TPSeNCE: Towards Artifact-Free Realistic Rain Generation for Deraining and Object Detection in Rain

Shen Zheng[1], Changjie Lu[2], Srinivasa G. Narasimhan[1]
[1]Carnegie Mellon University, [2]University of Illinois Urbana-Champaign
{shenzhen, srinivas}@andrew.cmu.edu, cl140@illinois.edu

Abstract

Rain generation algorithms have the potential to improve the generalization of deraining methods and scene understanding in rainy conditions. However, in practice, they produce artifacts and distortions and struggle to control the amount of rain generated due to a lack of proper constraints. In this paper, we propose an unpaired image-to-image translation framework for generating realistic rainy images. We first introduce a Triangular Probability Similarity (TPS) constraint to guide the generated images toward clear and rainy images in the discriminator manifold, thereby minimizing artifacts and distortions during rain generation. Unlike conventional contrastive learning approaches, which indiscriminately push negative samples away from the anchors, we propose a Semantic Noise Contrastive Estimation (SeNCE) strategy and reassess the pushing force of negative samples based on the semantic similarity between the clear and the rainy images and the feature similarity between the anchor and the negative samples. Experiments demonstrate realistic rain generation with minimal artifacts and distortions, which benefits image deraining and object detection in rain. Furthermore, the method can be used to generate realistic snowy and night images, underscoring its potential for broader applicability. Code is available at https://github.com/ShenZheng2000/TPSeNCE.

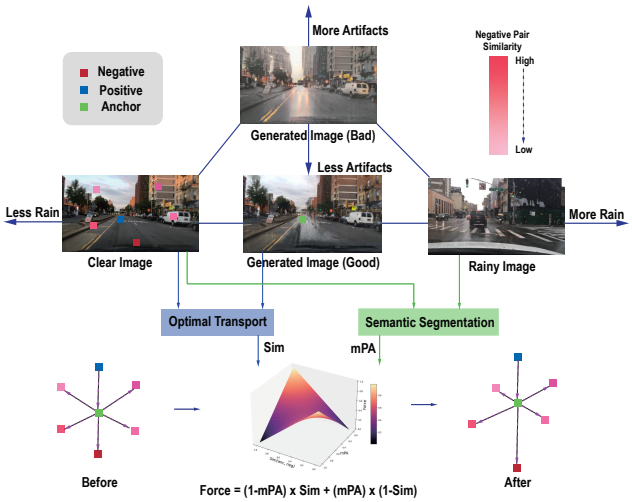

Figure 1: **Illustration for the proposed Triangular Probability Similarity (TPS) and Semantic Noise Contrastive Estimation (SeNCE) for rain generation**. TPS drives the generated rainy image towards the clear and the rainy image in the discriminator manifold to suppress artifacts and distortions. SeNCE collaboratively re-weights the pushing force of the negative patches based on their feature similarity with the anchor patch and refines that force with the semantic similarity between the clear and rainy image.

1. Introduction

Rain is a common bad weather condition that can significantly impair the quality of images and videos. Rain streaks, especially during heavy rain, obscure scene details and textures. Raindrops create a layer of water droplets on windshields, making objects appear blurry and distorted. Shiny wet roads create object reflections. Rain mist scatters ambient light, reducing the visibility of distant objects [42]. These visual manifestations of rain not only impair the perceptual quality of images but also pose challenges for scene understanding algorithms like object detection, which are typically trained using data captured under clear weather conditions [13].

One common way to improve object detection in the rain is to apply deraining (i.e. rain removal) as a preprocessing step. Ideally, deraining algorithms should remove rain from images before applying object detection models. However, most state-of-the-art deraining methods [5, 10, 25, 30, 37, 41, 47] rely on supervised training with paired synthetic clear/rainy images due to the intractability of obtaining real paired clear/rainy images. Unfortunately, these methods do not generalize well to real-world rainy images [42] because of the large domain gap between synthetic and natural rainy images. Although some deraining methods [16, 39, 43] use unpaired real clear/rainy images for unsupervised learning to improve generalization, it is challenging to integrate knowledge from the supervised and un-

supervised branches seamlessly to enhance deraining performance on real data [36].

Another approach to enhance object detection in rainy conditions is using rain generation techniques to create synthetic rainy images for training object detectors. However, traditional model-based rain generation approaches such as those presented in [6, 26, 32] rely on oversimplified assumptions and hand-crafted priors, which fail to accurately model the diverse types of real rain. In contrast, data-driven deep learning approaches such as unpaired image-to-image translation methods like UNIT [23] have demonstrated their ability to translate images across different weather conditions. Nevertheless, these methods often produce artifacts and distortions while generating rain due to the lack of proper constraints. Additionally, controlling the amount of rain produced is challenging, as generating too much rain can lead to overlapping of the background and feature loss while generating too little rain results in an unrealistic-looking image. The presence of unwanted artifacts, distortions, and uncontrollable rain amounts can decrease the perceptual quality and hinder detection algorithms.

In this paper, we address the above issues of rain generation approaches and propose an unpaired image-to-image translation framework for rain generation. Our analysis of the output matrix from the discriminator reveals a triangular relationship among the clear image, the generated rainy image, and the real rainy image on the discriminator manifold (Fig. 3). We observe that the generated rainy images with fewer artifacts and distortions are closer to the line segment connecting the clear and rainy images. Based on this observation, we propose a Triangular Probability Similarity (TPS) loss to guide the generated rainy images toward the real and clear images, thereby minimizing the artifacts and distortions. We then revisit the contrastive learning strategy of CUT [27] and find that the amount of rain generated can be controlled by regulating the pushing force of contrastive learning. For this, we propose a Semantic Noise Contrastive Estimation strategy (SeNCE) that reweights the pushing force of negative pairs based on the similarity between the negatives and anchor, and the mean Pixel Accuracy (mPA) (shown in Fig. 5) between the semantic segmentation maps of the clear and rainy images.

