Cooperative Image Segmentation and Restoration in Adverse Environmental Conditions

Weihao Xia, Zhanglin Cheng, Yujiu Yang

Abstract

Most state-of-the-art semantic segmentation or scene parsing approaches only achieve high accuracy rates in good environmental conditions. The performance decrease enormously if images with unknown disturbances occur, which is less discussed but appears more in real applications. Most existing research works cast the handling of the challenging adverse conditions as a post-processing step of signal restoration or enhancement after sensing, then feed the restored images for visual understanding. However, the performance will largely depend on the quality of restoration or enhancement. Whether restoration-based approaches would actually boost the semantic segmentation performance remains questionable. In this paper, we propose a novel net framework to tackle semantic Segmentation and image Restoration in adverse environmental conditions (SR-Net). The proposed approach contains two components: Semantically-Guided Adaptation, which exploits and leverages semantic information from degraded images then help to refine the segmentation; and Exemplar-Guided Synthesis, which synthesizes restored or enhanced images from semantic label maps given specific degraded exemplars. SR-Net exploits the possibility of building connections of low-level image processing and high level computer vision tasks, achieving image restoration via segmentation refinement. Extensive experiments on several datasets demonstrate that our approach can not only improve the accuracy of high-level vision tasks with image adaption, but also boosts the perceptual quality and structural similarity of degraded images with image semantic guidance.

1. Introduction

Robust and reliable visual sensing and understanding algorithms are crucial for emerging applications, such as UAVs, autonomous driving, search and rescue robots, security surveillance. Such systems concern a wide range of target tasks such as detection, recognition, segmentation, tracking, and parsing. However, since most current vision systems are designed to perform under optimal conditions, the performances of computer vision-based sensing and understanding of outdoor environments will be largely jeopardized by various challenging conditions in unconstrained and dynamic degraded environments, e.g., moving platforms, bad weathers, and poor illumination, as illustrated in Figure 1. Therefore, it is highly desirable to study how to cope with such challenging visual conditions for the goal of achieving robust visual sensing and understanding in the wild.

Conventionally, low-level image processing tasks, e.g., image restoration or enhancement, and high-level visual understanding problems are separately tackled by different frameworks. Most existing research works handle the challenging conditions as a independent post-processing step of image restoration or enhancement, and then feed the restored data for visual understanding. The performance of high-level visual understanding tasks will thus largely depend on the quality of restoration or enhancement. It remains questionable whether restoration-based approaches would actually improve the visual understanding performance, as the restoration or enhancement step might be suboptimal for the ultimate target task and may also bring in misleading information and lethal artifacts.

To address these bottlenecks, we present a novel framework for semantic segmentation and image restoration under adverse environmental conditions, called SR-Restore. We propose a new perspective for solving both the low-level image processing and high-level computer vision tasks in a single unified framework in a cooperative way. We connect the low-level (image restoration and enhancement) and high level computer vision tasks (semantic segmentation) by ex-
exploiting and leveraging mutual influence, achieving image restoration via segmentation refinement.

Our model contains two components: 1) The Segmentation Refinement Network, and 2) The Image Restoration Network. The segmentation refinement network, takes degraded image and its corresponding semantic map as input, and is trained to produce refined segmentation result by exploiting the intrinsic connection between raw images and semantic information. The image restoration network, restores degraded image with the complementary and auxiliary semantic information from refined segmentation.

Through the two-stage model, SR-Net explicitly formulates the translation from source (degradation) domain to target (refinement or restoration) domain as follows: Semantically-Guided Adaptation, which exploits and leverages semantic information from degraded images then help to refine the segmentation; and Exemplar-Guided Synthesis, which synthesizes restored or enhanced images from semantic label maps given specific degraded exemplars.

To sum up, our key contributions are three-fold:

- We propose a unified framework for cooperative image segmentation and restoration in adverse environmental conditions, including Semantically-Guided Adaptation and Exemplar-Guided Synthesis;
- We propose a Promotive Modification Algorithm for both components to promote the performance of each other during training;
- The systematical and numerous experiments demonstrate that our approach can not only improve the accuracy of high-level vision tasks with image adaption, but also boosts the perceptual quality and structural similarity of degraded images with image semantic guidance.

