Video Dehazing Based on Convolutional Neural Network Driven Using Our Collected Dataset

Wang Ziyan
Nanyang Technological University, 50 Nanyang Avenue, Singapore
WANG1351@e.ntu.edu.sg, (+65) 83598315

ABSTRACT With the fast development of automatic driving and video monitoring applications, video dehazing is a vital problem in computer vision because turbid images can have a large impact on their performance. As one of the remarkable techniques, convolutional neural networks have been greatly developed and shown great effectiveness in video dehazing at present. Inspired by deep learning, this paper introduces one of the convolutional neural networks named AOD-Net and collects a dataset to evaluate its dehazing performance. By comparison, the proposed algorithm achieves high SSIM and high PSNR on both our collected dataset and NYU Depth V2 dataset. AOD-Net not only generates well-dehazed images, but also learns very quickly during the training process. This All-in-One Dehazing Network improves visual quality of hazy images to a large extent.

1. Introduction
Nowadays, computers utilize images and videos to carry out complicated tasks. For example, the development of video surveillance and processing facilitates the process in the application of robotics. However, haze, raindrops and rain streaks may severely degrade the quality of images and videos. In particular, unlike other opaque objects, haze does not affect human recognition in general, but indeed causes problems in machine vision systems. As a result, the topic of dehazing system has attracted a large number of researchers.

Dehazing system can be divided into two categories: video dehazing and image dehazing. The most significant difference between image dehazing and other image enhancement technics is that image dehazing handles signal-dependent noises. The blurry areas in image depends on their depth value. Because dehazing process depends on input image, the distinct features of the images and frames must be taken into consideration by dehazing system. The dehazing model is generally used to collect information and represent hazed image data through both explicit and implicit way. For example, image-based dehazing system need to focus on at least lights, textures, and colors. Furthermore, in the dehazing process, an end-to-end system is a system that produces a dehazed image or a dehazed video directly from an input hazed image or video without intermediate outputs.

Recently, many ingenious dehazing algorithms have been designed. An end-to-end dehazing network designed by Li et al. [1] manages to optimize a reconstructed atmospheric scattering model which combines two parameters, global atmospheric light and transmission matrix, and the input image into one formula, \( K(x) \). This network uses a light-weight convolutional network which could easily be embedded in other dehazing models. This image dehazing network is AOD-Net, which stands for All-in-One Dehazing Net. It is the key element of this paper and will be discussed in detail. Li et al. [2] also introduced an end-to-end video dehazing network that maximizes the steadiness of the video. They also trained the network to detect objects and turned out to get more accurate and robust results. Several
criteria for the evaluation of dehazing algorithm are proposed by Li et al. [3] as a benchmarking of single image dehazing.

Many researchers have made dehazing systems in real life applications integrated with deep learning algorithms successfully. Liu et al. [4] recognized dehazing problem as image restoration problem and propounded a new loss function instead of the simple loss function mean squared error (MSE). They also introduced a domain-adaptive mask-RCNN. Zhang and Patel [5] presented a Multi-stream Dense Network (DID-MDN) that uses labels to classify the density of the raindrop and fulfills the de-raining job accordingly. Their network was tested on two synthetic dataset and one real-world dataset. They also introduced Densely Connected Pyramid Dehazing Net-work (DCPDN) [6] on single image, which learns transmission matrix and atmospheric light too. It is an end-to-end system with edge-preserving densely connected encoder-decoder structure that learns transmission map. A joint-discriminator is proposed to monitor transmission map and dehazed image. Qian et al. [7] advocated adversarial network for dehazing job. It has two segments. The generative network is trained to focus on raindrop areas and the discriminative network aims for evaluating local coherence of dehazed image. Ren et al. [8] introduced their VDH-Net to obtain information from close frames with temporal consistency. The net executes dehazing task on video frames by assuming a prior knowledge of global atmospheric light to predict transmission map. Kim et al. [9] presented a new cost function about image and video dehazing tasks. It includes the effect of local image contrast and information loss. As for video dehazing, they also removed flickering effect by making transmission value constant. Zhang et al. [10] proposed a new frame work that estimates optical flow and transmission map to refine video and single image. Especially, they used Markov Random Field (MRF) to obtain the spatial context in their algorithm. Chen et al. [11]’s work could do dehazing job while minimizing artifacts. In order to reduce the visual artifact, they came up with Gradient Residual Minimization (GRM).

