Improved Techniques for Learning to Dehaze and Beyond: A Collective Study

Yu Liu[1], Guanlong Zhao[2], Boyuan Gong[2], Yang Li[1], Ritu Raj[2], Niraj Goel[2], Satya Kesav[2], Sandeep Gottimukkala[2], Zhangyang Wang[2], Wenqi Ren[3], Dacheng Tao[4]
[1]Department of Electrical and Computer Engineering, Texas A&M University
[2]Department of Computer Science and Engineering, Texas A&M University
[3]Chinese Academy of Sciences [4] University of Sydney

Abstract

Here we explore two related but important tasks based on the recently released REalistic Single Image DEhazing (RESIDE) benchmark dataset: (i) single image dehazing as a low-level image restoration problem; and (ii) high-level visual understanding (e.g., object detection) of hazy images. For the first task, we investigated a variety of loss functions and show that perception-driven loss significantly improves dehazing performance. In the second task, we provide multiple solutions including using advanced modules in the dehazing-detection cascade and domain-adaptive object detectors. In both tasks, our proposed solutions significantly improve performance. GitHub repository URL: `https://github.com/guanlongzhao/dehaze`.

1. Introduction

Images taken in outdoor environments affected by air pollution, dust, mist, and fumes often contain complicated, non-linear, and data-dependent noise, also known as haze. Haze complicates many high-level computer vision tasks such as object detection and recognition. Therefore, dehazing has been widely studied in the fields of computational photography and computer vision. Early dehazing approaches often required additional information such as the provision or capture of scene depth by comparing several different images of the same scene [1, 2, 3]. Many approaches have since been proposed to exploit natural image priors and to perform statistical analyses [4, 5, 6, 7]. Most recently, dehazing algorithms based on neural networks [8, 9, 10] have delivered state-of-the-art performance. For example, AOD-Net [10] trains an end-to-end system and shows superior performance according to multiple evaluation metrics, improving object detection in the haze using end-to-end training of dehazing and detection modules.

2. Review and Task Description

Here we study two haze-related tasks: 1) boosting single image dehazing performance as an image restoration problem; and 2) improving object detection accuracy in the presence of haze. As noted by [11, 10, 12], the second task is related to, but is often unaligned with, the first.

While the first task has been well studied in recent works, we propose that **the second task is more relevant in practice and deserves greater attention**. Haze does not affect human visual perceptual quality as much as resolution, noise, and blur; indeed, some hazy photos may even have better aesthetics. However, haze in unconstrained outdoor environments could be detrimental to machine vision systems, since most of them only work well for haze-free scenes. Taking autonomous driving as an example, hazy and foggy weather will obscure the vision of on-board cameras and create confusing reflections and glare, creating problems even for state-of-the-art self-driving cars [12].

2.1. Haze Modeling and Dehazing Approaches

The atmospheric scattering model has been widely used to represent hazy images in haze removal works [13, 14, 15]:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where x indexes pixels in the observed hazy image, $I(x)$ is the observed hazy image, and $J(x)$ is the clean image to be recovered. The parameter A denotes the global atmospheric light, and $t(x)$ is the transmission matrix defined as:

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

where β is the scattering coefficient, and $d(x)$ represents the distance between the object and camera.

Conventional single image dehazing methods commonly exploit natural image priors (for example, the dark channel prior (DCP) [4, 5], the color attenuation prior [6], and the non-local color cluster prior [7]) and perform statistical analysis to recover the transmission matrix $t(x)$. More recently, convolutional neural networks(CNNs) have been applied for haze removal after demonstrating success in many other computer vision tasks. Some of the most effective

models include the multi-scale CNN (MSCNN) which predicts a coarse-scale holistic transmission map of the entire image and refines it locally [9]; DehazeNet, a trainable transmission matrix estimator that recovers the clean image combined with estimated global atmospheric light [8]; and the end-to-end dehazing network, AOD-Net [10, 16], which takes a hazy image as input and directly generates a clean image output. AOD-Net has also been extended to video [17].

2.2. RESIDE Dataset

We benchmark against the REalistic Single Image DEhazing (RESIDE) dataset [12]. RESIDE was the first large-scale dataset for benchmarking single image dehazing algorithms and includes both indoor and outdoor hazy images[1]. Further, RESIDE contains both synthetic and real-world hazy images, thereby highlighting diverse data sources and image contents. It is divided into five subsets, each serving different training or evaluation purposes. RESIDE contains $110,500$ synthetic indoor hazy images (ITS) and $313,950$ synthetic outdoor hazy images (OTS) in the training set, with an option to split them for validation. The RESIDE test set is uniquely composed of the synthetic objective testing set (SOTS), the annotated real-world task-driven testing set (RTTS), and the hybrid subjective testing set (HSTS) containing $1,000$, $4,332$, and 20 hazy images, respectively. The three test sets address different evaluation viewpoints including restoration quality (PSNR, SSIM and no-reference metrics), subjective quality (rated by humans), and task-driven utility (using object detection, for example).