2. Related Work

2.1. Image Restoration for Multiple Degradations

Most existing image restoration methods are designed for specific subtask like image denoising of certain noise levels. They lack scalability in learning a single model to non-blindly deal with multiple degradations. Some recent studies aim to learn a single model to effectively handle multiple and even spatially variant degradations.

Zhang et al. [39] learns a single convolutional super-resolution network SRMD with high scalability for multiple degradations. The proposed super-resolver takes both LR image and its degradation maps as input. The results on both synthetic and real LR images reconstruct visually plausible HR images. Liu et al. [20] propose a non-local recurrent network (NLRN) as the first attempt to incorporate non-local operations into a recurrent neural network (RNN) and introduce the correlation propagation, which achieves superior results to state-of-the-art methods with many fewer parameters on image denoising and super-resolution tasks. Ulyanov et al. [31] proposes a non-trained method that using a randomly-initialized neural network as the handcrafted prior in standard inverse problems such as denoising, super-resolution, and inpainting. Yu et al. [37] adopts am agent to select a toolchain to progressively restore an image corrupted by complex and mixed distortions by deep reinforcement learning.

2.2. Semantic Segmentation in Adverse Environmental Conditions

Most semantic segmentation and computer vision approaches focus on achieving high scores at well-known benchmarks, but they do not care much about the robustness of their methods in adverse weather conditions. In fact, there are only a few approaches which can deal with disturbances caused by diverse weather often focusing on one special noise type.

For instance, Sakaridis et al. [28] increase the segmentation accuracy in fog by adding simulated fog to real-world images, and by training with these synthetic data. Porav et al. [26] derain rainy images by means of a neural network and achieve better performance, when a further segmentation network is applied on the derained images. Bijelic et al. [1] introduces the Robust Learning Technique, by which an image classifier becomes more robust against unknown disturbances not occurring in the training set. The authors randomly chose one of the input channels (camera image of depth image) and replace the corresponding input data by an arbitrarily selected sensor data of another, different scene. Valada et al. [32] select RGB, depth and EVI (Enhanced Vegetation Index) as the modalities to build expert networks for segmenting images in adverse environments. Pfeuffer et al. [25] use a multiple sensor-setup, e.g. disturbed input images and corresponding depth images observed by the lidar sensor to increase the robustness of the semantic labeling approaches.
2.3. Label-Based Image Synthesis

Label-based image synthesis methods have progressed rapidly during the last few years. The goal of is to synthesize photo-realistic and faithful images semantically from abstract label maps, such as sparse landmarks or pixel-wise segmentation maps.

Isola et al. [11] proposes a pix2pix model, aiming to translate precise label maps to pleasing facades pictures using conditional GANs. Zhu et al. [42] proposes cycle-consistent loss to handle the paired training data limitation of [11]. Wang et al. [33] proposes a framework for 2048 × 1024 high-resolution photorealistic image synthesis from semantic label maps. They improve the original pix2pix framework by using a coarse-to-fine generator and a multi-scale discriminator. Chen et al. [5] proposes a cascade framework to synthesize high-resolution images from pixel-wise labeling maps. To preserve semantic information through stacks of convolution, normalization, and non-linearity layers, Park et al. [24] propose spatially-adaptive normalization layer for synthesizing photo-realistic images given an input semantic layout. The proposed normalization leads to the first semantic image synthesis model that can produce photorealistic outputs for diverse scenes including indoor, outdoor, landscape, and street scenes.

3. Methodology

3.1. Overview of the Proposed Approach

We propose an unified cooperative semantic segmentation and image restoration framework for adverse environmental conditions. Given semantic segmentation approaches optimized for good conditions and images sensed in unconstrained and severely degraded environments, we refine segmentation results first, and restore original images with these refined segmentation. The overview of our proposed SR-Net approach is illustrated in Figure [2].

SR-Net explicitly formulates the domain two separate but complementary stages: Semantically-Guided Adaptation, which exploits and leverages semantic information from degraded images then help to refine the segmentation, and Exemplar-Guided Synthesis, which synthesizes restored or enhanced images from semantic label maps given specific degraded exemplars.

**Semantically-Guided Adaptation** aims to produce refined segmentation maps of degraded images. Taking another look at this process of refinement, we can formulate it as the adaption of segmentation from favourable images to degraded ones. For this purpose, we adopt a refinement network $G_{ref}$ which takes a segmentation result of degraded image $S_d$ and corresponding image $I_d$ as input. The refinement network $G_{ref}$ `refine` segmentation by learning the adaption of segmentation from favourable images to degraded ones, which exploits the difference between images in two domains and leverages semantic information to refine the segmentation results.

