SINGLE-IMAGE DEHAZING ON AERIAL IMAGERY USING CONVOLUTIONAL NEURAL NETWORKS

M. Madadikhaljan [1,2], R. Bahmanyar [1], S. M. Azimi [1], P. Reinartz [1], U. Sörgel [2]

[1] DLR, German Aerospace Center, Earth Observation Center (EOC), Münchener Str. 20, 82234 Wessling, Germany-
(Mojgan.Madadikhaljan, Reza.Bahmanyar, Seyedmajid.Azimi, Peter.Reinartz)@dlr.de
[2] Institute of Photogrammetry (ifp), University of Stuttgart, Germany - soergel@ifp.uni-stuttgart.de

KEY WORDS: Single-image Dehazing, Convolutional Neural Networks, Aerial Imagery, Haze Removal, Hazy Image Generation

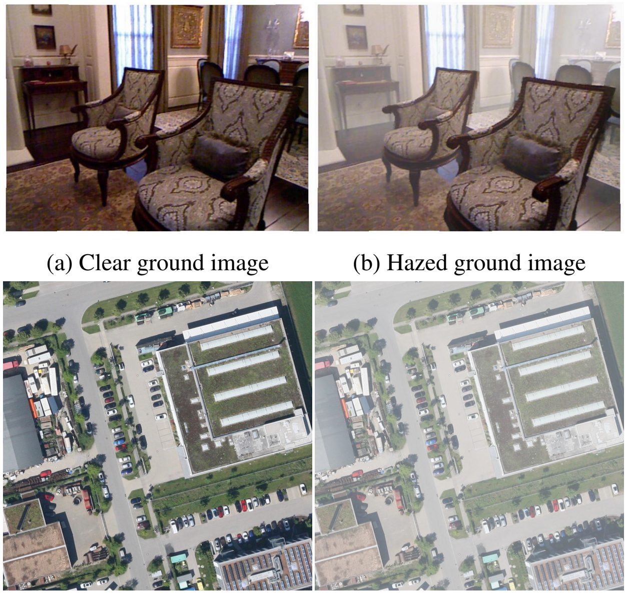
ABSTRACT:

Haze contains floating particles in the air which can result in image quality degradation and visibility reduction in airborne data. Haze removal task has several applications in image enhancement and can improve the performance of automatic image analysis systems, namely object detection and segmentation. Unlike rich haze removal literature in ground imagery, there is a lack of methods specifically designed for aerial imagery, considering the fact that there is a characteristic difference between the aerial imagery domain and ground one. In this paper, we propose a method to dehaze aerial images using Convolutional Neural Networks (CNNs). Currently, there is no available data for dehazing methods in aerial imagery. To address this issue, we have created a synthetically-hazed aerial image dataset to train the neural network on aerial hazy image dataset. We train All-in-One dehazing network (AOD-Net) as the base approach on hazy aerial images and compare the performance of our proposed approach against the classical model. We have tested our model on natural as well as the synthetically-hazed aerial images. Both qualitative and quantitative results of the adapted network show an improvement in dehazing results. We show that the adapted AOD-Net on our aerial image test set increases PSNR and SSim by 2.2% and 9%, respectively.

1. INTRODUCTION

Haze is an atmospheric phenomenon in which there are tiny particles coming from dust, volcanic ashes, foliage exudation, combustion products, etc. which have the size varying from 0.01 to 10 micro meters (McCartney, 1976). When utilizing aerial images for different applications, on the one hand, haze could be a reason to decrease the performance of automatic image analysis systems such as image recognition, object detection, segmentation and tracking (Li et al., 2017, Azimi et al., 2018b, Azimi et al., 2018a). On the other hand, it is not always possible to capture haze-free images as some amount of haze can always be floating in the air. Therefore, various dehazing methods have been proposed in order to reconstruct haze-free images from single hazy images.

Most of these algorithms deal with ground imagery, whereas the techniques focused on dehazing of aerial images are a few. The characteristics of the ground and aerial imagery scenes are very different. For instance, the sensor distance to the objects within the scenes (depth), the image contents, the ground resolution, and the ratio between the depth of each pixel and the height differences between the objects. Figure 1 exemplifies ground and aerial images with their hazy variants. As it can be observed, while haze is homogeneously spread over the aerial scene, in the room image, haze increases by the depth of the image pixels. In short, there is a need for adapting the ground imagery dehazing algorithms for developing new aerial image dehazing algorithms.



