Single Image Dehazing with a Generic Model-Agnostic Convolutional Neural Network

Zheng Liu, Botao Xiao, Muhammad Alrabeiah, Keyan Wang, and Jun Chen

***Abstract*—A simple convolutional neural network is proposed in this letter and is trained end-to-end to restore clear images from hazy inputs. The proposed network is generic and agnostic in the sense that it is not designed specifically for image dehazing and, in particular, it has no knowledge of the atmosphere scattering model. Remarkably, this network achieves record-breaking dehazing performance on several standard data sets that are synthesized using the atmosphere scattering model. This surprising finding suggests that there might be a need to rethink the predominant plug-in approach to image dehazing.**

***Index Terms*—Convolutional neural network, image dehazing, image restoration, residual learning.**

I. Introduction

MANY modern applications rely on analyzing visual data to discover patterns and make decisions. Notable examples include intelligent surveillance, tracking, and control systems, where good quality images or videos are essential for accurate results and reliable performance. However, such systems could be significantly affected by environmentally induced distortions, the most common of which are haze and smog. To mitigate this type of negative environmental impacts on visual analysis, a significant amount of research in the computer vision community (see, e.g., [1]–[4]) has been devoted to image dehazing, which aims to restore clear images from their hazy counterparts (see Fig. 1).

The relation between a hazy image and its clear version is approximately captured by the following equation known as the atmosphere scattering model [5] :

$$I_i(x) = J_i(x)t(x) + A(1 - t(x)) \quad i = 1, 2, 3. \tag{1}$$

Here $I_i(x)$ ($J_i(x)$) is the intensity of the ith color channel of pixel x in the hazy (clear) image; $t(x)$ is the medium transmission function, determined by the scene depth $d(x)$ and the atmospheric scattering coefficient β; parameter A is the atmosphere light intensity, which is assumed to be a global constant over the whole image. Since all variables in (1) are unknown except $I_i(x)$, (single) image dehazing is in general an undetermined problem.

Fig. 1. A hazy image and its clear version.

Over the past few decades, many methods have been proposed for image dehazing. Those methods could be loosely grouped into two categories: traditional and learning-based methods. Roughly speaking, the methods in the first category, such as [1], [3], and [6], solve the aforementioned underdetermined problem by exploiting some form of prior information while those in the second category, such as [2], [7]–[9], take a more data-driven approach by leveraging the advances in classic and deep learning technologies. Despite their many differences, the methods in both categories aim to recover the clear image by first estimating the unknown parameters A and $t(x)$ (or their variants) and then inverting (1) to determine $J_i(x)$.

From the viewpoint of estimation theory, all such methods fall under the umbrella of the plug-in approach[1] and consequently can be referred to as the plug-in methods. However, the optimality of the plug-in approach for image dehazing is not completely justified. Indeed, it is unlikely that the problem of lossy reconstruction of the clear image can be transformed equivalently to an estimation problem for parameters A and $t(x)$ (or their variants), at least when the two problems are subject to the same evaluation metric. Moreover, the actual relation between a hazy image and its clear version can be fairly complex and may not be fully captured by the atmosphere scattering model. Due to this potential mismatch, the plug-in methods do not guarantee desirable generalization to natural images even if they show good performance on synthetic datasets.

Manuscript received February 20, 2019; revised April 3, 2019; accepted April 4, 2019. Date of publication April 11, 2019; date of current version April 22, 2019. The work of K. Wang was supported by the National Natural Science Foundation of China under Grant 61301291. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Demetrio Labate *(Corresponding author: Keyan Wang.)*

Z. Liu, B. Xiao, M. Alrabeiah, and J. Chen are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: liuz156@mcmaster.ca; xiaob6@mcmaster.ca; alrabm@mcmaster.ca; chenjun@mcmaster.ca).