We evaluate the proposed method against multiple image-to-image translation approaches on various driving datasets including BDD100K [44], INIT [34], and Boreas [2]. The evaluation on the BDD100K encompasses image-to-image translation, image deraining, and object detection. whereas the evaluation on INIT and Boreas focus solely on image-to-image translation.

In summary, we present an unpaired image-to-image translation framework for generating realistic rainy images, with the following technical contributions:

- We introduce a Triangular Probability Similarity (TPS) loss to minimize the artifacts and distortions during rain generation.
- We propose a Semantic Noise Contrastive Estimation (SeNCE) strategy to regularize the contrastive learning force to optimize the amounts of generated rain.
- Our evaluation highlights the benefits of realistic rainy image generation for real rain removal and object detection in real rainy conditions.

Beyond rain, our method can be used to generate realistic snowy images and night images, underscoring its potential for a broader applicability.

2. Related Works

2.1. Unpaired Image-to-Image Translation

Unpaired image-to-image (I2I) methods aim to map between two domains of unpaired images. The main challenge is preserving source content while adopting the target style. CycleGAN [50] introduced cycle-consistency loss to ensure content is kept by minimizing post-translation differences. UNIT [23] built on this with the shared latent space assumption, suggesting that different domain images map to one latent code. However, these methods can hinder output style diversity. MUNIT [17] and DRIT [20] then disentangle images into domain-invariant content and domain-specific style codes, facilitating multimodal I2I translation.

Current I2I networks are designed for simple and common benchmarks, such as horse2zebra or apple2orange. When applied to more intricate tasks like rain generation, these general and weak constraints under unstable GAN training can misrepresent the intricate details, leading to unwanted artifacts and visual distortions.

2.2. Contrastive Learning

Contrastive learning aims to attract positive instances towards an anchor while repelling negative ones. CUT [27] was a pioneer to apply this to unpaired image-to-image translation by maximizing mutual information for random patches. However, CUT doesn't differentiate negative samples based on their similarity to the anchor. In response, QS-Attn [15] uses a query-selected attention module to pick critical negative patches. Meanwhile, NEGCUT [38] developed an instance-aware generator for hard negative examples, whereas MoNCE [45] employs an optimal-transport based approach to adjust the repulsion of negative pairs.

While these methods advance general I2I tasks, they overlook the high-level semantic similarity between source and target images, instead focusing on the image-level feature similarity between the source and the generated image. In rain generation, lacking high-level cues from the target rainy image makes it hard to determine the appropriate rain amount for a realistic output.

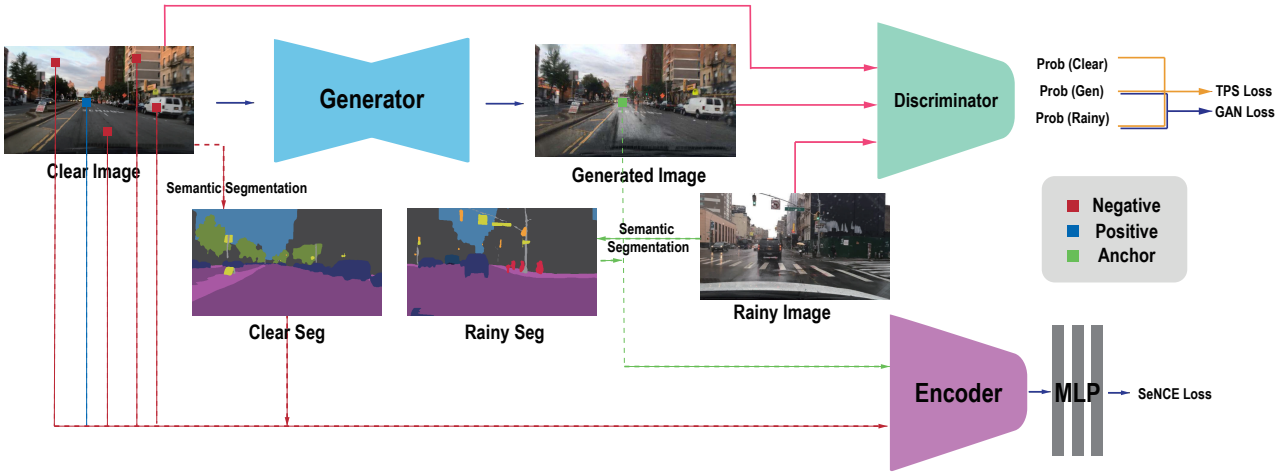

Figure 2: **Workflow for the proposed method**. The generator translates the clear image to generated (rainy) image. The discriminator receives the clear, generated, and rainy images to compute the TPS and GAN (adversarial) losses. Meanwhile, the encoder randomly selects and then embeds one positive patch and multiple negative patches from the clear image and the anchor patches from the generated image. The MLPs then process these embedded patches in a contrastive learning manner and output the SeNCE loss with the guidance of the semantic segmentation maps from the clear and rainy images.