The rest of the paper is organized as follows. Section 2 introduces related works, consisting of several existing datasets and our collected dataset, various deep learning models and their applications. Section 3 describes one of the representative convolutional neural networks. Section 4 presents the experimental results and makes a detailed statistical analysis. Section 5 draws to the conclusion that AOD-Net is an outstanding dehazing network and discusses the future works in the fields of research and application.

2. Related Work

2.1 Datasets
From NYU Depth V2 dataset [12]. The NYU Depth V2 dataset contains 27,256 images with RGB and depth parameters taken by Microsoft Kinect. Images of this dataset are from video sequences with 20 to 30 frames per second. Missing values of images are filled. Groups of images are labeled and numbered as chair1, chair2, chair3 and so on.

From TUM RGB-D dataset [13]. This dataset was built to evaluate SLAM (Simultaneous localization and mapping) system. It includes RGB, depth data captured by Microsoft Kinect and also ground truth data of the Kinect sensor’s path. The data comes from video with 680*680 resolution and 30 frames per second. They also provide an evaluation algorithm to assess the predicted camera path.

From ILSVRC2015 VID dataset [14]. This dataset is built to do large-scale object recognition. There are two labels in the dataset: 1) object detection: binary flag indicating the existence of an object 2) object localization: bounding box indicating the size and location of an object. For object detection, it has 200 fully labeled categories with 1.2 million training images, 50 thousand validation images and 100 thousand test images. For object localization, it has 1000 categories.

From RESIDE dataset [3]. The Realistic Single Image Dehazing (RESIDE) dataset uses clean images from NYU Depth V2 dataset and Middlebury stereo dataset to synthesis hazed image. When generating hazed image, different parameters in atmospheric scattering model are set. Overall it has 13,990 synthesized hazed images in training set. This contains 520 testing image that spans indoor to outdoor scenario.
From Middlebury stereo dataset[15]. The Middlebury stereo dataset contains 9 images of a stereo scene of a certain image size and 2 disparity maps as ground truth. It has three types of image size: full, half and quarter. This dataset overall has 2 scenes and 6 image sizes.

Our collected dataset. This dataset collects 200 outdoor images with high resolution. Majority of the images are natural scenes, including lakes, grasslands and mountains. Others are buildings in cities. They are manually added haze and then tested in All-in-One Dehazing Network for single image. Some examples of hazed images and dehazed images from our dataset is illustrated in Figure 1.

![Figure 1. Examples of our collected dataset.](image)

**2.2 Deep learning models**

Modern researchers combined dehazing systems with new deep learning algorithms. Zhang and Patel[6] designed DCPDN for the concurrent improvement of transmission map, atmospheric light and dehazing. Encoder-decoder structure and multi-level pooling within the network are used to deal with transmission map. They also introduced a new loss variable, edge-preserving loss. A joint-discriminator from Generative Adversarial Network (GAN) is used to estimate the structural correlation of transmission map and dehazed images.