K. Wang is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: kywang@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/LSP.2019.2910403

[1]Consider a parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and a mapping $\tau : \Theta \to \mathbb{R}$. Suppose the observation comes from P_{θ^*}. The plug-in approach refers to constructing an estimate of $\tau(\theta^*)$ by first deriving an estimate of θ^*, denoted by $\hat{\theta}$, then plugging $\hat{\theta}$ into $\tau(\cdot)$.

1070-9908 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

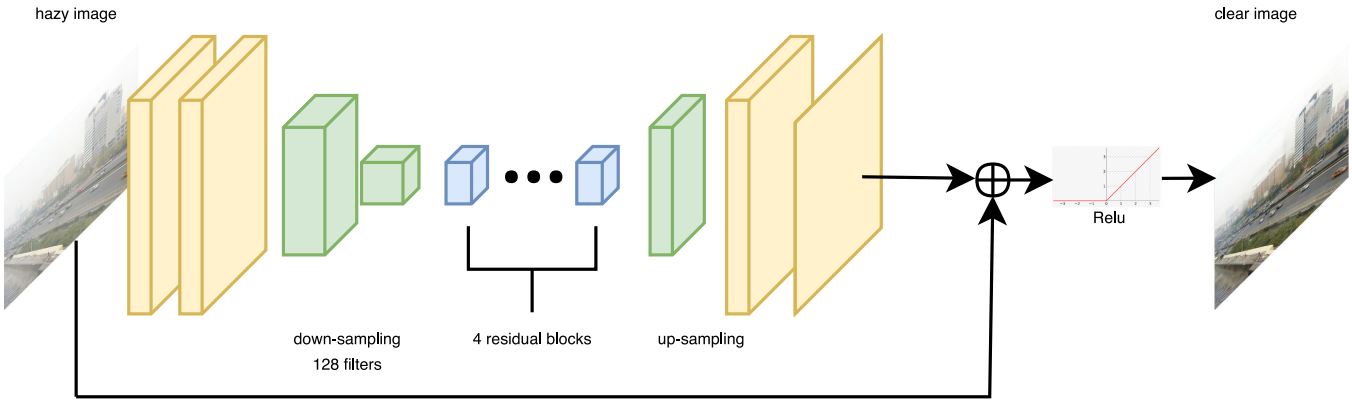


Fig. 2. The proposed generic model-agnostic convolutional neural network. The yellow blocks are convolutional layers and the green blocks are down-sampling/deconvolutional layers. The 4 concatenated residual blocks, shown in blue color, contain 2, 2, 3, 4 convolutional layers, respectively.

In this letter, we pursue a different approach to image dehazing, where no knowledge of the atmosphere scattering model is required and the parameter estimation step is completely circumvented. To the best of our knowledge, this model-agnostic approach was first put forward in [10] as a potential alternative to the plug-in approach; however, no attempt was made there to investigate its competitiveness. This new approach was also adopted in the recent work [11], which introduced a Gated Fusion Network (GFN) and demonstrated its promising performance for image dehazing. We move one step further along this line of research by proposing a new convolutional neural network that outperforms the state-of-the-art on several standard datasets. It will be seen that our proposed network has several advantages over GFN, especially in terms of architecture complexity and input-size flexibility; moreover, certain characteristics of GFN are specifically tailored to image dehazing whereas the architecture of our network is more generic and consequently more broadly applicable.

The rest of the letter is organized as follows. In Section II, we introduce a new fully convolutional neural network and give a detailed explanation of its overall architecture, important features, and main building blocks. The performance evaluation of the proposed network for image dehazing is conducted in Section III, which also includes a description of the datasets and the training procedure. Section IV contains some concluding remarks.

II. The Proposed Network

We shall introduce a Generic Model-Agnostic convolutional neural Network (GMAN) in this section. As its name suggests, this network is completely generic, with no special consideration of image dehazing in its design; in particular, it makes no attempt to exploit the atmosphere scattering model. The following subsections are devoted to a detailed explanation of GMAN in terms of its overall architecture, important features, and main building blocks as well as the loss function used for its training.