**Exemplar-Guided Synthesis** aims to generate restored images $I_r$ by deploying a restoration network $G_{res}$. $G_{res}$ takes original degraded image $I_d$ and refined segmentation map $S_r$ as input, and recover $I_d$ with the complementary and auxiliary semantic information from refined segmentation. Label-based image synthesis is a typical one-to-many translation problem, thus we use the original degraded image as external exemplar to control the global appearances of the output image. This recovery process can also be formulated as guided image synthesis.
3.2. Semantically-Guided Adaptation (SGAda)

The refinement network aims to modify segmentation results of input images. Let $S_d$ be segmentation result of degraded image. The ground truth segmentation will be denoted as $S_{gt}$. The adaptation from original segmentation maps $S_d$ to the modified ones $S_r$ can be denoted as $G_1 : \{S_d, I_d\} \rightarrow S_r$.

$$S_r = G_1(\{S_d, I_d\}, S_{gt}). \tag{1}$$

The total loss of refinement network combines an adversarial loss $L_{G_{ref}}$ and refinement loss $L_{ref}$ as

$$\min_{G_1} \max_{D_1} L_{G_1} = \min_{G_1} \left( \max_{D_1} \left( L_{G_{ref}} \right) + \lambda L_{ref} \right), \tag{2}$$

where $\lambda$ is regularization parameters controlling the importance of two terms. The adversarial loss $L_{G_1}$ is defined as

$$L_{G_1} = \mathbb{E}_{(S_{gt}, I_r)} [\log D_1(S_{gt}, I_r)] + \mathbb{E}_{(I_r)} \log [1 - D_1(S_r, I_r)]. \tag{3}$$

The discriminator $D_1$ is trained to distinguish between ground truth segmentation $S_{gt}$ and refined ones $S_r$ conditioned on degraded images $I_r$. $D_1$ is formulated as a SNGP-InstDisc [30] for its awareness of spatial contextual relations.

The refinement loss is defined by $L_{ref}$, learned by minimizing a standard cross-entropy loss:

$$L_{ref} = \mathbb{E} [CE(S_r, S_{gt})], \tag{4}$$

where $S_{gt}$ stands for ground-truth semantic labels, and $S_r$ stands for predicted labels.

For stable training, high image quality and considerable diversity, we use the least-squares GAN [22] in our experiment.

3.3. Exemplar-Guided Synthesis (EGSyn)

The refined segmentation result $S_r$ and original degraded image $I_d$ are then fed into the restoration network to generate restored image $I_r$. This translation process from the refined segmentation $S_r$ to the restoration $I_r$ can be defined as $G_2 : \{S_r, I_d\} \rightarrow I_r$. The refined segmentation $I_r$ helps to restore the original degraded image $I_d$, while sharing the same resolution with the input image:

$$I_r = G_2(\{S_r, I_d\}, I_{gt}) \tag{5}$$

We train this image synthesis network with a joint loss, which consists of five terms: an adversarial loss $L_{G_2}$, $\ell_1$ reconstruction loss $L_{\ell_1}$, perceptual loss $L_{percep}$, style loss $L_{style}$, and total variation loss $L_{tv}$:

$$L_{G_2} = \lambda_1 L_{\ell_1} + \lambda_2 L_{G_{ref}} + \lambda_3 L_{percep} + \lambda_4 L_{style} + L_{tv}. \tag{6}$$

The adversarial loss is defined similar to Equation 3 as

$$L_{G_2} = \mathbb{E}_{(I_{ref})} [\log D_2(I_{gt})] + \mathbb{E}_{(I_r)} \log [1 - D_2(I_r)]. \tag{7}$$

The perceptual loss is proposed by Johnson et al. [12] based on perceptual similarity. It is originally defined as the distance between two activated features of a pre-trained deep neural network. Here we adopt a more effective perceptual loss which uses features before activation layers [34]. These features are more dense and thus provide relatively stronger supervision, leading to better performance:

$$L_{percep} = \mathbb{E}[\|\phi_i(I_{gt}) - \phi_i(I_r)\|_1], \tag{9}$$

where $\phi_i$ donates the feature maps before activation of the VGG-19 network pre-trained for image classification.