(a) Clear ground image (b) Hazed ground image

(c) Clear aerial image (d) hazed aerial image

Figure 1: Sample ground image and aerial image with homogeneous hazy versions

There are several single-image dehazing methods based on prior knowledge and *convolutional neural network (CNN)*, which can yield a dehazed and clear image from a single hazy image as input. As one of the primary and effective methods, dark channel prior-based dehazing technique (Kaiming He, 2011) benefits from the statistics of the outdoor scenes, in which at least in one color channel there exist pixels with very low intensities. Taking this prior as a starting point and together with estimating the depth information, the dark channel method recovers the haze-free image from a single input hazy image. Recently, *conditional generative adversarial networks (cGAN)* was proposed for single-image dehazing. The authors in (Li et al., 2018) used a cGAN network architecture with a generator to create haze-free images and a discriminator to identify whether the image is realistic or not. The authors also added a perceptual regularization based on the VGG network's feature-maps and applied L1 regularized gradient prior in order to avoid artifacts and color distortions in the dehazed images. A proposed method based on CycleGAN (Engin et al., 2018) uses two generators and two discriminators in the network to add or remove haze to the images. This method further takes advantage of cycle-consistency and perceptual losses to refine texture information and improve the final results. The authors in (Zhang et al., 2018) proposed a perceptual pyramid deep network for image dehazing, where the network architecture benefits from dense and residual blocks. It is composed of an encoder to map the image into a latent feature space and a decoder to transfer the features back from the latent space and generate the haze-free image. Not only the deep neural networks, but the shallow neural networks with quite simple structures have shown promising results in the image dehazing scenarios. For example, the widely-used All-in-One Dehazing Network (AOD-Net) (Li et al., 2017) achieves very high accuracy by its simple network design and a reformed mathematical formulation of the haze equation.

Due to the outperforming results of the AOD-Net between state of the art methods, we have used this dehazing network in our experiments. The original AOD-Net was trained on a synthetically hazed ground image datasets such as NYU depth V2 (Silberman et al., 2012). In order to apply AOD-Net to aerial imagery, one could either use the pre-trained model or train the model on an aerial image haze dataset. Due to the aforementioned differences between the ground and aerial images, training the network should result in a more accurate haze removal.

In order to train AOD-Net, we need an aerial image haze dataset containing hazy images and their ground truth. To the best of our knowledge, there is no publicly available dehazing aerial image dataset. Therefore, for our experiments, we create a synthetically-hazed aerial image dataset. It is not practically possible to have naturally hazed aerial image dataset. In the aerial imagery, the atmospheric conditions of the different parts of the imaging area could be different. Thus, the captured images are either with or without haze. While for the haze-free images there is no hazy equivalent, for the hazy images there is no haze-free image as ground truth. Having a second acquisition to fulfill the required data is very complicated and demanding.

Our aerial haze dataset consists of haze-free aerial images acquired by German Aerospace Center (DLR) in 2018 in the framework of VABENE++ project using the 3K camera system (Kurz et al., 2011) mounted on a helicopter flying over the city of Munich, Germany.

As a prerequisite to create synthetic hazy images, the depth of each pixel (the euclidean distance of the corresponding pixel to the object point on the ground) should be computed. To this end, the camera information from the flight and the co-linearity equation are used. Using the computed depth map, the hazy images are generated based on the atmospheric scattering equation which is explained in Section 2.1. In the next step, we train AOD-Net on the created aerial haze dataset. In order to evaluate the trained models, we split the dataset into train and test sets. The performance of the dehazing models is then evaluated using the images in the test set and some natural hazy images. Training AOD-Net on our dataset shows a significant improvement in both PSNR and SSim metrics as compared to directly applying the pre-trained model.

We have organized our paper as follows. In Section 2, we explain the generation of new dataset together with assumptions and formulations in detail. We shortly describe AOD-Net in Section 3 and then we present the experiments and results followed by discussion in Section 4. Finally, we conclude the paper in Section 5.