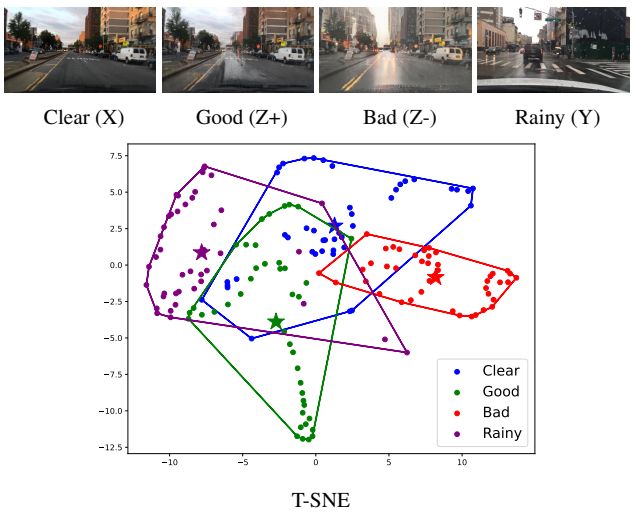

Figure 3: **T-SNE [35] visualization of the output matrices from the discriminator** D. Output matrices from different images are visualized with different colors. The centroid for each group is visualized as a star. Compared with $D(Z-)$ from the 'Bad' generated image $Z-$ (i.e., images with more artifacts), $D(Z+)$ from the 'Good' generated rainy image $Z+$ (i.e., images with fewer artifacts) have more area overlap with $D(X)$ and $D(Y)$. Designing a loss that brings $D(Z)$ near $D(X)$ and $D(Y)$ would benefit artifacts reduction, resulting in high-quality generated rainy images.

Figure 4: **Comparing Point To Line (PTL) with Triangular Probability Similarity (TPS)**: PTL employs unconstrained optimization, pushing the generated image 'Gen' onto the straight line connecting 'Clear' and 'Rainy', frequently ends up on the extension cords. In contrast, TPS uses constrained optimization, adjusting the ellipse level set to position the generated image on the **line segment** connecting 'Clear' and 'Rainy'. Zoom in for better view.

3. Proposed Methods

In this section, we first explain the Triangular Probability Similarity (TPS) loss and revisit the Noise Contrastive Estimation (NCE) [11] schemes of CUT [27] and MoNCE [45]. Extending these NCEs, we derive our Semantic Noise Contrastive Estimation (SeNCE) strategy. Last, we display the loss functions for model training. An overview of the proposed method's workflow is shown in Fig. 2.

3.1. Triangular Probability Similarity (TPS)

Generating realistic rainy images while minimizing artifacts and distortions is a challenging task. Due to the ill-posed nature of rain generation and the instability of GAN training [7], the generated rainy images often suffer from artifacts and distortions.

We show a T-SNE visualization in Fig. 3, which explains the motivation for our Triangular Probability Similarity (TPS) loss. Let X be the clear image, Y be the rainy image, and Z be the generated rainy image. The TPS loss is based on the output representation from the discriminator D. It constrains $D(Z)$ to lie in the space spanned by $D(X)$ and $D(Y)$, ensuring that the generated rainy image follows a similar distribution as the clear image and the real rainy image. This strategy effectively mitigates undesired artifacts and distortions, as the information for the generated images is sourced exclusively from the clear image (providing background) and the real rainy image (providing rain).

A potential issue with calculating TPS based on the distance between $D(Z)$ and the **straight line** connecting the centroid of $D(X)$ and $D(Y)$ is that it may guide $D(Z)$ towards the extension cord of either $D(X)$ or $D(Y)$, leading to *too much or too little rain* in the generated image (as shown in Fig. 1). Additionally, $D(Z)$ may end up too far from $D(X)$ and $D(Y)$, resulting in *artifacts and distortions* in the generated image that do not belong to either X or Y.

As shown in Fig. 4, we address these issues with a loss function based on the distance from $D(Z)$ to the **line segment** between the centroid of $D(X)$ and $D(Y)$. Based on the triangular inequality, we use the following TPS loss:

$$\mathcal{L}_{\text{TPS}}(X,Y,Z) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}(|D(X)_{i,j} - D(Z)_{i,j}| + |D(Y)_{i,j} - D(Z)_{i,j}| - |D(X)_{i,j} - D(Y)_{i,j}|) \tag{1}$$

Where H and W represent the height and width of the probability matrix $D(Z)$, respectively.

3.2. Revisiting NCEs

PatchNCE Patch Noise Contrastive Estimation (PatchNCE) [27] aims to maximize the mutual information between the corresponding input and output patches as below:

$$\mathcal{L}_{\text{PatchNCE}}(X,Z) = -\sum_{i=1}^{N}\log\frac{e^{\frac{x_i\cdot z_i}{\tau}}}{e^{\frac{x_i\cdot z_i}{\tau}} + \sum_{\substack{j=1\\ j\neq i}}^{N} e^{\frac{x_i\cdot z_j}{\tau}}} \tag{2}$$

Where N is the number of patches, $[x_1, x_2, \cdots, x_N]$ and $[z_1, z_2, \cdots, z_N]$ are encoded patch features, and τ is the temperature hyperparameter.