The traditional method to describe hazing process is building the atmospheric scattering model [16]. A hazed image uniquely maps to its clean image, atmospheric light and transmission matrix. AOD-Net of single image dehazing proposed by Li et al. [1] combines atmospheric light and transmission matrix into one feature K by utilizing the information of clean image to reduce error. It is an end-to-end model consisting of two parts: a K-estimation module with 5 convolutional layers which regresses clean image from hazed image, and a clean image generation module which produces dehazed image directly from an input hazed image without intermediate outputs. They regarded NYU depth dataset as ground truth to synthesize hazed image for the training and testing dataset. EVDD-Net Model of video dehazing in Li et al. [2] gets inspiration from Li et al. [1]. It has 3 levels of fusion from input and K-estimation to output. This light-weight network, integrated with a video object detection model, converges very fast. The result shows that considering 5 consecutive frames can get the best result on object detection. K-level fusion turned out to perform better than I-level fusion and J-level fusion. This fusion level separately five input frames and then concatenate them after convolutional layer 2 if 5 frames are taken. Both natural and synthetic video dataset are used.

To achieve higher precision and better results, Li et al. [3] used PSNR/SSIM as full-reference parameters, spatial-spectral entropy-based quality (SSEQ) [17] blind image integrity notator using DCT statistics (BLIINDS-II) [17] as no-reference parameters to evaluate each state-of-the-art algorithm. Qian et al. [7] used attentive GAN to remove raindrops. The key idea to cope with the problem is that using attentive-recurrent network to generate an attention map as raindrop areas are not given. Then contextual autoencoder focuses on rainy regions and refines the image. The discriminator of the network uses several convolutional layers to get and check the result. DID-MDN network proposed by Zhang and Patel [5] contains following two parts: firstly, a residual-aware rain-density classifier which classifies
the severity of a certain image into three levels, and secondly, a multi-stream densely connected de-raining network which makes use of the information from the classifier to generate outputs of the network, clean images without rain streaks. Liu et al. [4] proposed several loss functions and they provided two solution sets to solve dehazing for detection.

Apart from dehazing video frames by assuming global semantic as prior, Ren et al. [8] showed that consecutive frames could maintain consistence without any assistance. They created a synthesis dataset based on NYU depth V2 dataset. The network they introduced uses a stack of 5 video frames to predict transmission maps of 3 video frames in the middle. They applied encoder-decoder structure in this network. Kim et al. [9] introduced an algorithm that optimizes contrast of hazy images and videos. They first used quadtree-based subdivision to identify atmospheric light, and then predicted transmission values for dehazing. Zhang et al. [10] used human visualization system (HVS) and Markov Random Field (MRF) to obtain temporal consistency and spatial consistency to implement the dehazing job. Their algorithm is very efficient as it reduces the amount of interacting data and constrains minimum input data. Chen et al. [11] utilized local prior of the hazy image to predict its transmission map. In addition, they used GRM to alleviate the exaggeration of artifacts when producing clean image. They also presented Total Generalized Variation (TGV) [20] regularization to improve image refining skills. In all these papers, a variety of novel deep learning algorithms and models are designed and presented. They provide different methods to train the weights with a visible layer and a hidden layer that correspond to the input hazed image and the output clean image respectively.

2.3 Applications
Recently, dehazing systems have been successfully implemented in many real-life applications. The network introduced by Ren et al. [8] generates outstanding results of video dehazing without parameter tuning or image aligning. Their VHD-Net estimates transmission map and semantic segmentation to improve dehazing capability. The fast algorithm designed by Kim et al. [9] is able to do dehazing on either single image or a video. This algorithm avoids information loss while dehazing. AOD-Net proposed by Li et al. [1] puts efforts on object recognition and detection on single image with parameter tuning. They combine global atmospheric light and transmission matrix together. Their group’s EVD-net [2] on video dehazing also boosts object detection task and the network achieves higher average precision. The result can be further used on applications like autonomous driving and video monitoring.