A. Network Architecture

As shown in Fig. 2, the proposed GMAN takes a hazy image as the input and outputs its clear version. Functionally speaking,

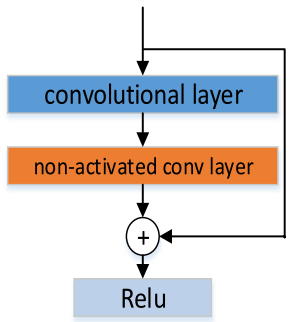


Fig. 3. A residual block used in the middle layer of the proposed GMAN. The number of convolutional layers can be different from block to block. ReLU is used as the activation function after the addition operator in each block.

GMAN is an end-to-end generative network that employs an encoder-decoder structure with down- and up-sampling factor of 2. Its first two layers are constructed with 64-channel convolutional blocks. Following them are two-step down-sampling layers that encode the input image into a $56 \times 56 \times 128$ volume. The encoded image is then fed to a residual layer built with 4 residual blocks, each containing a shortcut connection as shown in Fig. 3. This layer represents the transition from encoding to decoding, for it is followed by the deconvolutional layer that up-samples the residual layer output and reconstructs a new $224 \times 224 \times 64$ volume, which is finally added to the input image and thresholded with a ReLU to produce the haze-free image.

B. Residual Learning

The network uses residual learning on two levels, local and global. In the middle layer and just right after down-sampling, the residual blocks are used to build the local residual layer. It takes advantage of the hypothesized and empirically proven [12]–[15] easy-to-train property of residual blocks (see [16]). Residual learning also appears in the overall architecture of the proposed GMAN. Specifically, the input image is fed along with the output of the final convolutional layer to a sum operator, creating one global residual block as shown in Fig. 2. The purpose

of introducing this global residual block is to compensate the potential information loss in the intermediate processing steps.

C. Encoder-Decoder Structure

The proposed GMAN employs the popular encoder-decoder structure, which is composed of three parts: encoder, residual layer, and decoder. The input is transformed and down-sampled by the encoder and then fed to the residual layer to extract certain features based on which the decoder generates an approximation of the desired output via up-sampling and transformation. The down-sampling and up-sampling process helps to retain the critical features and discard the inessential ones so that the network has good generalization performance.

D. Loss Function

The MSE loss is used to train the proposed GMAN. Specifically, we define

$$L = \frac{1}{N}\sum_{x=1}^{N}\sum_{i=1}^{3} \parallel \hat{J}_i(x) - J_i(x) \parallel^2, \tag{2}$$

where $\hat{J}_i(x)$ ($J_i(x)$) is the intensity of the ith color channel of pixel x in the output image (the ground truth), and N is the total number of pixels.

III. Experimental Results

In this section, we first provide a description of the datasets and the training procedure, then present the evaluation results of the proposed GMAN, benchmarked against several existing methods:[2] DCP [3], DehazeNet [2], MSCNN [8], AOD-Net [9], and GFN [11].

A. Datasets

According to the atmosphere scattering model, the haze level is controlled by the transmission map $t(x)$ and the atmosphere light intensity A. Setting these two factors properly is important for building a training dataset of synthetic hazy images. In this work we adopt the Ondoor Training Set (OTS) and the Indoor Training Set (ITS) from RESIDE [17] for network training. The OTS contains 313,950 synthetic hazy images, generated from 8,970 outdoor clear images through the atmosphere scattering model by varying the values of A and β (the depth information $d(x)$ is estimated using the algorithm in [18]). We notice that the 1000 ground-truth images and their 35,000 synthetic hazy counterparts in the Synthetic Objective Testing Set (SOTS), which is the testing set from RESIDE, are all contained in the OTS. These testing images are removed from the training dataset, leading to a reduced-size dataset of 278,950 hazy images (generated from 7,970 ground-truth images). The ITS contains 13,390 synthetic hazy images, generated from 1,339 indoor clear images in a way similar to that for the OTS (the depth information $d(x)$ is obtained from NYU Depth dataset V2 [19] and Middlebury Stereo datasets [20]).