The style loss is adopted in the same form as in the original work [12] [9], which aims to measure differences between covariance of activation features:

$$L_{style} = \mathbb{E}_j \left[ \|G_{\phi_j}^\phi(I_{gt}) - G_{\phi_j}^\phi(I_r)\|_1 \right], \tag{10}$$

where $G_{\phi_j}^\phi$ represents the Gram matrix constructed from feature maps $\phi_j$. It’s shown to be an effective tool to alleviate ‘checkerboard’ artifacts caused by transpose convolution layers [27].

We also add total variation loss to remove unwanted noises and encourage spatial smoothness in the generated images. It is useful to mitigate checkerboard artifacts from the perceptual loss term. It is defined on the basis of the absolute gradient of generated images:

$$L_{tv} = \|\nabla_x I_r - \nabla_y I_r\|_1. \tag{11}$$

3.4. Promotive Modification Algorithm

Our model consists of two constituent components: SGAda and EGSyn. SGAda aims to exploit the difference of distributions between degradation and degradation-free images and adapt the segmentation. EGSyn tries to reconstruct the restoration from the label and leverage the original degraded image as exemplar. To get an optimal model, we design a training strategy named Gradient Promotive Modification Algorithm. More specifically, the training processes are divided into three stages. Firstly, we train our SGAda module including $G_1$ and $D_1$ using the ground truth label as supervision. Meanwhile, we train the EGSyn module $G_2$ and $D_2$ using the refined map together with original degraded image as input and ground truth images as supervision. We then jointly train both SGAda and EGSyn in an end-to-end way until convergence.
The idea behind this is that the quality of image restoration results mainly depends on the refined labels produced by SGAda. Though the EGSyn has the degradation as exemplar, the layout of restoration mainly depends on the generated label and the degradation exemplar decides other elements like textures. If the segmentation itself were not precise enough, the restoration would be consistent with the wrong pixels of the label. The EGSyn back-propagate the errors to guide the training to the direction that learns more precise segmentation results. Both components promote each other.

4. Experiments

4.1. Experiment Settings

We conduct experiments on several datasets and different degradation types. We also conduct experiments on Cityscapes under different and illumination conditions (e.g. fog, rain, reflection). These datasets [6, 10, 29] and corresponding results can be found in supplementary material.

4.1.1 Datasets

- **COCO-Stuff** [3] is derived from the COCO dataset [18]. It has 118,000 training images and 5,000 validation images captured from diverse scenes with 182 semantic classes.
- **PASCAL VOC 2012** [8] contains nearly 10K images annotated with pixel-wise segmentation of each object present.
- **ADE20K** [41] contains more than 20K scene-centric images exhaustively annotated with objects and object parts. There are totally 150 semantic categories, which include stuffs like sky, road, grass, and discrete objects like person, car, bed.
- **CelebAMask-HQ** [15] is a large-scale face image dataset that has 30,000 high-resolution face images selected from the CelebA-HQ [13]. The masks were manually-annotated with the size of $512 \times 512$ and 19 classes including all facial components and accessories such as skin, nose, eyes.

We use pre-trained models to compute the input segmentation or parsing masks, *i.e.* ResNet50dilated model for Semantic Segmentation on MIT ADE20K dataset, DeepLabV2 [4] model on COCO-Stuff / PASCAL VOC 2012 dataset, and BiSeNet [35] model for face parsing on CelebAMask-HQ dataset.

4.1.2 Degradation Types

We use six types of degradation for experiments. These degradations are caused by moving platforms, bad weather, poor illumination or transmission procedure, and are common in both cases of daily life and emerging applications like autonomous driving.

- **Irregular Mask.** To generate corrupted images for image inpainting, we use irregular mask dataset provided by [21] and Quick Draw Irregular Mask Dataset [1] which is combination of 50 million strokes drawn by human hand. To a certain extent, irregular mask can simulate the case that camera lens are stained by impurities like mud.
- **Random Watermark.** The angle, size, location of watermark is random. The opacity of logo images are set to be $[0.25, 0.5, 0.75, 0.98]$.
- **Gaussian Blur.** 2D circularly symmetric Gaussian blur kernels are applied with standard deviations set to be $[1.2, 2.5, 6.5, 15.2]$.
- **Gaussian Noise.** The local variance of the Gaussian noise added is set to be $[0.05, 0.09, 0.13, 0.2]$.
- **JPEG compression.** The quality factor that determines the DCT quantization matrix is set to be $[43, 12, 7, 4]$.
- **Chromatic Aberrations.** Color images are often degraded by the residual chromatic aberrations of the optical system, causing chromatism errors and a decrease of resolution. For chromatic aberration recovery, the mutual shifting of in R and B channels is set to be $[2, 6, 10, 14]$ and $[1, 3, 5, 7]$, respectively.