Most notably, RTTS is the only existing public dataset that can be used to evaluate object detection in hazy images, representing mostly real-world traffic and driving scenarios. Each image is annotated with object bounding boxes and categories (person, bicycle, bus, car, or motorbike). $4,807$ unannotated real-world hazy images are also included in the dataset for potential domain adaptation.

For Task 1, we used the training and validation sets from ITS + OTS, and the evaluation is based on PSNR and SSIM. For Task 2, we used the RTTS set for testing and evaluated using mean average precision (MAP) scores.

3. Task 1: Dehazing as Restoration

Most CNN dehazing models [8, 9, 10] refer to the mean-squares error (MSE) or ℓ_2 norm-based loss functions. However, MSE is well-known to be imperfectly correlated with human perception of image quality [18, 19]. Specifically, for dehazing, the ℓ_2 norm implicitly assumes that the degradation is additive white Gaussian noise, which is oversimplified and invalid for haze. Conversely, ℓ_2 treats the impact of noise independently of the local image characteristics such as structural information, luminance and contrast. However, according to [20], the sensitivity of the Human Visual System (HVS) to noise depends on the local properties and structure of a vision.

Here we aimed to identify loss functions that better match human perception to train a dehazing neural network. We used AOD-Net [10] (originally optimized using MSE loss) as the backbone but replaced its loss function with the following options:

- ℓ_1 **loss**: The ℓ_1 loss for a patch P can be written as:

$$\mathcal{L}^{\ell_1}(P) = \frac{1}{N}\sum_{p\in P}|x(p) - y(p)|. \tag{3}$$

  where N is the number of pixels in the patch, p is the index of the pixel, and $x(p)$ and $y(p)$ are the pixel values of the generated image and the ground truth image respectively.

- **SSIM loss**: Following [19], we write the SSIM for pixel p as:

$$\begin{aligned} SSIM(p) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ &= l(p) \cdot cs(p). \end{aligned} \tag{4}$$

  The means and standard deviations are computed using a Gaussian filter with standard deviation σ_G. The loss function for SSIM can then be defined as:

$$\mathcal{L}^{SSIM}(P) = \frac{1}{N}\sum_{p\in P} 1 - SSIM(p). \tag{5}$$

- **MS-SSIM loss**: The choice of σ_G would impact the training performance of SSIM. Here we adopt the idea of multi-scale SSIM [19], where M different values of σ_G are pre-chosen and fused:

$$\mathcal{L}^{MS-SSIM}(P) = l_M^{\alpha}(p) \cdot \prod_{j=1}^{M} cs_j^{\beta_j}(P). \tag{6}$$

- **MS-SSIM+ℓ_2 Loss**: using a weighted sum of MS-SSIM and ℓ_2 as the loss function:

$$\mathcal{L}^{MS-SSIM-\ell_2} = \alpha\cdot\mathcal{L}^{MSSSIM} + (1-\alpha)\cdot G_{\sigma_G^M}\cdot\mathcal{L}^{\ell_2}, \tag{7}$$

  a point-wise multiplication between $G_{\sigma_G^M}$ and $\mathcal{L}^{\ell_2}$ is added for the ℓ_2 loss function term, because MS-SSIM propagates the error at pixel q based on its contribution to MS-SSIM of the central pixel $\widetilde{q}$, as determined by the Gaussian weights.

[1] The RESIDE dataset was updated in March 2018, with some changes made to dataset organization. Our experiments were all conducted on the original RESIDE version, now called RESIDE-v0.

Models	PSNR		
	Indoor	Outdoor	All
AOD-Net Baseline	**21.01**	24.08	22.55
ℓ_1	20.27	25.83	23.05
SSIM	19.64	26.65	23.15
MS-SSIM	19.54	**26.87**	23.20
MS-SSIM+ℓ_1	20.16	26.20	23.18
MS-SSIM+ℓ_2	20.45	26.38	**23.41**
MS-SSIM+ℓ_2 (fine-tuned)	20.68	26.18	**23.43**

Table 1. Comparison of PSNR results (dB) for Task 1.