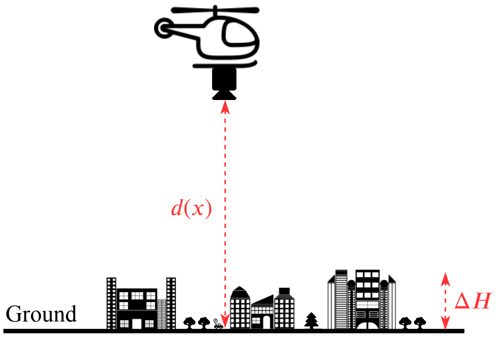


Figure 2: Depth of the pixels ($d(x)$) and the height variations on the ground ΔH in aerial imagery (note that $\Delta H \ll d(x)$).

2. SYNTHETICALLY HAZED AERIAL IMAGE DATASET GENERATION

Haze in the image can be seen in the locations where there is a gray effect due to the existing particles which cause the light to be scattered. Depending on how the image has been captured, the haze appears in the image differently. The objects in the ground imagery are relatively closer to the camera compared to the aerial case. Hence, the spatial resolution of a ground image is significantly higher than aerial one. On the other hand, the scene coverage in aerial imagery is often larger than ground images. When capturing aerial images on the airplanes, the cameras are moving most of the time, so blurring effects are unavoidable. When flying over a city, the weather or atmospheric conditions may differ in different parts of the city, especially when flying over megacities or even farmlands. Similar objects in the ground imagery appear differently in aerial imagery. In fact, one of the most important dissimilarities between these two imagery domains is the ratio of height (or depth) differences of the objects in the scene and their distance to the camera. For instance, in flights with 1000m altitude, the height difference between building and roads in the image would be around 5 meters on average in the cities with not much of skyscrapers. However, when taking a photo from a scene in a room, the ratio of height/depth differences and the distance to the camera is close to 1:1 (see Figure 2).

All of the aforementioned deviations of the two imagery domains create the need to revise the hazy image generation pattern. Therefore, we first go through the atmospheric scattering equation which models the generation of haze on image in Section 2.1. Afterward, we mathematically compute the depth map of the images as a prerequisite to hazy image generation in Section 2.2. We finally provide the generated hazy images in Section 2.3.

2.1 Atmospheric Scattering Model

In order to reconstruct the haze in an image, the Atmospheric Scattering Model is defined as

$$I_{x,y} = J_{x,y}t_{x,y} + A(1 - t_{x,y}) \quad (1)$$

where I = Received radiance of a hazy image,
J = True scene radiance,
t = Medium transmission,
A = Global atmospheric light,
x, y = Pixel location.

Equation (1) shows that the true scene radiance is attenuated when traveling through the air to reach the camera. The illumination from the atmosphere affects the traveling light beam by adding the Airlight (A) term. In homogeneous atmospheres, the medium transmission term can be calculated by

$$t_{x,y} = e^{-\beta d_{x,y}}, \tag{2}$$

where β = Scattering coefficient of the atmosphere
$d_{x,y}$ = Depth of each pixel.

Almost all of the single-image dehazing algorithms (Li et al., 2017, Li et al., 2018, Engin et al., 2018, Zhang et al., 2018, Min et al., 2019) use this model to reconstruct haze-free image by estimating the transmission map, airlight or combination of them.

2.2 Depth Image Generation

As shown in Equation (2), the key to generate hazy images is the depth map. Assuming that the depth map is provided, given β value, the transmission map for each pixel can be calculated. Then, we put the transmission values into Equation (2) and for different values of A, we construct different levels of hazy images.

In order to have the depth for each pixel, we need the 3D world coordinates of each pixel (X, Y, Z) on the ground. Then, the depth or Euclidean distance of the object point to the principal point of the camera can be calculated using

$$d_{x,y} = \sqrt{(X_{x,y} - X_c)^2 + (Y_{x,y} - Y_c)^2 + (Z_{x,y} - Z_c)^2}, \tag{3}$$

where $(X_{x,y}, Y_{x,y}, Z_{x,y})$ = 3D world coordinates of each pixel,
(X_c, Y_c, Z_c) = 3D world coordinates of principal point,
$d_{x,y}$ = Depth of each pixel.

