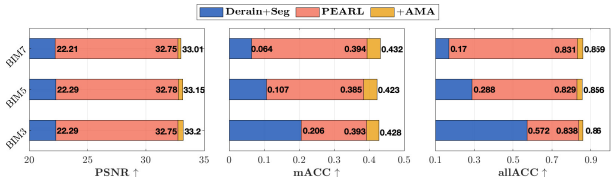



**Figure 7: Illustrating the evaluation results of Derain + Seg, PEARL and PEARL with AMA under different attack intensities of BIM ($K = 3, 5, 7$).**

## 4.4 Extension

Last but not least, we also validate the generalization performance of the proposed framework across different datasets. Specifically, we transfer the deraining model in Table 1 (trained on PSPNet and Cityscapes by PEARL framework) directly to PASCAL VOC dataset (the segmentation modeltrained on PASCAL VOC are also employed), and report the results in Table 4. It can be seen that even in the face of unseen input data distribution and new downstream segmentation model, PEARL framework with the assistance of AMA, can still obtain significant performance in comparison with NAT.

**Table 4: Results of the defense performance on PASCAL VOC dataset. The derain model is the same as the one used in Table 1, while the segmentation model was replaced.**

| Methods | Rain | | Rain + BIM | | Rain + PGD | | Rain + CW | |
|---|---|---|---|---|---|---|---|---|
| | mIoU | Acc | mIoU | Acc | mIoU | Acc | mIoU | Acc |
| NAT | 53.05 | **79.71** | 36.59 | 66.62 | 36.28 | 66.03 | 36.28 | 65.91 |
| PEARL | **58.41** | 72.84 | 43.16 | 73.50 | 43.32 | 73.89 | 43.41 | 73.86 |
| +AMA | 58.38 | 72.89 | **43.54** | **73.95** | **43.88** | **74.42** | **43.87** | **74.22** |

## 5 CONCLUSION

In this paper, we have addressed the robustness of semantic segmentation tasks in a general application scenario where the input image is affected by both natural degradation factors (i.e., rain streaks) and artificially generated degradation factors (i.e., adversarial attacks). Based on the unified understanding of the above degradation factors and analysis of proposed NAT framework, we introduced the PEARL framework, which leverages the adversarial robustness by transferring it to the derain model to simultaneously eliminate the influence of both rain streaks and adversarial perturbation. Moreover, we introduced the AMA generator to the PEARL framework, which provides positive information prior as opposed to the NAA generator. We have shown the significant improvement of the PEARL framework for handling both degradation factors based on different derain and segmentation models. Furthermore, we have verified the generalization performance of the PEARL framework with AMA across different datasets.

## ACKNOWLEDGEMETS

This work is partially supported by the National Key R&D Program of China (No. 2022YFA1004101), the National Natural Science Foundation of China (No. U22B2052).

# REFERENCES


[1] Shashank Agnihotri and Margret Keuper. 2023. CosPGD: a unified white-box adversarial attack for pixel-wise prediction tasks. *arXiv preprint arXiv:2302.02213* (2023).

[2] Anurag Arnab, Ondrej Miksik, and Philip HS Torr. 2018. On the robustness of semantic segmentation models to adversarial attacks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 888–897.

[3] Anish Athalye and Nicholas Carlini. 2018. On the robustness of the cvpr 2018 white-box adversarial example defenses. *arXiv preprint arXiv:1804.03286* (2018).

[4] Andreas Bar, Fabian Huger, Peter Schlicht, and Tim Fingscheidt. 2019. On the robustness of redundant teacher-student frameworks for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 0–0.

[5] Nicholas Carlini and David Wagner. 2017. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*. Ieee, 39–57.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 4 (2017), 834–848.

[7] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).

[8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*. 801–818.

[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3213–3223.

[10] Francesco Croce and Matthias Hein. 2020. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *International conference on machine learning*. PMLR, 2206–2216.

[11] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. 2020. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14560–14569.