During the evaluation, Qian et al. [7] focused on the topic of removing raindrops on single image even in extremely bad conditions. DID-MDN network proposed by Zhang and Patel [5] also aims for raindrop removal. Their key idea is to classify the seriousness of rain streaks into 3 levels before de-raining. DCPDN designed by Zhang and Patel [6] solves the problem of single image dehazing with different methods. Liu et al. [4] came up with a new loss function and they trained their network to improve object detection performance on single image with domain-adaptive detector. Li et al. [3] worked on the same dataset as Liu et al. [4] They summarized several algorithms and announced that further work would focus on no-reference metric developing. Zhang et al. [10] improved the way predicting transmission map, and their network behaves not satisfactory when the colour of image objects is close to atmospheric light. Chen et al. [11] introduced a new image refinement algorithm to dehaze image to suppress visual artifacts. In conclusion, all these proposed state-of-the-art approaches in dehazing applications mentioned in the previous part outperform other methods in terms of accuracy and efficiency.

3. Convolutional Neural Network: AOD-Net
In this section, AOD-Net as one of the convolutional neural networks for single image dehazing and implementation will be introduced. Details about physical models, network design, code and its environment will be provided.

3.1 Transformed Atmospheric Scattering Model
The classical characterization of hazy image generation is the atmospheric scattering model denoted by
\[ I(x) = J(x)t(x) + A[1 - t(x)] \]  \hspace{1cm} (1)

where \( I(x) \) is the hazy observed image, \( J(x) \) is the scene radiance (clean image), \( A \) is the global atmospheric light [3]. The critical medium transmission matrix \( t(x) \) is calculated by

\[ t(x) = e^{-\beta d(x)} \]  \hspace{1cm} (2)

where \( d(x) \) is the scene depth and \( \beta \) is the scattering coefficient. \( t(x) \) represents the light that reaches the camera without being scattered by particles in the atmosphere. The key of image dehazing is to solve \( J(x), A \) and \( t(x) \) given \( I(x) \).

Equation (3) is the identical transformation of equation (1) to obtain clean image. AOD-Net applies a tricky method to minimize reconstruction errors. Both \( A \) and \( t(x) \) are combined into one variable \( K(x) \) [3] denoted by Equation (4).

\[ K(x) = \frac{1}{t(x)} I(x) - A \frac{1}{t(x)} + A \]  \hspace{1cm} (3)

\[ K(x) = \frac{1}{t(x)} (I(x) - A)(A - b) \]  \hspace{1cm} (4)

\( K(x) \) is dependent on \( I(x) \). \( J(x) \) becomes \( J(x) = K(x)I(x) - K(x) + b \), where \( b \) is a constant bias which has default value 1. The model is an input-adaptive deep model, which indicates that as input images change, parameters change accordingly. The model is trained to minimize the errors between the output \( J(x) \) and ground truth clean image.

3.2 Network Design

The AOD-Net has two modules: a k-estimation module and a clean image generation module. AOD-Net is implemented by convolutional layers of deep learning method. The Figure 2 shows how a hazy image is purified into a clean image through this end-to-end system.

![Figure 2. Overview of AOD-Net structure.](image)

3.2.1 K-estimation Module

The k-estimation module is the crucial part of the network. It estimates the image depth and haze level. To obtain accurate \( K(x) \), 5 parallel structured convolutional layers and 3 concatenation layers are used. The CNN uses ReLU as activation

![Figure 3(a). Structure of K-estimation Module.](image)
Figure 3(b). Code of K-estimation Module.

function. The structure and its corresponding code are shown in Figure 3, we label them as $conv(i), i = 1, 2, 3, 4, 5$ and $concat(j), j = 1, 2, 3$. Concat(1) concatenates the features from $conv(1)$ and $conv(2)$, Concat(2) concatenates the features from $conv(2)$ and $conv(3)$, Concat(3) concatenates those from all the preceding four layers, $conv(1)$, $conv(2)$, $conv(3)$ and $conv(4)$. Finally, $conv(5)$ is the estimated result of K.