TABLE I
Performance Comparison on the Outdoor SOTS

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
PSNR	18.54	26.84	21.73	24.08	21.67	28.19
SSIM	0.710	0.826	0.831	0.873	0.852	0.964
SSEQ	27.41	69.84	71.21	85.38	68.31	77.26
B-II	89.12	83.60	86.72	87.43	79.30	89.31

B. Training

The proposed GMAN is trained end-to-end by minimizing the loss L given by (2). All layers in GMAN have 64 filters (kernels), except for the down-sampling ones which have 128 filters, with spatial size of 3×3. The network requires an input with size 224×224, so every image in the training dataset is randomly cropped in order to fit the input size.[3] The batch size is set to 35 to balance the training speed and the memory consumption on the GPU. For accelerated training, the Adam optimizer [21] is used with the following settings: the initial learning rate of 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The training process is implemented using TensorFlow and is carried out on four NVIDIA GTX 1080Ti GPUs. After 20 epochs of training, the loss function drops to 0.0004, which is considered a good stopping point.

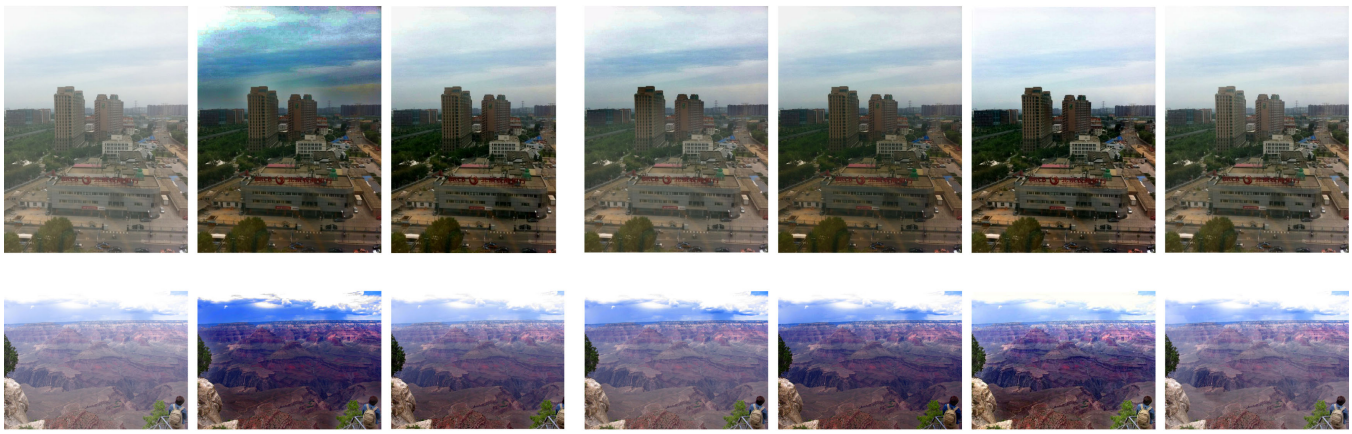
C. Evaluation Results

Although the proposed GMAN is not designed specifically for image dehazing and does not exploit the atmosphere scattering model, its dehazing capability is surprisingly strong; in particular, it achieves superior performance relative to many state-of-the-art plug-in methods even on the datasets that are synthesized using the atmosphere scattering model. According to Table I, GMAN clearly outperforms all other competing methods under consideration in terms of PSNR and SSIM values on the outdoor SOTS. Moreover, as shown in Fig. 4, GMAN avoids darkening the image color as well as the excessive sharpening of object edges. In contrast, it can be seen from Fig. 4 that DCP dims the light intensity of the dehazed image and causes color distortions in high-depth-value regions (e.g., sky); though MSCNN does well in these high-depth-value regions, its performance degrades in medium-depth areas of the target image. Hence, the proposed GMAN can overcome many of these issues and generate better haze-free images. We have also tested GMAN on the indoor SOTS (see Table II). In this case, its performance is again very impressive, outperforming other competing methods in terms of PSNR and SSIM values by a big margin. These results suggest that the model-agnostic approach is not only feasible, but also capable of delivering record-breaking dehazing results for both outdoor and indoor images.