[12] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. 2018. Boosting adversarial attacks with momentum. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9185–9193.

[13] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88 (2010), 303–338.

[14] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. 2017. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3855–3863.

[15] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. 2019. Lightweight pyramid networks for image deraining. *IEEE transactions on neural networks and learning systems* 31, 6 (2019), 1794–1807.

[16] Santiago González Izard, Ramiro Sánchez Torres, Oscar Alonso Plaza, Juan Antonio Juanes Mendez, and Francisco José García-Peñalvo. 2020. Nextmed: automatic imaging segmentation, 3D reconstruction, and 3D model visualization platform using augmented and virtual reality. *Sensors* 20, 10 (2020), 2962.

[17] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[18] Jindong Gu, Hengshuang Zhao, Volker Tresp, and Philip HS Torr. 2022. SegPGD: An Effective and Efficient Adversarial Attack for Evaluating and Boosting Segmentation Robustness. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIX*. Springer, 308–325.

[19] Shixiang Gu and Luca Rigazio. 2014. Towards deep neural network architectures robust to adversarial examples. *arXiv preprint arXiv:1412.5068* (2014).

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[21] Jan Hendrik Metzen, Mummadi Chaithanya Kumar, Thomas Brox, and Volker Fischer. 2017. Universal adversarial perturbations against semantic image segmentation. In *Proceedings of the IEEE international conference on computer vision*. 2755–2764.

[22] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. 2019. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 8022–8031.

[23] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. 2020. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8346–8355.

[24] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. 2011. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing* 21, 4 (2011), 1742–1755.

[25] Taeheon Kim, Youngjoon Yu, and Yong Man Ro. 2022. Defending Physical Adversarial Attack on Object Detection via Adversarial Patch-Feature Energy. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1905–1913.

[26] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. 2016. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236* (2016).

[27] Alexey Kurakin, Ian J Goodfellow, and Samy Bengio. 2018. Adversarial examples in the physical world. In *Artificial intelligence safety and security*. Chapman and Hall/CRC, 99–112.

[28] Guanbin Li, Yuan Xie, Liang Lin, and Yizhou Yu. 2017. Instance-level salient object segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2386–2395.

[29] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1633–1642.

[30] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. 2019. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3838–3847.

[31] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*. 254–269.

[32] Fangzhou Liao, Ming Liang, Yinpeng Dong, Tianyu Pang, Xiaolin Hu, and Jun Zhu. 2018. Defense against adversarial attacks using high-level representation guided denoiser. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1778–1787.

[33] Xiaofeng Liu, Yuzhuo Han, Song Bai, Yi Ge, Tianxing Wang, Xu Han, Site Li, Jane You, and Jun Lu. 2020. Importance-aware semantic segmentation in self-driving with discrete wasserstein training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 11629–11636.

[34] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2017. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083* (2017).

[35] Chengzhi Mao, Amogh Gupta, Vikram Nitin, Baishakhi Ray, Shuran Song, Junfeng Yang, and Carl Vondrick. 2020. Multitask learning strengthens adversarial robustness. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 158–174.

[36] Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z Berkay Celik, and Ananthram Swami. 2017. Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia conference on computer and communications security*. 506–519.

[37] Aaditya Prakash, Nick Moran, Solomon Garber, Antonella DiLillo, and James Storer. 2018. Deflecting adversarial attacks with pixel deflection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8571–8580.

[38] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3937–3946.

[39] Abhinav Sagar and RajKumar Soundrapandiyan. 2021. Semantic segmentation with multi scale spatial attention for self driving cars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2650–2656.

[40] Chuanbiao Song, Kun He, Liwei Wang, and John E Hopcroft. 2018. Improving the generalization of adversarial training with domain adaptation. *arXiv preprint arXiv:1810.00740* (2018).