MoNCE The problem with PatchNCE [27] is that it indiscriminately pushes all negative patches from the anchor, leading to sub-optimal performance for tasks with mixed easy and hard negatives patches. Modulated Noise Contrastive Estimation (MoNCE) [45] address this issue by reweighting the pushing force for the negative patches based on their similarity with the anchor as below.

$$\mathcal{L}_{\text{MoNCE}}(X,Z) = -\sum_{i=1}^{N}\log\frac{e^{\frac{x_i\cdot z_i}{\tau}}}{e^{\frac{x_i\cdot z_i}{\tau}} + Q(N-1)\sum_{\substack{j=1\\ j\neq i}}^{N} w_{ij}\cdot e^{\frac{x_i\cdot z_j}{\tau}}} \tag{3}$$

Where Q is a hyperparameter, and $w_{ij}(j \neq i)$ represents a weighting strategy. MoNCE proposes a hard weighting strategy w_{ij}^{+} and an easy weighting strategy w_{ij}^{-} to address unpaired and paired image-to-image translation, respectively. w_{ij}^{+} and w_{ij}^{-} are written as below.

$$w_{ij}^{+} = \text{softmax}\left(\frac{x_i\cdot z_j}{\beta}\right)_j \quad w_{ij}^{-} = \text{softmax}\left(\frac{1 - x_i\cdot z_j}{\beta}\right)_j \tag{4}$$

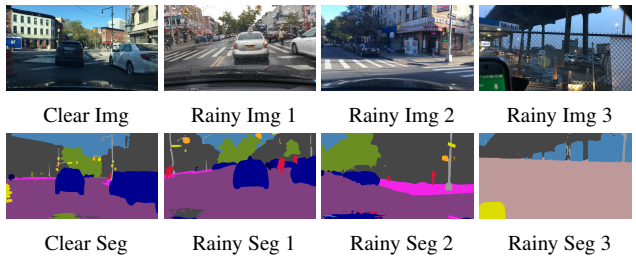

Figure 5: **We prefer mPA and mIoU for segmentation maps over PSNR and SSIM for images in the context of regularizing contrastive learning.** For *Clear image* paired with *Rainy Image 1*, image scores are: [PSNR/SSIM] = [27.945/0.154] and segmentation scores are [mPA/mIoU] = [0.592/0.140]. For *Clear image* paired with *Rainy Image 2*, image scores are: [PSNR/SSIM] = [27.889/0.111] and for segmentation [mPA/mIoU] = [0.466/0.029]. Notably, mPA and mIoU capture semantic similarities in segmentation more effectively, with higher scores for more aligned pairs, whereas PSNR and SSIM show limited variance across scenarios.

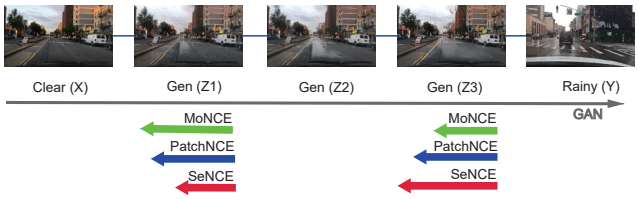

Figure 6: **SeNCE outperforms others in optimizing rain amount to produce realistic rainy images**. The length of the arrow here represents the magnitude of the NCE losses.

3.3. Semantic Noise Contrastive Estimation (SeNCE)

While MoNCE [45] has improved upon PatchNCE [27] on some benchmark image-to-image translation datasets, two issues impair its performance.

First, MoNCE employs distinct weighting strategies for paired and unpaired settings without strong justification. In fact, in unpaired scenarios, images may appear paired, such as the same house from different angles or identical parking lots with different cars. In such cases, w_{ij} should lean towards w_{ij}^{-} rather than w_{ij}^{+}. To sum, a smooth transition between these weighting strategies would be more ideal.

Second, MoNCE uses only image-level information from random patches for reweighting. Yet, in rain generation, many target domain pixels are compromised by droplets, streaks, wetness and mist. Such pixels can't provide precise guidance for contrastive learning [3,4]. Hence, it's crucial to move beyond image-level details and seek a deeper understanding with minimal rain interference.

It is evident from Fig. 5 that semantic-level metrics like mPA and mIoU more accurately capture the similarity between unpaired clear and rainy images than image-level metrics like PSNR and SSIM. This is because semantic-level metrics doesn't depend on perfectly aligned, corruption-free pixels. Even with unaligned images or pixels marred by rain, we can discern their differences using a comprehensive understanding of the segmentation maps. While we occasionally encounter pairs with very low mPA (e.g., 0.162 for clear and rainy img 3), such cases are rare and have minimal impact on training. Thus, we still include them in our training dataset. Below is our formulation for Semantic Noise Contrastive Estimation (SeNCE).

$$\mathcal{L}_{\text{SeNCE}}(X,Y,Z) = -\sum_{i=1}^{N} \log \frac{e^{\frac{x_i \cdot z_i}{\tau}}}{e^{\frac{x_i \cdot z_i}{\tau}} + Q(N-1)\sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} \cdot e^{\frac{x_i \cdot z_j}{\tau}}} \tag{5}$$

Our weight $w_{i,j}$ in the above equation is:

$$w_{ij} = \text{softmax}\left(\frac{F(i,j)}{\beta}\right)_j \tag{6}$$

Where $F(i,j)$ can be expressed as:

$$F(i,j) = (1 - mPA(X,Y))(x_i \cdot z_j) + (mPA(X,Y))(1 - x_i \cdot z_j) \tag{7}$$

$F(i,j)$ represents a semantic-based contrastive learning force derived from mPA. It adjusts between easy weights w_{ij}^{-} and hard weights w_{ij}^{+} of MoNCE. High mPA, indicating semantically similar clear and rainy images, results in a shift towards w_{ij}^{-}. In contrast, low mPA, indicating dissimilar images, leans towards hard weights w_{ij}^{+}.