Equation (3) shows that to have the depth of each pixel, it is necessary to have 3D coordinates of each pixel together with the 3D coordinate of the principal point of the camera in the same coordinate system, namely world coordinate system. There are two strategies to obtain the 3D position of each pixel on the ground:

1. The first one is to use Digital Surface Model (DSM) of the corresponding image. DSM is a geo-referenced raster data of the same size and resolution to the image containing ground height of each pixel. As it is geo-referenced, we have X and Y coordinates of the pixels and Z is stored inside the pixel values of the DSM.
2. As the second option, we can use the camera information such as focal length, flight height, rotation angles, etc. together with the ground height of each pixel (Z) to calculate X and Y for each pixel (using well-known "Collinearity Equation").

In our case, the DSM of the whole city is available, but we face the problem of matching the position and resolution of each image with its corresponding area on whole DSM data. Hence, we need to crop each area of the DSM to cover each image location and match the spatial resolution of the pixels inside the DSM with the ones of image. Therefore, we use a simple strategy to obtain the ground coordinates and therefore the depth map. As can be seen from Figure 2, the flight height is significantly larger than the height differences of objects *e.g.*, buildings on the ground. Consequently, when computing the Euclidean distance, there is no significant difference in depth for the neighboring objects. Thus, the only varying depth parameter is the distance from the nadir line to the ground. Therefore, it is logical to consider the height of all areas to be similar.

According to Figure 2 we have

$$\Delta H \ll d(x). \tag{4}$$

Therefore,

$$H_i = Z_{avrg}, \forall H_i, \tag{5}$$

where H_i = Ground height of the image pixels,
$Z_{avrg.}$ = The average ground height of the region.

Considering the ground height of all pixels in image similar to each other, i.e., the average height of the region, X and Y ground coordinates of the pixels can be computed by

$$X_{x,y} = X_c + (Z_{x,y} - Z_c)\frac{r_{11}(x - x_c) + r_{21}(y - y_c) - r_{31}c}{r_{13}(x - x_c) + r_{23}(y - y_c) - r_{33}c}, \tag{6}$$

$$Y_{x,y} = Y_c + (Z_{x,y} - Z_c)\frac{r_{12}(x - x_c) + r_{22}(y - y_c) - r_{32}c}{r_{13}(x - x_c) + r_{23}(y - y_c) - r_{33}c}, \tag{7}$$

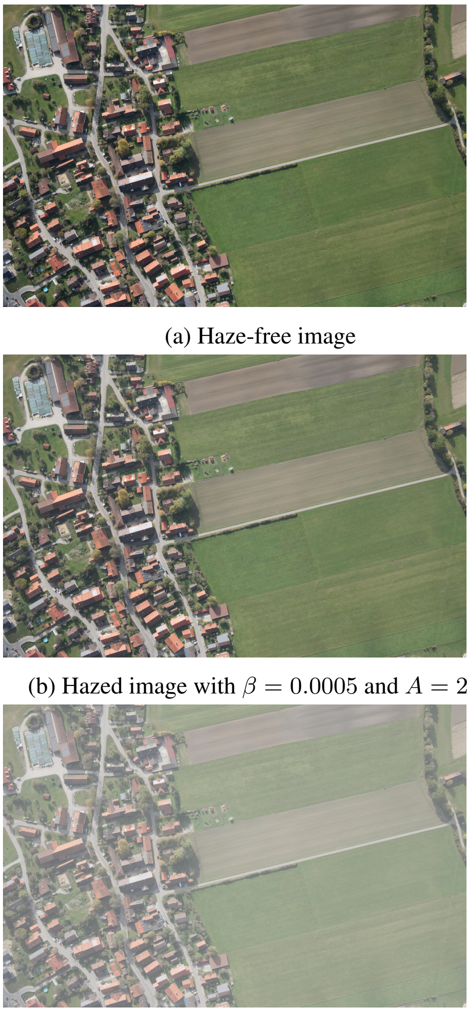
where (x_c, y_c) = pixel coordinates of principal point,
$r_{11}, r_{12}, ..., r_{33}$ = Rotations of the camera at the capture time,
c = Focal length of the camera,
$Z_{x,y} = Z_{avrg.}, \forall x, y$ in the image.

Since all the camera parameters are available for all the images, by taking the average altitude of the region as height of all pixels, we can calculate the world coordinates of each pixel and use them in Equation (3) to obtain the depth map.