4.2. Results

We first evaluate the proposed method for semantic segmentation and image restoration under adverse environmental conditions on COCO-Stuff [3], PASCAL VOC 2012 [8], ADE20K [41], and CelebAMask-HQ [15]. Then we compare the proposed SR-Net with state-of-the-art image restoration for multiple degradations. Results for Cityscapes under different and illumination condition (e.g. fog, rain, reflection) can be found in supplementary material.

4.2.1 Qualitative Comparison

Figure 4 and Figure 5 shows a sample of images generated by our method. Degradation types in Figure 4 from top to bottom are CA, GB, GN and JPEG, respectively. We also use the proposed method to remove watermark or text overlaid on an image. The segmentation as illustrated in Figure 5 our approach leads to satisfactory results with virtually no artifacts.

---

1[https://github.com/karfly/qd-imd](https://github.com/karfly/qd-imd)
Table 1. Quantitative results of different degradation type on CelebAMask-HQ[15]. CA, GB, GN and JPEG means Chromatic Aberrations, Gaussian Blur, Gaussian Noise and JPEG Compression, respectively. The best performance is in bold.

| Degradation | Degraded PSNR | Degraded SSIM | Deep Prior [31] PSNR | Deep Prior [31] SSIM | RL-Restore [37] PSNR | RL-Restore [37] SSIM | NLRN [20] PSNR | NLRN [20] SSIM | SR-Net PSNR | SR-Net SSIM |
|-------------|---------------|---------------|----------------------|----------------------|----------------------|----------------------|----------------|----------------|----------------|-------------|
| CA          | 23.05 0.7407  | 23.25 0.7539  | 31.17 0.9025         | 31.83 0.9214         | 32.03 0.9271         |                      |                 |                 |                |             |
| GB          | 24.11 0.7562  | 24.58 0.7608  | 28.40 0.8827         | 29.71 0.8864         | 30.97 0.8931         |                      |                 |                 |                |             |
| GN          | 22.32 0.5946  | 22.45 0.6167  | 26.36 0.8011         | 28.11 0.8339         | 28.37 0.8388         |                      |                 |                 |                |             |
| JPEG        | 25.67 0.7482  | 26.11 0.7523  | 27.09 0.8115         | 27.52 0.8295         | 27.73 0.8384         |                      |                 |                 |                |             |

For visualization purposes, we follow the color encoding scheme of CelebAMask-HQ[15] to colorize the label map. As shown, the obtained semantic maps of images degraded by irregular masks are terrible since the segmentation method we adopted are designed to work under optimal conditions.

From these results, we observe that the segmentation and restoration results generally get improved with our SR-Net approach. Due to the semantically-guided adaptation, semantic information from degraded images is exploited and the information is then leveraged to refine the original segmentation. Exemplar-guided synthesis learns how to generate restored images from semantic label maps together with specific degraded exemplars. Extensive qualitative results of COCO-Stuff, PASCAL VOC 2012, and ADE20K are demonstrated in Figure 6.

4.2.2 Quantitative Comparison

The quantitative comparison results are shown in Table 2. For semantic segmentation, we use popular Pixel Accuracy (PA), mean Pixel Accuracy (mPA), mean Intersection over Union (mIoU), Frequency Weighted Intersection over Union (FWIoU) as evaluation metrics. For image restoration, we use PSNR and SSIM as evaluation metrics.