Variants	TrainA	TestA	TrainB	TestB	Height x Width
BDD100K (clear2rainy)	27,988	4,025	12,798	3,301	720 x 1,280
BDD100K (clear2snowy)			4,025	422	
BDD100K (day2night)			22,884	3,274	
INIT (clear2rainy)	18,112	3,197	3,330	588	1,208 x 1,920 3,000 x 4,000
Boreas (clear2snowy)	8,356	2,089	3,649	913	720 x 860

Table 1: **Training and testing images for the task A2B.**

3.4. Analysis of NCEs

Using the notation $Sim(X,Z) = x_i \cdot z_i$, we analyze the three NCEs with the help of Fig. 6:

- For insufficient rain (Z_1), $Sim(X,Z)$ does not align with $mPA(X,Y)$, and the weight w_{ij} is large based on Eq. 6 and Eq. 7 . According to Eq. 5, this results in a low absolute value for $\mathcal{L}_{\text{SeNCE}}(X,Y,Z)$, allowing the GAN loss to dominate and drive the image towards the desired state (Z_2).
- For excessive rain (Z_3), $Sim(X,Z)$ closely matches $mPA(X,Y)$, and the weight w_{ij} is small. This increases the magnitude of $\mathcal{L}_{\text{SeNCE}}(X,Y,Z)$, enabling SeNCE to overpower the GAN loss to guide the image back towards Z_2.

In essence, SeNCE adjusts the NCE loss based on the comparability of $mPA(X,Y)$ and $Sim(X,Z)$, refining the generated rain. Unlike PatchNCE, which lacks weight adjustment, and MoNCE, which doesn't account for semantic similarity, SeNCE ensures more realistic results.

3.5. Final Objective

The training objective of the proposed method is:

$$\mathcal{L}_{(X,Y)} = \lambda_1 \mathcal{L}_{\text{GAN}}(X,Y) + \lambda_2 \mathcal{L}_{\text{SeNCE}}(X,Y,Z) + \lambda_3 \mathcal{L}_{\text{TPS}}(X,Y,Z) \tag{8}$$

Similar to CUT [27], we set both λ_1 and λ_2 as 1. Since TPS is an auxiliary loss [49], we set λ_3 as 0.1.

4. Experiments

4.1. Implementation Details and Metrics

Implementation Details: We trained our model with PyTorch [28] on 8 RTX 3090 GPUs with an Intel Xeon Gold 6330 Processor CPU, using the Adam optimizer [19]. The model was trained for 200 epochs, with a learning rate of $2e^{-4}$ for the first 100 epochs, decreased to $2e^{-5}$ for the remaining epochs. Training utilized randomly cropped

Model	Constraints		NCEs				Scores			
	PTL	TPS	Patch-NCE	Mo-NCE	SeNCE (mIoU)	SeNCE (mPA)	Content↑	Style↑	KID↓	FID↓
M1			✓				3.32	3.38	85.29	21.90
M2	✓		✓				2.93	2.83	88.15	22.12
M3		✓	✓				3.51	3.58	**70.93**	20.90
M4				✓			3.10	3.40	75.66	**18.60**
M5						✓	3.52	3.56	74.20	20.64
M6		✓			✓		3.27	3.33	80.37	21.12
M7		✓				✓	**3.58**	**3.70**	72.19	19.34

Table 2: **Quantitative ablation for rain generation on BDD100K dataset.** The best scores are in **bold**, and the second best scores are in blue.

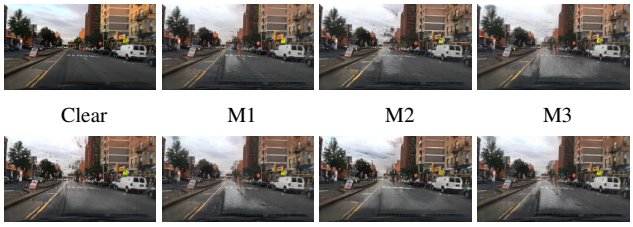

Figure 7: **Qualitative ablation for rain generation on BDD100K dataset.**

Methods	BDD100K Dataset (clear → rainy)				INIT Dataset (clear → rainy)			
	Content↑	Style↑	KID↓	FID↓	Content↑	Style↑	MMD↓	ED↓
UNIT	3.24	3.48	88.85	**18.099**	2.58	2.66	34.231	35.702
MUNIT	2.44	2.80	189.12	26.538	2.80	2.72	34.425	36.458
CUT	3.32	3.38	85.29	21.901	3.16	2.90	33.704	34.777
QS-Attn	3.34	3.58	85.59	21.614	2.46	2.66	33.836	34.853
MoNCE	3.10	3.30	75.66	18.595	2.18	2.24	33.579	34.814
Ours	**3.58**	**3.70**	**72.19**	19.341	**3.42**	**3.04**	**33.535**	**34.774**

Table 3: **Quantitative comparison for image rain generation on the BDD100K and INIT datasets.**

Methods	EffDerain		VRGNet		PreNet		SAPNet			
	DBCNN	MUSIQ	DBCNN	MUSIQ	DBCNN	MUSIQ	DBCNN	MUSIQ	Qual	Perf
Rain100H	39.33	49.51	36.10	49.65	47.34	53.72	46.59	53.27	3.20	2.08
UNIT	36.56	48.73	43.07	51.73	44.16	55.05	45.91	56.04	3.37	3.49
MUNIT	38.54	52.55	37.22	50.21	37.44	54.28	41.92	55.07	3.04	2.20
CUT	36.34	48.82	37.22	50.21	55.00	60.67	54.49	59.97	2.88	3.92
QS-Attn	37.02	49.62	36.93	50.81	34.61	43.02	57.31	61.06	2.94	3.22
MoNCE	37.30	49.63	37.44	50.96	59.49	60.68	37.08	48.95	2.78	2.37
Ours	**39.82**	**54.51**	**47.87**	**54.10**	**60.17**	**61.79**	**57.83**	**61.85**	**3.53**	**4.33**