2.3 Hazy Images

We can then insert different values for β and A and create different hazy images. However, there are two points to be noted. Firstly, the depth in our images is around 1000 meters, which is totally different from the ground imagery scenarios. As a result, the values that should be inserted for the medium transmission and global atmospheric light should be different as well. In our case, we chose these values $\beta = \{0.0005, 0.0015, ..., 0.002\}$ and $A = \{230, 240, 250\}$.

Secondly, by assuming the atmosphere to be homogeneous and the heights to be similar, the haze on the images appear homogeneously. In our case, in the images collected from different flight campaigns, the haze is almost always homogeneous. This is valid for the cases where there is no pollution



(a) Haze-free image

(b) Hazed image with $\beta = 0.0005$ and $A = 230$

(c) Hazed image with $\beta = 0.001$ and $A = 250$

Figure 3: Synthetically and homogeneously hazed aerial image

source around because when this is not the case. For example, when there is a factory in the area, the smoke coming out of the smokestack of the factory causes the density of the haze or pollution to be greater in some parts; therefore, the assumption of homogeneous atmosphere or haze does not hold anymore. This applies also for the flights over mountains and valleys, which there might be haze on valley, but on top of the mountain is clear, making the haze not homogeneous In Figure 3a, Figure 3b and Figure 3c, you can see a haze-free image, together with two hazy versions of it.

The training images in aerial hazy image dataset contain hazy images for a single image in 9 different levels. In our experiments, 140 images are chosen as training and 17 images as test sets. Flight height is around 1600 meters above the sea level, where the average ground height is 600 meters. The view points are almost nadir looking and the images consist of urban, rural, and some forest areas. The camera in this flight is a hyperspectral camera with 30 cm ground spatial resolution and 88 mm focal length.

3. INTRODUCTION TO AOD-NET

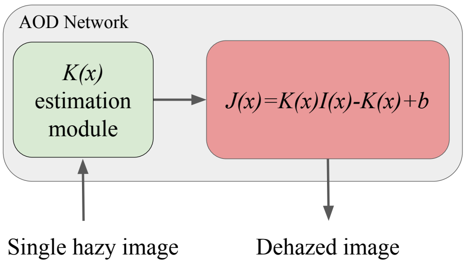


Figure 4: The General Structure Of AOD-Net

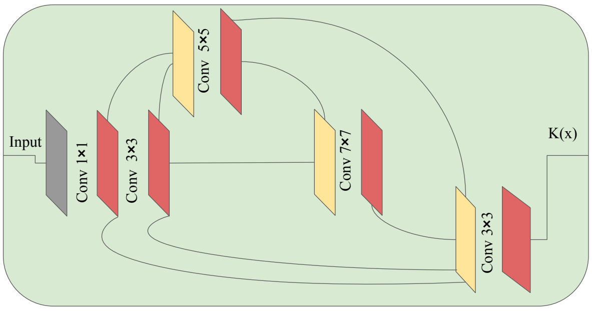


Figure 5: The $K(x)$ estimation module

All-in-One Dehazing Network called AOD-Net (Li et al., 2017) is a lightweight CNN-based method to dehaze single-images. It consists of two modules, first, there is a CNN network estimating the $K(x)$ value and then revisit the haze-free image from the estimated $K(x)$. In the second step, the haze-free image is restored using