3.2.2 Clean Image Generation Module

Clean image generation module is simpler. It carries on the result K from the previous K estimation module and uses the equation $J(x) = K(x)I(x) - K(x) + b$ (assign default value 1 to $b$) to calculate $J(x)$. The code uses an element-wise multiplication layer and two element-wise addition layers as shown in Figure 4.

```
def forward(self, x):
    source = []
    source.append(x)

    clean_image = self.relu(x3 + x) - x3 + 1
    return clean_image
```

Figure4. Code for clean image generation module.

3.3 Code

AOD-Net is an extremely lightweight model which takes less than 10KB, it is publicly available at Github1. The network contains 4 modules: train, net, dataloader and dehaze. NYU Depth V2 dataset[12] is used as training set. After each training period, data will be recorded so the performance of the net improves gradually. The flowchart of data processing and details of the code will be presented in the following session.

3.3.1 Net

Net module, consisting of K-estimation module and clean image generation module, has been discussed before. It is the key of the AOD-Net which focuses on unified parameter $K(x)$ to minimize reconstruction errors. Figure3(b) and Figure4 contain the main part of the net module.

3.3.2 Dataloader

Dataloader module prepares images for training and dehazing and stores images in appropriate location. In training dataset, several hazy images which share the same prefix in their file names are mapped to one clean image. Code in Figure5 maps hazy images as values into a dictionary whose keys are file names of corresponding clean images. The module divides training images into two groups as shown in Figure5: 90 percent of hazy images for training, and the rest 10 percent for testing.

1https://github.com/TheFairBear/PyTorch-Image-Dehazing.
In addition, as shown in Figure 6, dataloader module also resizes images and applies a permutation matrix for data consistency. Finally, hazy images are transferred to the network in a unified form for training and testing.

3.3.3 Train

Train module contributes to the excellent performance of AOD-Net. It extracts training dataset from default path. After each epoch of training, the record will be stored thus the network makes progress in dehazing performance. The source code is written in GPU functions (CUDA) shown in Figure 7. The network uses MSE (mean squared error) function as its loss function. After certain epochs, value of loss function is printed thus user can observe the improvement of the network. Users are allowed to adjust parameters such as batch size and the number of epochs between two printing instructions.

---

```python
for image in image_list_haze:
    image = image.split('/')[-1]
    key = image.split('_')[0] + '_' + image.split('_')[1] + '.jpg'
    if key in tmp_dict.keys():
        tmp_dict[key].append(image)
    else:
        tmp_dict[key] = [image]
        tmp_dict[key].append(image)
len_keys = len(tmp_dict.keys())
for i in range(len_keys):
    if i < len_keys/10:
        train_keys.append(list(tmp_dict.keys())[i])
else:
    val_keys.append(list(tmp_dict.keys())[i])
```

---

Figure 5. Part of code in dataloader module.

---

```python
def __getitem__(self, index):
    data_orig_path, data_hazy_path = self.data_list[index]
    data_orig = Image.open(data_orig_path)
    data_hazy = Image.open(data_hazy_path)
    data_orig = data_orig.resize((400, 640), Image.ANTIALIAS)
    data_hazy = data_hazy.resize((400, 640), Image.ANTIALIAS)
    data_orig = (np.array(data_orig)/255.0)
    data_hazy = (np.array(data_hazy)/255.0)
    data_orig = torch.from_numpy(data_orig).float()
    data_hazy = torch.from_numpy(data_hazy).float()
    return data_orig.permute(2,0,1), data_hazy.permute(2,0,1)
```

---

Figure 6. Data management in dataloader module.

---

```python
img_orig = img_orig.read()
img_haze = img_haze.read()

clean_image = dense_settings(img_haze)
loss = criterion(clean_image, img_orig)
optimizer.zero_grad()
loss.backward()
torch.nn.utils.clip_grad_norm_(denet.parameters(), config.clip_grad_norm)
optimizer.step()
if (iteration) % config.display_iter == 0:
    print("loss at iteration", iteration, ".", loss.item())
if (iteration) % config.snapshot_iter == 0:
    torch.save(denet.state_dict(), config.snapshot_folder + "epoch" + str(epoch) + ".pth")
```

---

Figure 7. Part of train module and its details.
3.3.4 Dehazing

Having been thoroughly trained, the network generates an output of clean image given an input of a hazy image. This module manipulates input image as dataloader module does. The flowchart (Figure 8) below introduces how a clean image is generated through training and testing.