Models	SSIM		
	Indoor	Outdoor	All
AOD-Net Baseline	**0.8372**	0.8726	0.8549
ℓ_1	0.8045	0.9111	0.8578
SSIM	0.7940	0.8999	0.8469
MS-SSIM	0.8038	0.8989	0.8513
MS-SSIM+ℓ_1	0.8138	**0.9184**	0.8661
MS-SSIM+ℓ_2	0.8285	0.9177	**0.8731**
MS-SSIM+ℓ_2 (fine-tuned)	0.8229	**0.9266**	**0.8747**

Table 2. Comparison of SSIM results for Task 1.

- **MS-SSIM+ℓ_1 loss**: using a weighted sum of MS-SSIM and ℓ_1 as the loss function:

$$\mathcal{L}^{MSSSIM-\ell_1} = \alpha \cdot \mathcal{L}^{MSSSIM} + (1-\alpha) \cdot G_{\sigma_G^M} \cdot \mathcal{L}^{\ell_1}, \quad (8)$$

the ℓ_1 loss is similarly weighted by $G_{\sigma_G^M}$.

We selected 1,000 images from ITS + OTS as the validation set and the remaining images for training. The initial learning rate and mini-batch size of the systems were set to 0.01 and 8, respectively, for all methods. All weights were initialized as Gaussian random variables, unless otherwise specified. We used a momentum of 0.9 and a weight decay of 0.0001. We also clipped the ℓ_2 norm of the gradient to be within [-0.1, 0.1] to stabilize network training. All models were trained on an Nvidia GTX 1070 GPU for around 14 epochs, which empirically led to convergence. For SSIM loss, σ_G was set to 5. C_1 and C_2 in (4) were 0.01 and 0.03, respectively. For MS-SSIM losses, the multiple Gaussian filters were constructed by setting $\sigma_G^i = \{0.5, 1, 2, 4, 8\}$. α was set as 0.025 for MS-SSIM+ℓ_1, and 0.1 for MS-SSIM+ℓ_2, following [19].

As shown in Tables 1 and 2, simply replacing the loss functions resulted in noticeable differences in performance. While the original AOD-Net with MSE loss performed well on indoor images, it was less effective on outdoor images, which are usually the images needing to be dehazed in practice. Of all the options, MS-SSIM-ℓ_2 achieved both the highest overall PSNR and SSIM results, resulting in 0.88 dB PSNR and 0.182 SSIM improvements over the state-of-the-art AOD-Net. We further fine-tuned the MS-SSIM-ℓ_2 model, including using a pre-trained AOD-Net as a warm initialization, adopting a smaller learning rate (0.002) and a larger minibatch size (16). Finally, the best achievable PSNR and SSIM were 23.43 dB and 0.8747, respectively. Note that the best SSIM represented a nearly 0.02 improvement over AOD-Net.

4. Task 2: Dehazing for Detection

4.1. Solution Set 1: Enhancing Dehazing and/or Detection Modules in the Cascade

In [10], the authors proposed a cascade of AOD-Net dehazing and Faster-RCNN [21] detection modules to detect objects in hazy images. We therefore considered it intuitive to try different combinations of more powerful dehazing/detection modules in the cascade. Note that such a cascade could be subject to further joint optimization, as many previous works [22, 23, 10]. However, **to be consistent with the results in** [12], all detection models used in this section were the original pre-trained versions, *without any re-training or adaptation*.

Our solution set 1 considered several popular dehazing modules including DCP [4], DehazeNet [8], AOD-Net [10], and the recently proposed densely connected pyramid dehazing network (DCPDN) [24]. Since hazy images tend to have lower contrast, we also included a contrast enhancement method called contrast limited Adaptive histogram equalization (CLAHE). Regarding the choice of detection modules, we included Faster R-CNN [21][2], SSD [26], RetinaNet [27], and Mask-RCNN [28].

The compared pipelines are shown in Table 3. In each pipeline, "X+Y" by default means applying Y directly on the output of X in a sequential manner. The most important observation is that simply applying more sophisticated detection modules is unlikely to boost the performance of the dehazing-detection cascade, due to the domain gap between hazy/dehazed and clean images (on which typical detectors are trained). The more sophisticated pre-trained detectors (RetinaNet, Mask-RCNN) may have overfitted the clean image domain, again highlighting the demand of handling domain shifts in real-world detection problems. Moreover, a better dehazing model in terms of restoration performance does not imply better detection results on its pre-processed images (e.g., DPDCN). Further, adding dehazing pre-processing does not always guarantee better de-

[2]We replace the backbone of Faster R-CNN from VGG 16 as used by [12] with the ResNet101 model [25] to enhance performance.