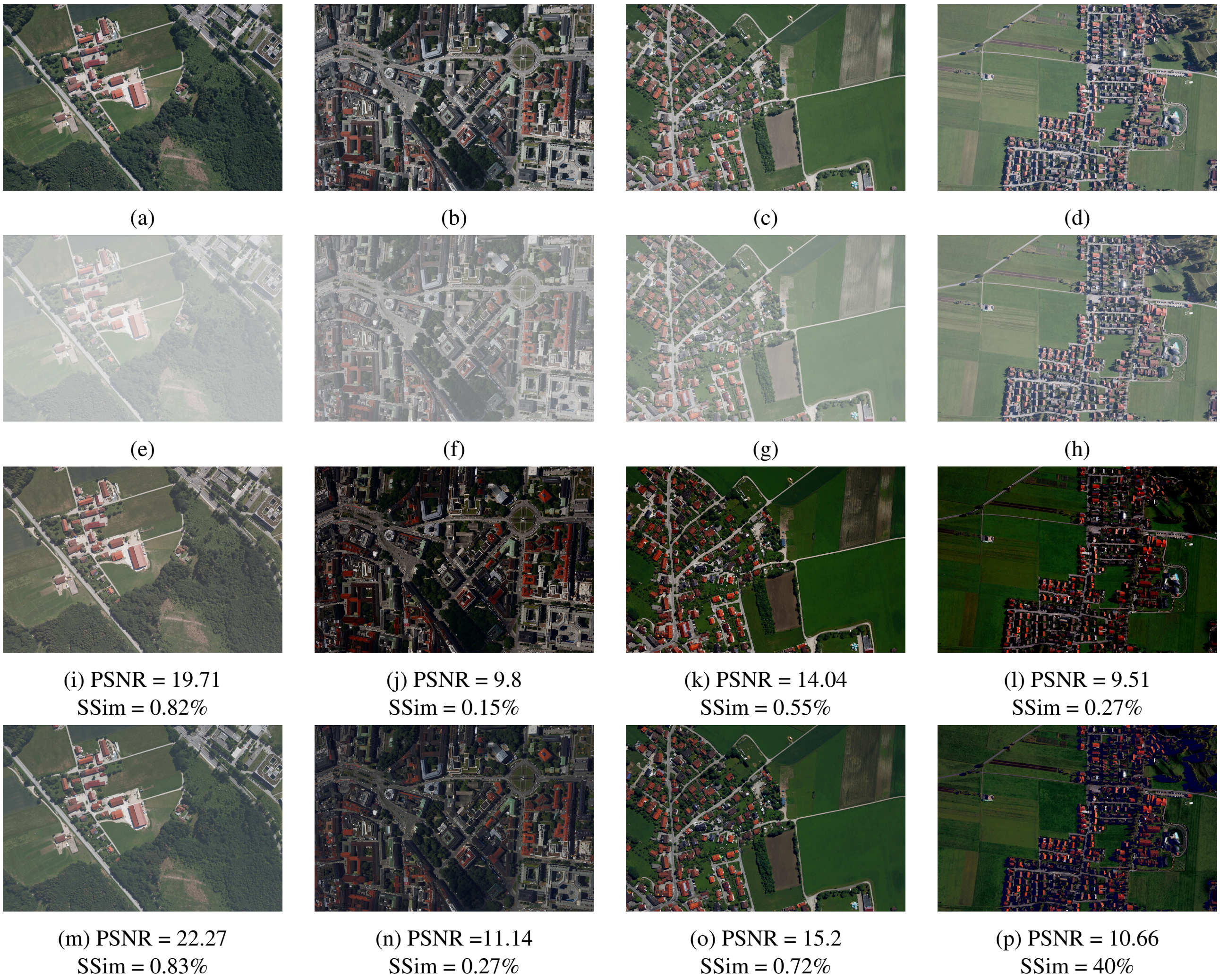
$$J(x) = K(x)I(x) - k(x) + b. \tag{8}$$

The network structure includes two main modules as illustrated in Figure 4. The $K(x)$ estimation module which is shown in Figure 5 unifies two parameters of A and $t_{x,y}$ to be estimated at once. To train the dehazing network, there should be hazy images from the scene as well as the haze-free or ground truth images from the same scene under the same conditions (illuminations, light, camera positions, etc.) available. As discussed in section Section 2.1, the depth of each pixel is a prerequisite to generate synthetic hazy images. There are several image datasets available that provide images from different indoor scenes and the corresponding depth image which is mostly produced using either laser scanners or stereo-matching techniques. The AOD-Net authors have used the well-known NYU depth V2 (Silberman et al., 2012) and Middlebury stereo database (Scharstein, Szeliski., 2003) dataset to train their network. They use using 27,256 and 3,170 images from NYU as training and validation sets respectively with β differing from {0.4,0.6,0.8,1.0,1.2,1.4,1.6} and A in range of [0.6,1.0] during the AOD-Net training. The network converges after 10 epochs. In our experiments, we use 10 epoch also for the sake of comparisons.