![Flowchart](image)

Figure 8. Flowchart, where dehaze module calls the well trained network. Details of the network are provided in Figure 2 and Figure 3(a).

3.4 Experimental Setting

The deep learning module AOD-Net is implemented on pytorch, which requires at least python 3 and pytorch version 0.4. It is recommended to use GPU at least 8GB as the code runs much faster than using CPU. If the computer operating system does not support CUDA (MacOS), researchers are required to modify the code to allow CPU to execute the instructions. CPU of 2.4Hz Inter Core i5 is capable of running the code with 16GB random access memory.

4. Experimental results and analysis

In this section, the performance of AOD-Net on both our own dataset and NYU Depth V2 dataset will be presented. PSNR, peak signal-to-noise ratio and SSIM, structural similarity, are used for the evaluation between ground truth images and dehazed images by the network. Average PSNR and SSIM results on two datasets are recorded in Table 1. AOD-Net is an end-to-end system which is trained under MSE loss, so the PSNR value should be high. The outstanding performance of measuring global atmospheric light promises AOD-Net a high SSIM value. We also plotted the graph of loss versus training iteration number to show that AOD-Net converges fast as well as provides good results.

| Dataset               | PSNR    | SSIM    |
|-----------------------|---------|---------|
| NYU Depth V2          | 20.4987 | 0.9683  |
| Our collected dataset | 22.7822 | 0.9811  |

Table 1. Average PSNR and SSIM results on two datasets

As shown in Table 1, a high PSNR value fairly means high quality of the images. It has an inverse proportional logarithmic relationship to the mean squared error. SSIM measures the general similarity between ground truth image and the generated dehazed image. The PSNR and SSIM are approximately equal to those showed by Li et al [1], which means that AOD-Net is acting better than other traditional dehazing methods.

The flaw of using only quantitative evaluation is that images with same PSNR may differ significantly. Besides, human perception of image quality is not as sensitive as machine calculated values. SSIM presents the overall performance of dehazing network more accurate than PSNR to some extent, but qualitative analysis is still indispensable for evaluating results.
In Figure 9, three images in the first row are dehazed examples from our collected dataset, and three images in the second row are picked from NYU Depth V2 dataset. From the comparison between hazy images and dehazed images, we are able to derive that AOD-Net generates images with high resolution, eliminates blurry and hazy areas and turns these areas into rich colors. The output images tend to be even brighter than original images.

Apart from producing excellent outputs, another advantage of AOD-Net is that the network learns very quickly and has an extremely lightweight model. AOD-Net is able to carry out good enough results after 10 training epochs. The line chart in Figure 10 shows that AOD-Net is almost well trained after only two epochs (one epoch is around 3000 iterations; 6000 iterations means 2 epochs). MSE loss drops to around 0.01 shortly. In the further training epochs, the network gradually converges to a perfect structure. It is observable that the network is already able to produce good results after two epochs’ training, reaching an MSE loss of less than 0.01.

However, AOD-Net and other traditional image dehazing networks are not omnipotent. For example, facing an image with a small piece of stained area while all other parts are exactly the same as ground truth image, AOD-Net does not perform well. The reason is that AOD-Net calculates atmospheric light, which tends to be uniform. The terrible results in Figure 11 show that AOD-Net, as well as other traditional methods, has difficulty in refining images with abrupt stains. In these cases, AOD-Net turns white objects into other colors incorrectly and stains are not cleared.