Pipelines	mAP
Faster R-CNN	0.541
SSD	0.556
RetinaNet	0.531
Mask-RCNN	0.457
DehazeNet + Faster R-CNN	0.557
AOD-Net + Faster R-CNN	0.563
DCP + Faster R-CNN	0.567
DehazeNet + SSD	0.554
AOD-Net + SSD	0.553
DCP + SSD	0.557
AOD-Net + RetinaNet	0.419
DPDCN + RetinaNet	0.543
DPDCN + Mask-RCNN	0.477
AOD-Net + DCP + Faster R-CNN	0.568
CLACHE + DCP + Mask-RCNN	0.551

Table 3. Solution set 1 mAP results on RTTS. Top 3 results are colored in red, green, and blue, respectively.

tection (e.g, comparing RetinaNet versus AOD-Net + RetinaNet), consistent with the conclusion made in [12]. In addition, AOD-Net tended to generate smoother results but with lower contrast than the others, potentially compromising detection. Therefore, we created two three-stage cascades as in the last two rows of Table 3, and found that using DCP to process AOD-Net dehazed results (with greater contrast) further marginally improved results.

4.2. Solution Set 2: Domain-Adaptive Mask-RCNN

Motivated by the observations made on solution set 1, we next aimed to more explicitly tackle the domain gap between hazy/dehazed images and clean images for object detection. Inspired by the recently proposed domain adaptive Faster-RCNN [29], we applied a similar approach to design a domain-adaptive mask-RCNN (DMask-RCNN).

In the model shown in Figure 1, the primary goal of DMask-RCNN is to mask the features generated by feature extraction network to be as domain invariant as possible, between the source domain (clean input images) and the target domain (hazy images). Specifically, DMask-RCNN places a domain-adaptive component branch after the base feature extraction convolution layers of Mask-RCNN. The loss of the domain classifier is a binary cross entropy loss:

$$-\sum_i (y_i log(p_i) + (1-y_i) log(1-p_i)), \qquad (9)$$

where y_i is the domain label of the i_{th} image, and p_i is the prediction probability from the domain classifier. The overall loss of DMask-RCNN can therefore be written as:

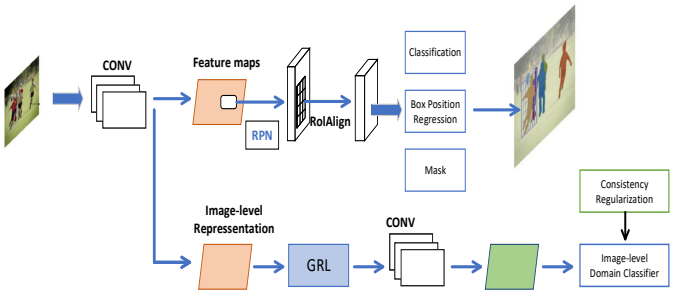

Figure 1. DMask-RCNN structure.

$$\begin{aligned} L(\theta_{res}, \theta_{head}, \theta_{domain}) = & L_{C,B}(C, B|\theta_{res}, \theta_{head}, x \in D_s) \\ & - \lambda L_d(G_d|\theta_{res}, x \in D_s, D_t) \\ & + \lambda L_d(G_d|\theta_{domain}, x \in D_s, D_t), \end{aligned} \qquad (10)$$

where x is the input image, and D_s and D_t represents the source and target domain, respectively. θ denotes the corresponding weights of each network component. G represents the mapping function of the feature extractor; I is the feature map distribution; B is the bounding box of an object and C is the object class. Note that when calculating the $L_{C,B}$, only source domain inputs will be counted in since the target domain has no labels.

As seen from Eqn. (10), the negative gradient of the domain classifier loss needs to be propagated back to ResNet, whose implementation relies on the gradient reverse layer [30] (GRL, Fig. 1). The GRL is added after the feature maps generated by the ResNet and feeds its output to the domain classifier. This GRL has no parameters except for the hyper-parameter λ, which, during forward propagation, acts as an identity transform. However, during back propagation, it takes the gradient from the upper level and multiplies it by $-\lambda$ before passing it to the preceding layers.

Experiments To train DMask-RCNN, MS COCO (clean images) were always used as the source domain, while **two target domain options** were designed to consider two types of domain gap: (1) all unannotated realistic haze images from RESIDE; and (2) dehazed results of those unannotated images, using MSCNN [9]. The corresponding DMask-RCNNs are called DMask-RCNN1 and DMask-RCNN2, respectively.