Table 4: **Deraining comparisons with different deraining methods trained on images from different rain generation methods**. For all metrics in this table, higher is better.

patches of size 256 × 256. All rain generation baselines were trained under that setting for fair comparison.[1]

Evaluation Metrics: We evaluated rain generation models using FID [12] and KID [1] on the BDD100K and Boreas datasets. We employed MMD [9] and ED [31] for the INIT dataset, due to its small size [33]. In all tables, KID values are scaled by 10^{-4} and MMD by 10^{-5}. Additionally, a 50-participant user study rated generated image content and style on a 1-5 scale. For deraining, we train multiple deraining methods [10, 30, 37, 47] on images generated from different rain generation models, evaluated them using MUSIQ [18] and DBCNN [14], and conducted another user study for image quality of the derained images and the performance of the deraining methods. For detection, we used Yolov3 [29], evaluating performance with mAP across different IoU thresholds and object sizes.

[1]The implementation details for benchmark datasets, as well as deraining and detection algorithms, can be found in the supplementary material.

Pretrained	Finetuned	mAP	mAP_50	mAP_75	mAP_s	mAP_m	mAP_l
COCO	None	0.171	0.373	0.134	0.093	0.300	0.409
	Clear	0.231	0.492	0.190	0.151	0.366	0.454
	UNIT	0.215	0.462	0.178	0.113	0.404	0.546
	MUNIT	0.178	0.394	0.147	0.108	0.291	0.438
	CUT	0.243	0.512	0.203	0.149	0.410	0.548
	QS-Attn	0.248	0.513	0.213	0.151	0.401	0.574
	MoNCE	0.246	0.510	0.211	0.151	0.400	0.570
	Ours	**0.262**	**0.526**	**0.237**	**0.158**	**0.419**	**0.646**

Table 5: **Quantitative comparison for Yolov3 object detection on the BDD100K test rainy dataset.** The Yolov3 object detector are pretrained on COCO [22], and finetuned on generated rainy images using different methods. mAP is computed on the most challenging 100 rainy images.

4.2. Ablation Studies

TPS and SeNCE: We perform an ablation study on the TPS and SeNCE modules. As shown in Fig. 7, the inclusion of TPS **(M3)** enhances rain generation by mitigating artifacts and distortions. Additionally, the usage of SeNCE **(M5)** optimizes the contrast and surface water area, leading to more realistic road reflections. Quantitative analysis presented in Tab. 2 further demonstrates the efficacy of both TPS and SeNCE in improving rain generation.

TPS vs. PTL: We examine the effects of replacing TPS with the related modules Point To Line Distance (PTL). As depicted in Fig. 7, the inclusion of PTL **(M2)** results in a degradation of rain generation, evident from numerous artifacts and distortions in the background. Conversely, the adoption of TPS leads to high-quality rain generation. Furthermore, Tab. 2 validates that PTL yields worse scores, while the proposed TPS achieves significantly better scores.

SeNCE vs. Other NCEs: We investigate the effectiveness of the proposed SeNCE against other NCEs, including PatchNCE [27] and MoNCE [45]. Fig. 7 shows that SeNCE **(M5)** leads to high-quality wet surfaces and road reflections, compared with MoNCE **(M4)** and PatchNCE **(M1)**. This is confirmed at Tab. 2, where SeNCE leads to generally better scores than MoNCE or PatchNCE.

Semantic Metrics Selection in SeNCE: We justify the choice of using mPA compared with widely used mIoU. Fig. 7 shows that TPSeNCE with mPA **(M7)** surpass TPSeNCE with mIoU **(M6)** in generating realistic surface water and road reflections. This is agreed by Tab. 2, where mPA leads to much better scores than mIoU.

Figure 8: **Qualitative comparison for clear2rainy (i.e. rain generation) on the BDD100K and INIT dataset.**

Figure 9: **Qualitative comparison for rainy2clear (i.e. deraining) on the BDD100K dataset.** The deraining model SAPNet is trained on images from different rain generation methods.

4.3. Experiment Results

Rain Generation: Tab. 3 shows quantitative comparison for rain generation on BDD100K and INIT, where our method attained the best score on all metrics, except for FID on BDD100K. One possible reason is that FID is insensitive to local variations, such as small regions containing artifacts and distortions [8]. Fig. 8 showcases a qualitative comparison, revealing that our method generates the most realistic rainy images, with a perfect balance between raindrops, wet surfaces, and road reflections. Besides, our method effectively preserves the content of the clear image without introducing noticeable artifacts or distortions.

Deraining: Tab. 4 presents a quantitative comparison of deraining performance on BDD, indicating that our method achieves the highest score on all deraining metrics. Meanwhile, Fig. 9 provides a qualitative comparison, highlighting that our method outperforms other models in removing raindrops and mists while retaining feature details, restoring color balance, and suppressing noise, blur, and artifacts.

Detection: Tab. 5 presents a quantitative comparison of Yolov3 finetuned on images from different rain generation methods. The most challenging 100 images in heavy rain or poor lighting were selected for computing mean average precision (mAP) since most rainy images in BDD100K are captured in light rain and good illumination conditions and cannot distinguish the performance of different models. Our proposed method achieved the highest mAP scores across all IoU thresholds and object sizes. Also, as shown in Fig. 10, our method outperforms others in detecting objects under heavy raindrops, with the smallest number of false positives and false negatives across multiple object classes, such as cars, traffic lights, and stop signs.