5. Conclusion

Image dehazing is a vital technique in computer vision field. A variety of state-of-the-art dehazing methods based on convolutional neural network have been designed, and AOD-Net, All-in-One
Dehazing Network proposed by Li et al [1] turns out to be very effective when it deals with uniform spots, for example, haze. We used both our collected dataset and NYU Depth V2 dataset to evaluate the AOD-Net. Its end-to-end structure reduces intermediate error, promises quick convergence and guarantees excellent results.

References
[1] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. 2017. AOD-Net: All-in-One Dehazing Network. Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob, 4780–4788.
[2] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. 2017. End-to-End United Video Dehazing and Detection, (1), 7016–7023.
[3] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. 2019. Benchmarking Single Image Dehazing and beyond. IEEE Transactions on Image Processing, 28(1), 492–505.
[4] Liu, Y., Zhao, G., Gong, B., Li, Y., Raj, R., Goel, N., … Tao, D. 2018. Improved Techniques for Learning to Dehaze and Beyond: A Collective Study.
[5] Zhang, H., & Patel, V. M. 2018. Density-Aware Single Image De-raining Using a Multi-stream Dense Network. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 695–704.
[6] Zhang, H., & Patel, V. M. 2018. Densely Connected Pyramid Dehazing Network. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3194–3203.
[7] Qian, R., Tan, R. T., Yang, W., Su, J., & Liu, J. 2018. Attentive Generative Adversarial Network for Raindrop Removal from A Single Image. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2482–2491. https://doi.org/10.1109/CVPR.2018.00263
[8] Ren, W., Cao, X., Zhang, J., Xu, X., Meng, G., Ma, L., & Liu, W. 2019. Deep Video Dehazing with Semantic Segmentation. IEEE Transactions on Image Processing, 28(4), 1895–1908. https://doi.org/10.1109/TIP.2018.2876178
[9] Kim, J. H., Jang, W. D., Sim, J. Y., & Kim, C. S. 2013. Optimized contrast enhancement for real-time image and video dehazing. Journal of Visual Communication and Image Representation, 24(3), 410–425. https://doi.org/10.1016/j.jvci.2013.02.004
[10] Zhang, J., Li, L., Zhang, Y., Yang, G., Cao, X., & Sun, J. 2011. Video dehazing with spatial and temporal coherence. Visual Computer, 27(6–8), 749–757. https://doi.org/10.1007/s00371-011-0569-8
[11] Chen, C., Do, M. N., & Wang, J. 2016. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9906 LNCS, 576–591. https://doi.org/10.1007/978-3-319-46475-6_36
[12] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. 2012. Indoor segmentation and support inference from RGBD images. In European Conference on Computer Vision, pages 746–760. Springer.
[13] Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; and Cremers, D. 2012. A benchmark for the evaluation of RGB-D slam systems. In Proc. of the International Conference on Intelligent Robot Systems.
[14] Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. 2015. ImageNet large scale visual recognition challenge. International Journal of Computer Vision 115(3):211–252.
[15] D. Scharstein and R. Szeliski.2003. “High-accuracy stereo depth maps using structured light,” in Computer Vision and Pattern Recognition, 2003. Proceedings 2003 IEEE Computer Society Conference on, vol. 1. IEEE, 2003, I–I.
[16] E. J. McCartney. 1976. Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421
[17] L. Liu, B. Liu, H. Huang, and A. C. Bovik. 2014. “No-reference image quality assessment based on spatial and spectral entropies,” *Signal Processing: Image Communication*, vol. 29, no. 8, 856–863.

[18] M. A. Saad, A. C. Bovik, and C. Charrier. 2012. “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE transactions on Image Processing*, vol. 21, no. 8, 3339–3352.

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

[20] Bredies, K., Kunisch, K., Pock, T. 2010. Total generalized variation. *SIAM Journal on Imaging Sciences* 3(3), 492–526.

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

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

[23] S. G. Narasimhan and S. K. Nayar. 2002. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3), 233–254.