We initialized the Mask-RCNN component of DMask-RCNN with a pre-trained model on MS COCO. All models were trained for 50, 000 iterations with learning rate 0.001, then another 20, 000 iterations with learning rate 0.0001. We used a naive batch size of 2, including one image randomly selected from the source domain and the other from the target domain, noting that larger batches may further

Pipelines	mAP
DMask-RCNN1	0.612
DMask-RCNN2	0.617
AOD-Net + DMask-RCNN1	0.602
AOD-Net + DMask-RCNN2	0.605
MSCNN + Mask-RCNN	0.626
MSCNN + DMask-RCNN1	0.627
MSCNN + DMask-RCNN2	0.634

Table 4. Solution set 2 mAP results on RTTS. Top 3 results are colored in red, green, and blue, respectively.

benefit performance. We also tried to concatenate dehazing pre-processing (AOD-Net and MSCNN) with DMask-RCNN models to form new dehazing-detection cascades.

Table 4 shows the results of solution set 2 (the naming convention is the same as in Table 3), from which we can conclude that:

- the domain-adaptive detector presents a very promising approach, and its performance significantly outperforms the best results in Table 3;[3]
- the power of strong detection models (Mask-RCNN) is fully exploited, given the proper domain adaptation, in contrast to the poor performance of vanilla Mask RCNN in Table 3;
- DMask-RCNN2 is always superior to DMask-RCNN1, showing that the choice of dehazed images as the target domain matters. We make the reasonable hypothesis that the domain discrepancy between dehazed and clean images is smaller than that between hazy and clean images, so DMask-RCNN performs better when the existing domain gap is narrower; and
- the best result in solution set 2 is from a dehazing + detection cascade, with MSCNN as the dehazing module and DMask-RCNN as the detection module and highlighting: **the joint value of dehazing pre-processing and domain adaption**.

5. Conclusion

This paper tackles the challenge of single image dehazing and its extension to object detection in haze. The solutions are proposed from diverse perspectives ranging from novel loss functions (Task 1) to enhanced dehazing-detection cascades as well as domain-adaptive detectors (Task 2). By way of careful experiments, we significantly improve the performance of both tasks, as verified on the RESIDE dataset. We expect further improvements as we continue to study this important dataset and tasks.

[3]By saying that, we also emphasize that Table 3 results have not undergone joint tuning as in [31, 10], so there is potential for further improvements.

Acknowledgements

The study was initially performed as a team project effort in the Machine Learning course (Spring 2018, CSCE 633) of CSE@TAMU, taught by Dr. Zhangyang Wang. We acknowledge Texas A&M High Performance Research Computing (HPRC) for providing some of the computing resources used in this research.

References

[1] K. Tan and J. P. Oakley, "Enhancement of color images in poor visibility conditions," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 2, pp. 788–791, IEEE, 2000.

[2] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–I, IEEE, 2001.

[3] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, *Deep photo: Model-based photograph enhancement and viewing*, vol. 27. ACM, 2008.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[5] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2995–3000, 2014.

[6] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[7] D. Berman, S. Avidan, *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1674–1682, 2016.

[8] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on computer vision*, pp. 154–169, Springer, 2016.

[10] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4770–4778, 2017.

[11] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang, "Studying very low resolution recognition using deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4792–4800, 2016.

[12] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Reside: A benchmark for single image dehazing," *arXiv preprint arXiv:1712.04143*, 2017.

[13] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[14] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1, pp. 598–605, IEEE, 2000.

[15] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[16] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "An all-in-one network for dehazing and beyond," *arXiv preprint arXiv:1707.06543*, 2017.

[17] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "End-to-end united video dehazing and detection," *arXiv preprint arXiv:1709.03919*, 2017.

[18] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pp. 1477–1480, IEEE, 2012.

[19] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[22] D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When image denoising meets high-level vision tasks: A deep learning approach," *arXiv preprint arXiv:1706.04284*, 2017.

[23] B. Cheng, Z. Wang, Z. Zhang, Z. Li, D. Liu, J. Yang, S. Huang, and T. S. Huang, "Robust emotion recognition from low quality and low bit rate video: A deep learning approach," *arXiv preprint arXiv:1709.03126*, 2017.

[24] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2017.

[28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 2980–2988, IEEE, 2017.

[29] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3339–3348, 2018.

[30] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *arXiv preprint arXiv:1409.7495*, 2014.

[31] D. Liu, B. Cheng, Z. Wang, H. Zhang, and T. S. Huang, "Enhance visual recognition under adverse conditions via deep networks," *arXiv preprint arXiv:1712.07732*, 2017.