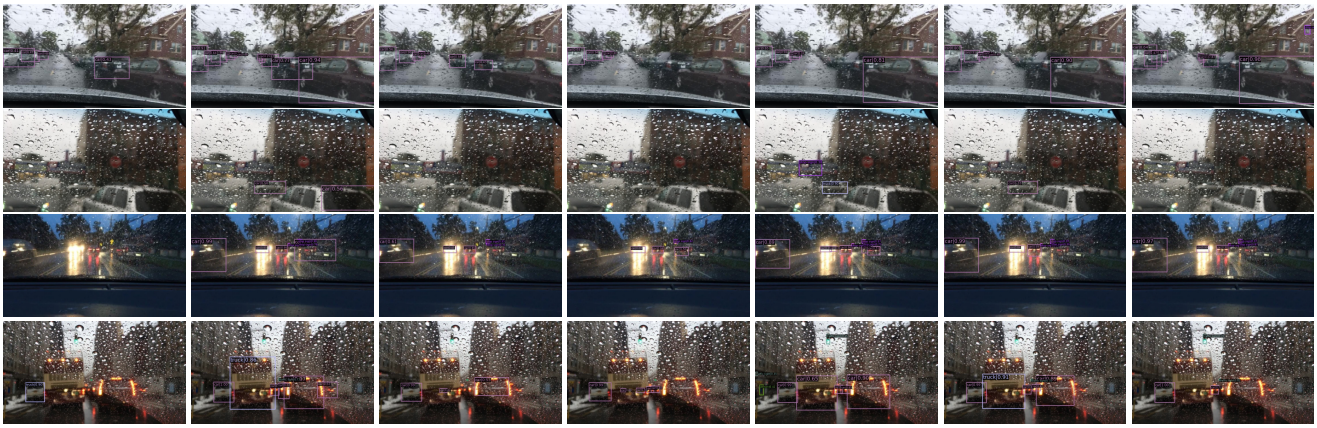

Figure 10: **Qualitative comparison for Yolov3 object detection on the BDD100K dataset.** Yolov3 was pretrained on COCO, finetuned on generated rainy images from different rain generation methods, and tested on BDD100K test rainy.

Methods	BDD100K Dataset (day → night)				Boreas Dataset (clear → snowy)			
	Content↑	Style↑	KID↓	FID↓	Content↑	Style↑	KID↓	FID↓
CUT	3.10	3.73	147.26	16.522	2.80	1.87	170.34	36.98
QS-Attn	3.13	3.37	158.83	17.544	3.50	3.70	142.96	35.42
MoNCE	2.93	3.30	142.97	17.003	3.37	3.57	158.30	34.62
Ours	**3.47**	**4.03**	**142.10**	**15.901**	**4.17**	**4.37**	**144.40**	**33.83**

Table 6: **Quantitative comparison of day2night translation on BDD100K and clear2snowy on Boreas datasets.**

4.4. Extension to Snowy and Night

Beyond rain generation, our method works for **Clear2Snowy** and **Day2Night** translation. We show qualitative comparison for snow generation in Fig. 11. Our method effectively produces authentic snowy road surfaces with realistic contrast, lighting, reflections, and textures. On the other hand, CUT and QS-Attn produce insufficient snow, while MoNCE's snow lacks realism due to minimal surface reflections and limited contrast variations. Moreover, in Tab. 6, our method achieves the best scores for day2night and clear2snowy translation. [2]

4.5. Limitations and Future Work

Our method is trained on benchmark datasets that compromise mostly of mild rainy images with weak light sources. Consequently, it cannot address extremely heavy rain images effectively. Addressing heavy rain is crucial for improving object detection, given the significant challenges it poses, such as occlusion, reflection, motion blur, low contrast, and noise [42, 46]. In the future, we intend to gather a large dataset of heavy rain images featuring strong light sources, use physics-based models [21, 24] for training, and explore joint preprocessing and finetuning [48].

[2]More day2night, clear2snowy translation results are in supplementary.

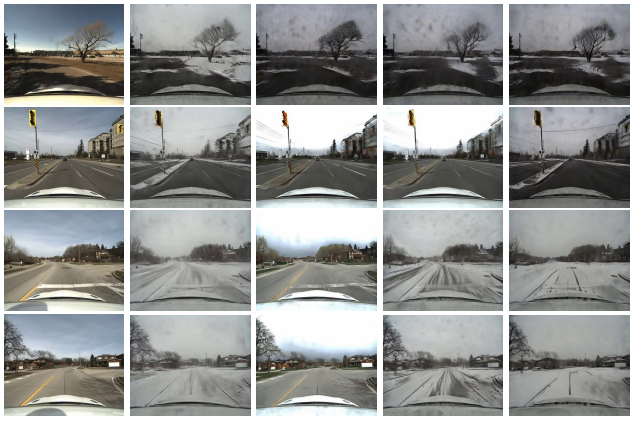

Figure 11: **Qualitative comparison for clear2snowy translation (i.e. snow generation) on Boreas dataset.**

5. Conclusion

This paper presented a novel unpaired image-to-image translation framework that generates highly realistic rainy images with minimal artifacts. Our approach utilizes Triangular Probability Similarity (TPS) and Semantic Noise Contrastive Estimation (SeNCE) to minimize artifacts and distortions and optimize the amount of generated rain. Experiments demonstrate that the proposed method outperforms state-of-the-art in generating realistic rainy images with minimal artifacts, which can benefit image deraining and object detection in rain. Additionally, our method can generate high-quality snowy and night images, highlighting its capability for diverse weather and lighting conditions.