4. EXPERIMENT AND RESULTS

In this section, we provide more details on the experiments and the achieved outcomes. The results are compared by their PSNR and SSim with their corresponding ground truth images. To have a fair comparison, the number of images and the batch size of training data remain similar when training the network on aerial hazy image dataset and NYU Depth V2 datasets. The experimental results on the test set show 9% quantitat-

https://doi.org/10.5194/isprs-archives-XLII-4-W18-687-2019



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) PSNR = 19.71 SSim = 0.82%

(j) PSNR = 9.8 SSim = 0.15%

(k) PSNR = 14.04 SSim = 0.55%

(l) PSNR = 9.51 SSim = 0.27%

(m) PSNR = 22.27 SSim = 0.83%

(n) PSNR =11.14 SSim = 0.27%

(o) PSNR = 15.2 SSim = 0.72%

(p) PSNR = 10.66 SSim = 40%

Figure 6: Ground truth images, together with their hazy and dehazed versions. The first row: Haze-free images, the second row: Hazed version of the same images in different levels, the third row: Dehazed images using AOD-Net which is trained on NYU dataset, and the fourth row: Dehazed images using adapted AOD-Net

ive improvement on PSNR and 2.2 on SSim, achieving a more natural-looking dehazed image.

4.1 Experiment Setup

The size of the original aerial images is 5616×3744 px. The empirical results show that when trying to balance the speed of the network in training, the GPU consumption capacity and the information flow into the network, it is optimal to crop the images into 1024×1024 px patches. We crop 140 training and validation images 24 times with 1024×1024 px patch size and haze in 9 levels. In total, we obtain 30,240 training and validation patches. We use Titan X GPUs in our experiments. It takes around 5 days to train the network on 27,240 training images and validate on 3,000 images for 10 epochs with the learning rate of 0.0001. The loss of the network is mean squared error and we keep the Adam optimizer same as the classical AOD-Net structure during the training.

In order to have a fair comparison, we train both AOD-net on NYU depth V2 datasets as well as our aerial hazy image dataset with the same training configurations and parameters. The test set includes 17 images from different flight campaigns, collected from different years and locations in Germany. Some of these images together with their dehazed versions have been illustrated in Figure 6. We use the Peak Signal To Noise Ratio (PSNR) and Structural Similarity (SSim) indicators as widely used metrics for dehazing performance evaluation (Li et al., 2017, Li et al., 2018, Engin et al., 2018, Zhang et al., 2018, Min et al., 2019). For the sample dehazed images, the PSNR and SSim comparisons are mentioned in captions of Figure 6. In Table 1, we calculate the average PSNR and SSim for the entire test set after being dehazed by two AOD-Net dehazing algorithms. Based on the results, we can infer that the dehazing results of the adapted AOD-Net are significantly superior.

4.2 Results And Discussion

In Figure 6, the dehazed images together with their hazy versions are shown. From (e) to (h), we can that the haze level is decreasing. Despite the fact that there are color distortions in both dehazed results, the images which are dehazed by adapted AOD-Net have more a natural-looking appearance and are more clear (see image Figure 6 (i) and (m)). As mentioned in the description of the Figure 6, both PSNR and SSim values of the sample images increase during by adapted AOD-Net. The highest jump in PSNR and SSim belongs to the image (m) and (o), respectively.

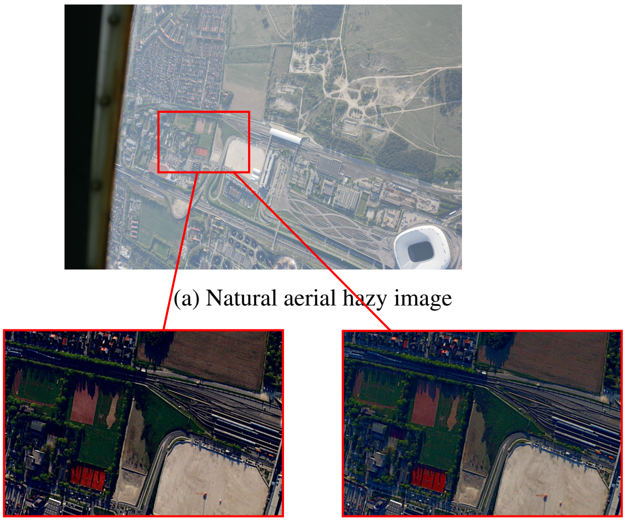
https://doi.org/10.5194/isprs-archives-XLII-4-W18-687-2019

	AOD-Net Classic	AOD-Net Adapted
Average PSNR	14.08	16.28
Average SSim	0.50	0.59

Table 1: The average PSNR and SSim of the test set when tested on Classical and Adapted AOD-Net

To have a general improvement record, we compute the average PSNR and SSim metrics as shown in Table 1. The dehazing performance on the test set of our aerial hazy image dataset has been refined by improving PSNR from 14.08 to 16.28 and SSim from 0.50 to 0.59.