[41] Yang Song, Taesup Kim, Sebastian Nowozin, Stefano Ermon, and Nate Kushman. 2017. Pixeldefend: Leveraging generative models to understand and defend against adversarial examples. *arXiv preprint arXiv:1710.10766* (2017).

[42] Shangquan Sun, Wenqi Ren, Tao Wang, and Xiaochun Cao. 2022. Rethinking Image Restoration for Object Detection. *Advances in Neural Information Processing Systems* 35 (2022), 4461–4474.

[43] Florian Tramèr, Alexey Kurakin, Nicolas Papernot, Ian Goodfellow, Dan Boneh, and Patrick McDaniel. 2017. Ensemble adversarial training: Attacks and defenses. *arXiv preprint arXiv:1705.07204* (2017).

[44] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M Patel. 2022. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2353–2363.

[45] Di Wang, Hao Tang, Jinshan Pan, and Jinhui Tang. 2021. Learning a tree-structured channel-wise refinement network for efficient image deraining. In *2021 IEEE international Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[46] Yuxuan Wang, Jiakai Wang, Zixin Yin, Ruihao Gong, Jingyi Wang, Aishan Liu, and Xianglong Liu. 2022. Generating transferable adversarial examples against vision transformers. In *Proceedings of the 30th ACM International Conference on Multimedia*. 5181–5190.

[47] Eric Wong, Leslie Rice, and J Zico Kolter. 2020. Fast is better than free: Revisiting adversarial training. *arXiv preprint arXiv:2001.03994* (2020).
[48] Chaowei Xiao, Ruizhi Deng, Bo Li, Fisher Yu, Mingyan Liu, and Dawn Song. 2018. Characterizing adversarial examples based on spatial consistency information for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 217–234.
[49] Cihang Xie, Jianyu Wang, Zhishuai Zhang, Yuyin Zhou, Lingxi Xie, and Alan Yuille. 2017. Adversarial examples for semantic segmentation and object detection. In *Proceedings of the IEEE international conference on computer vision*. 1369–1378.
[50] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems* 34 (2021), 12077–12090.
[51] Ke Xu, Xin Tian, Xin Yang, Baocai Yin, and Rynson WH Lau. 2021. Intensity-aware single-image deraining with semantic and color regularization. *IEEE Transactions on Image Processing* 30 (2021), 8497–8509.
[52] Xiaogang Xu, Hengshuang Zhao, and Jiaya Jia. 2021. Dynamic divide-and-conquer adversarial training for robust semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7486–7495.
[53] Maoke Yang, Kun Yu, Chi Zhang, Zhiwei Li, and Kuiyuan Yang. 2018. Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3684–3692.
[54] Changqian Yu, Jingbo Wang, Changxin Gao, Gang Yu, Chunhua Shen, and Nong Sang. 2020. Context prior for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12416–12425.
[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14821–14831.
[56] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. 2018. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 7151–7160.
[57] He Zhang and Vishal M Patel. 2018. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 695–704.
[58] Jiaming Zhang, Qi Yi, and Jitao Sang. 2022. Towards Adversarial Attack on Vision-Language Pre-training Models. In *Proceedings of the 30th ACM International Conference on Multimedia*. 5005–5013.
[59] Kaihao Zhang, Wenhan Luo, Wenqi Ren, Jingwen Wang, Fang Zhao, Lin Ma, and Hongdong Li. 2020. Beyond monocular deraining: Stereo image deraining via semantic understanding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*. Springer, 71–89.
[60] Yihua Zhang, Guanhua Zhang, Prashant Khanduri, Mingyi Hong, Shiyu Chang, and Sijia Liu. 2022. Revisiting and advancing fast adversarial training through the lens of bi-level optimization. In *International Conference on Machine Learning*. PMLR, 26693–26712.
[61] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2881–2890.
[62] Quan Zhou, Yu Wang, Yawen Fan, Xiaofu Wu, Suofei Zhang, Bin Kang, and Longin Jan Latecki. 2020. AGLNet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network. *applied soft computing* 96 (2020), 106682.