Acknowledgements: This work was supported in part by an NSF CPS Grant CNS-2038612, a DOT RITA Mobility-21 Grant 69A3551747111 and by General Motors Israel.

References

[1] Mikołaj Bińkowski, Danica J Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*, 2018. 6

[2] Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023. 2

[3] Minghui Chen, Zhiqiang Wang, and Feng Zheng. Benchmarks for corruption invariant person re-identification. *arXiv preprint arXiv:2111.00880*, 2021. 5

[4] Yi-Lei Chen and Chiou-Ting Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Proceedings of the IEEE international conference on computer vision*, pages 1968–1975, 2013. 5

[5] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017. 1

[6] Kshitiz Garg and Shree K Nayar. Photorealistic rendering of rain streaks. *ACM Transactions on Graphics (TOG)*, 25(3):996–1002, 2006. 2

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 3

[8] Diego Gragnaniello, Davide Cozzolino, Francesco Marra, Giovanni Poggi, and Luisa Verdoliva. Are gan generated images easy to detect? a critical analysis of the state-of-the-art. In *2021 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2021. 7

[9] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012. 6

[10] Qing Guo, Jingyang Sun, Felix Juefei-Xu, Lei Ma, Xiaofei Xie, Wei Feng, Yang Liu, and Jianjun Zhao. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1487–1495, 2021. 1, 6

[11] Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 297–304. JMLR Workshop and Conference Proceedings, 2010. 3

[12] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 6

[13] Mazin Hnewa and Hayder Radha. Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. *IEEE Signal Processing Magazine*, 38(1):53–67, 2020. 1

[14] Weilong Hou, Xinbo Gao, Dacheng Tao, and Xuelong Li. Blind image quality assessment via deep learning. *IEEE transactions on neural networks and learning systems*, 26(6):1275–1286, 2014. 6

[15] Xueqi Hu, Xinyue Zhou, Qiusheng Huang, Zhengyi Shi, Li Sun, and Qingli Li. Qs-attn: Query-selected attention for contrastive learning in i2i translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18291–18300, 2022. 2, 7, 8

[16] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7732–7741, 2021. 1

[17] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018. 2, 7, 8

[18] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021. 6

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[20] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018. 2

[21] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1633–1642, 2019. 8

[22] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 6

[23] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017. 2, 7, 8

[24] Changjie Lu, Shen Zheng, Zirui Wang, Omar Dib, and Gaurav Gupta. As-introvae: Adversarial similarity distance makes robust introvae. *arXiv preprint arXiv:2206.13903*, 2022. 8

[25] Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2704–2713, 2021. 1

[26] Yoshiki Mizukami, Katsuhiro Sasaki, and Katsumi Tadamura. Realistic rain rendering. In *GRAPP*, pages 273–280, 2008. 2

[27] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 2, 3, 4, 5, 6, 7, 8

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 5

[29] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 6

[30] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3937–3946, 2019. 1, 6

[31] Maria L Rizzo and Gábor J Székely. Energy distance. *wiley interdisciplinary reviews: Computational statistics*, 8(1):27–38, 2016. 6

[32] Pierre Rousseau, Vincent Jolivet, and Djamchid Ghazanfarpour. Realistic real-time rain rendering. *Computers & Graphics*, 30(4):507–518, 2006. 2

[33] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016. 6

[34] Zhiqiang Shen, Mingyang Huang, Jianping Shi, Xiangyang Xue, and Thomas S Huang. Towards instance-level image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3683–3692, 2019. 2

[35] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 3

[36] Hong Wang, Yichen Wu, Minghan Li, Qian Zhao, and Deyu Meng. A survey on rain removal from video and single image. *arXiv preprint arXiv:1909.08326*, 2019. 2

[37] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14791–14801, 2021. 1, 6

[38] Weilun Wang, Wengang Zhou, Jianmin Bao, Dong Chen, and Houqiang Li. Instance-wise hard negative example generation for contrastive learning in unpaired image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14020–14029, 2021. 2

[39] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3877–3886, 2019. 1

[40] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 8

[41] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017. 1

[42] Wenhan Yang, Robby T Tan, Shiqi Wang, Yuming Fang, and Jiaying Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on pattern analysis and machine intelligence*, 43(11):4059–4077, 2020. 1, 8

[43] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2726–2736, 2020. 1

[44] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 2

[45] Fangneng Zhan, Jiahui Zhang, Yingchen Yu, Rongliang Wu, and Shijian Lu. Modulated contrast for versatile image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18280–18290, 2022. 2, 3, 4, 5, 6, 7, 8

[46] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, pages 581–590, 2022. 8

[47] Shen Zheng, Changjie Lu, Yuxiong Wu, and Gaurav Gupta. Sapnet: Segmentation-aware progressive network for perceptual contrastive deraining. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 52–62, 2022. 1, 6

[48] Shen Zheng, Yiling Ma, Jinqian Pan, Changjie Lu, and Gaurav Gupta. Low-light image and video enhancement: A comprehensive survey and beyond. *arXiv preprint arXiv:2212.10772*, 2022. 8

[49] Shen Zheng, Yuxiong Wu, Shiyu Jiang, Changjie Lu, and Gaurav Gupta. Deblur-yolo: Real-time object detection with efficient blind motion deblurring. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2021. 5

[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2