The segmentation results of original images without any degradation are presented in the first row. Performances rapidly deteriorate when the input images are corrupted with different degradation types. We quantify the extent of segmentation for different degradation types. Specifically, images corrupted by gaussian blur and irregular mask get the lowest and the second lowest segmentation performance. Other degradation types have relatively little influence.
Table 2. Quantitative results on CelebAMask-HQ [15]. IM, CA, GB, GN and JPEG means Irregular Mask, Chromatic Aberrations, Gaussian Blur, Gaussian Noise and JPEG Compression, respectively. IM_Re means Refinement or Restoration results of Irregular Masked images. For Semantic Segmentation, we use Pixel Accuracy (PA), mean Pixel Accuracy (mPA), mean Intersection over Union (mIoU), Frequency Weighted Intersection over Union (FWIoU) as evaluation metrics. For Image Restoration, we use PSNR and SSIM as evaluation metrics.

For corrupted images of different degradation types, the refinement network improves the segmentation or parsing performance with a large margin. The experiments demonstrate the universality and scalability of our proposed method. In the case of Gaussian Noise, the performance even outperforms the original images.

We further compare the proposed method with state-of-the-art image restoration methods for multiple degradations. Deep Prior [31] is non-trained image restoration method for image denosing, deJPEG and super-resolution. RL-Restore [37] adopts an agent to select a toolchain to progressively restore an image corrupted by complex and mixed distortions. DnCNN [38], VDSR [14], MemNet [30] and NLRN [20] are state-of-the-art learning-based models for handling multiple degradations. Since RL-Restore and NLRN have claimed their superiority over DnCNN, VDSR and MemNet, we didn’t compare with these methods.

The results of all methods are summarized in Table 1. Bold indicates the best performance. Visual comparison are illustrated in Figure 3. The restoration results of Deep Prior [31] are not good, which perhaps can be due to the sensi-
Figure 6. Qualitative results on COCO-Stuff [3] (GN), PASCAL VOC 2012 [8] (JPEG), and ADE20K [41] (IM). From left to right: input image, segmentation result of input image, refined segmentation, ground truth label, and restored image. For visualization purposes, we use the color encoding scheme to colorize the label map.

Activity of hyper-parameters that we didn’t deliberately select for each degraded image. In most cases of RL-Restore [37], the results are quite satisfactory. However, the performance drops, sometimes image barely ameliorates when the distortions are extremely severe. NLRN [20] performs good on image denoising and image dejpeg as in the original paper, we surprisingly found that it also produce comparatively satisfactory results on image deblur task. Our method produces best quantitative and visual results.

4.2.3 Effect of Promotive Modification Algorithm

To show how the promotive modification algorithm train the model in a cooperative way and how the low/high-level representations promote the performance for each task, we train the two components SGAda and EGSyn separately, and no information are transformed during training. We use mIoU and PSNR to evaluate segmentation and restoration, respectively. The results are shown in Table 3 which clearly corroborates that not only image adaptation helps improve the accuracy of semantic segmentation, but also image semantic guidance boosts the perceptual quality and structural similarity of degraded images.

5. Conclusion and Future Work

In this paper, we have introduced a novel SR-Net framework to tackle cooperative semantic segmentation and image restoration under adverse environmental conditions. SR-Net explicitly formulates the problem two separate but complementary stages: Semantically-Guided Adaptation and Exemplar-Guided Synthesis, which exploits the possibility of building connections between low-level computer vision tasks and high-level visual understanding tasks. The experiments demonstrate that our approach can not only improve the accuracy of semantic segmentation with image adaptation, but also boosts the perceptual quality and structural similarity of degraded images with image semantic guidance.

For now, one model has to be trained for each degradation. In the future, we will ameliorate the limitation by using the multi-domain image translation frameworks [16][19] or multiple source domain adaption methods [40][17]. Furthermore, recent works [11] have inspired us to better utilize the constraints of the guidance image. We are also intend to further exploit the connection between image restoration and other high-level visual cues like depth.

Table 3. Effect of our proposed Promotive Modification Algorithm. We use mIoU and PSNR to evaluate segmentation and restoration, respectively.

| Ablation | CelebA | ADE20K |
|----------|--------|--------|
|          | Separate | Joint | Separate | Joint |
| IM       | Seg. | 0.5839 | 0.6143 | 0.2975 | 0.3746 |
|          | Res.  | 25.67  | 30.68  | 22.14  | 28.33  |
| GN       | Seg. | 0.7036 | 0.7505 | 0.3052 | 0.3719 |
|          | Res.  | 23.91  | 28.37  | 23.85  | 29.71  |

References

[1] Badour AlBahar and Jia-Bin Huang. Guided image-to-image translation with bi-directional feature transformation. In The IEEE International Conference on Computer Vision, 2019.
[2] Mario Bijelic, Christian Muench, Werner Ritter, Yuri Kalnishkan, and Klaus Dietmayer. Robustness against unknown noise for raw data fusing neural networks. In Inter-
[3] Holger Caesar, Jasper R. R. Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In Conference on Computer Vision and Pattern Recognition, pages 1209–1218, 2018.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell., 40(4):834–848, 2018.