We also test the networks on naturally hazed aerial images, The example for this comparison is shown in Figure 7. Since there is no ground truth available, we have to limit the comparing results to the appearance of the dehazed image. As can be seen, the removal seems to be almost similar in these two methods, but if we zoom into the image, due to the color distortion, the shadow areas using original AOD-Net in left Figure 7(b) becomes black, which is not the case for the adapted AOD-Net in right Figure 7(b).



(a) Natural aerial hazy image

(b) Left: classic AOD-Net dehazing. Right: adapted AOD-Net dehazing

Figure 7: Dehazing natural hazy image

In most of the dehazing results so far, the artifacts like structural damage, color distortion, and unrealistic appearance of the colors due to color range shift or over-enhancement appear (Min et al., 2019). Even though in our case, the unrealistic sharpening problem has been improved a bit, it still present in the image and it may seem pleasing for human eye to look at the images with nicely sharped colors, but when it comes to the similarity comparison of the image to its ground truth, we can consider it as a disadvantage for the dehazing algorithms.

5. CONCLUSION AND FUTURE WORK

One of the ways to enhance aerial image quality is to use dehazing algorithms. In this work, we propose to use a deep learning method, but a dataset with aerial images is needed to develop a devoted algorithm for aerial imagery. Due to the different characteristics of the aerial and terrestrial imageries, we created a new dataset containing aerial hazy images with homogeneously- synthetically hazed images to train the dehazing network. We create our aerial hazy image dataset by obtaining the depth map first and afterward hazy images. We train AOD-Net on our aerial hazy image dataset as aerial imagery domain and compare the results with the original dehazing results. Both qualitative results on the test set as well as the natural hazy images show significant improvements and the quantitative indicators of PSNR and SSim enhance by 2.2% and 9% respectively.

The dehazing results can be further improved using more images and more haze levels, as well as the images with different flight height and inhomogeneous distribution of the haze on it. Adding further layers to the network may also improve the results and several other loss functions such as perceptual loss may overcome the color distortion and artifact problems.

REFERENCES

Azimi, S. M., Fischer, P., Körner, M., Reinartz, P., 2018a. Aerial LaneNet: Lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5), 2920–2938.

Azimi, S. M., Vig, E., Bahmanyar, R., Körner, M., Reinartz, P., 2018b. Towards multi-class object detection in unconstrained remote sensing imagery. *ACCV*.

Engin, D., Genç, A., Ekenel, H. K., 2018. Cycle-dehaze: Enhanced cyclegan for single image dehazing. *CVPRW*.

Kaiming He, Jian Sun, X. T., 2011. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33, 2341-2353.

Kurz, F., Rosenbaum, D., Leitloff, J., Meynberg, O., Reinartz, P., 2011. Real time camera system for disaster and traffic monitoring. *International Conference on Sensors and Models inPhotogrammetry and Remote Sensing*.

Li, B., Peng, X., Wang, Z., Xu, J., Feng, D., 2017. AOD-Net: All-In-One Dehazing Network.

Li, R., Pan, J., Li, Z., Tang, J., 2018. Single Image Dehazing via Conditional Generative Adversarial Network. *CVPR*.

McCartney, E. J., 1976. *Optics of the Atmosphere, Scattering by Molecules and Particles*. wiley.

Min, X., Zhai, G., Gu, K., Zhu, Y., Zhou, J., Guo, G., Yang, X., Guan, X., Zhang, W., 2019. Quality evaluation of image dehazing methods using synthetic hazy images. *IEEE Transactions on Multimedia*.

Scharstein, D., Szeliski., R., 2003. High-accuracy stereo depth maps using structured light. *CVPR*.

Silberman, N., Hoiem, D., Kohli, P., Fergus, R., 2012. Indoor segmentation and support inference from rgbd images. *European Conference on Computer Vision*, Springer, 746–760.

Zhang, H., Sindagi, V., Patel, V. M., 2018. Multi-scale Single Image Dehazing Using Perceptual Pyramid Deep Network. *CVPRW*.