[5] Qifeng Chen and Vladlen Koltun. Photographic image synthesis with cascaded refinement networks. In IEEE International Conference on Computer Vision, pages 1520–1529, 2017.

[6] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset. In CVPR Workshop on The Future of Datasets in Vision, 2015.

[8] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John M. Winn, and Andrew Zisserman. The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 88(2):303–338, 2010.

[9] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2414–2423, 2016.

[10] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In IEEE Conference on Computer Vision and Pattern Recognition, pages 8022–8031, 2019.

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In Computer Vision and Pattern Recognition, 2017.

[12] Justin Johnson, Alexandre Alahi, and Fei Fei Li. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision, pages 694–711, 2016.

[13] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In International Conference on Learning Representations, 2018.

[14] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In Conference on Computer Vision and Pattern Recognition, pages 1646–1654, 2016.

[15] Cheng-Han Lee, Zhiwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. Technical Report, 2019.

[16] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Kumar Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. arXiv preprint arXiv:1905.01270, 2019.

[17] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Zi Huang. Cycle-consistent conditional adversarial transfer networks. In ACM MM. ACM, 2019.

[18] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In ECCV, pages 740–755, 2014.

[19] Alexander H. Liu, Yen-Cheng Liu, Yu-Ying Yeh, and Yu-Chiang Frank Wang. A unified feature disentangler for multi-domain image translation and manipulation. In Advances in Neural Information Processing Systems, pages 2595–2604, 2018.

[20] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S. Huang. Non-local recurrent network for image restoration. In Advances in Neural Information Processing Systems, pages 1680–1689, 2018.

[21] Guilin Liu, Fitosim A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In ECCV, pages 89–105, 2018.

[22] Xudong Mao, Qiong Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In IEEE International Conference on Computer Vision, pages 2794–2802, 2017.

[23] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In Advances in Neural Information Processing Systems, pages 2802–2810, 2016.

[24] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[25] Andreas Pfeuffer and Klaus Dietmayer. Robust semantic segmentation in adverse weather conditions by means of sensor data fusion. CoRR, abs/1905.10117, 2019.

[26] Horia Porav, Tom Bruls, and Paul Newman. I can see clearly now: Image restoration via de-raining. CoRR, abs/1901.00893, 2019.

[27] Mehdi S. M. Sajjadi, Bernhard Schölkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In IEEE International Conference on Computer Vision, pages 4501–4510, 2017.

[28] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision, 126(9):973–992, 2018.

[29] C. Sakaridis, D. Dai, and L. Van Gool. Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision, 2018.

[30] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In IEEE International Conference on Computer Vision, pages 4549–4557, 2017.
[31] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. In IEEE Conference on Computer Vision and Pattern Recognition, pages 9446–9454, 2018.

[32] Abhinav Valada, Johan Vertens, Ankit Dhall, and Wolfram Burgard. Adapnet: Adaptive semantic segmentation in adverse environmental conditions. In International Conference on Robotics and Automation, pages 4644–4651, 2017.

[33] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[34] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In The European Conference on Computer Vision Workshops (ECCVW), September 2018.

[35] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisonet: Bilateral segmentation network for real-time semantic segmentation. In ECCV, pages 334–349, 2018.

[36] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. arXiv preprint arXiv:1806.03589, 2018.

[37] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2443–2452, 2018.

[38] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans. Image Processing, 26(7):3142–3155, 2017.

[39] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In IEEE Conference on Computer Vision and Pattern Recognition, pages 3262–3271, 2018.

[40] Han Zhao, Shanghang Zhang, Guanhang Wu, José MF Moura, Joao P Costeira, and Geoffrey J Gordon. Adversarial multiple source domain adaptation. In Advances in Neural Information Processing Systems, pages 8568–8579, 2018.

[41] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[42] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In IEEE International Conference on Computer Vision, 2017.