We further compare the dehazing results using two no-reference Image Quality Assessment (IQA) metrics: Spatial-Spectral Entropy-based Quality (SSEQ) and BLind Image Integrity Notator using DCT Statistics (BLIINDS-II or simply B-II). It can be seen from Table I and Table II that the evaluation results based on no-reference IQA metrics are not fully

[2]The evaluation results of these methods are either quoted from [17] or obtained using the publicly available source codes.

[3]This restriction is only for the training phase. The trained network can be applied to images of arbitrary size



(a) Hazy (b) DCP (c) Dehazenet (d) MSCNN (e) AOD-Net (f) GFN (g) GMAN

Fig. 4. Comparison of different dehazing methods for a synthetic image (first row) and a natural image (second row).

TABLE II
PERFORMANCE COMPARISON ON THE INDOOR SOTS

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
PSNR	18.87	22.66	20.01	21.01	22.44	27.94
SSIM	0.794	0.833	0.791	0.837	0.884	0.897
SSEQ	54.92	66.73	66.78	69.90	69.18	66.40
B-II	74.41	71.71	74.49	78.94	75.19	75.12

TABLE III
PERFORMANCE COMPARISON ON HEAVY HAZE IMAGES FROM THE OUTDOOR SOTS

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
PSNR	15.43	23.09	18.65	20.02	20.24	26.00
SSIM	0.771	0.894	0.818	0.891	0.769	0.936

TABLE IV
PERFORMANCE COMPARISON ON THE CITYSCAPES OUTDOOR DATASET

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
PSNR	17.26	14.94	18.49	16.11	15.69	18.64
SSIM	0.747	0.482	0.879	0.728	0.625	0.813

consistent with those based on the full-reference PSNR/SSIM metrics. A detailed discussion regarding this inconsistency can be found in [17].

We also conduct a performance comparison on heavy haze images from the outdoor SOTS. Specifically, we select those images with the maximum β value, i.e., $\beta = 0.2$ (there are totally 106 such images). As shown by Table III, the proposed GMAN is highly effective in removing heavy haze, which is known to be a difficult task.

To demonstrate its robustness, the proposed GMAN (pre-trained on the OTS) is further evaluated on the Cityscapes outdoor dataset [22] without re-training or fine-tuning. We have also tested our network (pre-trained on the ITS) directly on a new indoor synthetic dataset.[4] It can be seen from Table IV and Table V that the proposed GMAN still shows highly competitive performance.

[4]To create this dataset, we randomly select 100 images (accompanied by their depth information) from the the Sun RGB-D dataset [23] and use them to generate hazy images through the atmosphere scattering model with some arbitrary $\beta \in [0.2, 0.4]$ (which yields very heavy haze) and $A \in [0.8, 1]$.

TABLE V
PERFORMANCE COMPARISON ON THE SUN RGB-D INDOOR DATASET

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
PSNR	15.24	21.18	19.63	20.63	20.32	22.98
SSIM	0.862	0.932	0.905	0.927	0.784	0.913

TABLE VI
DETECTION RESULTS ON THE RTTS (IN %)

	DCP	DehazeNet	MSCNN	AOD-Net	GFN	GMAN
Person	60.67	60.94	60.97	60.82	60.11	61.03
Bicycle	40.76	40.58	41.40	40.22	37.76	41.14
Car	34.94	41.48	42.31	34.93	34.86	41.74
Bus	24.06	25.11	25.08	20.47	24.98	20.86
Motorbike	30.21	33.68	35.56	30.20	27.93	35.46

Finally, we compare the detection results of Faster R-CNN on the Real-world Task-driven Testing Set (RTTS) from RESIDE [17]. As shown by Table VI, the proposed GMAN demonstrates competitive performance across all categories except for "Bus".

IV. CONCLUSION

We have proposed a simple convolutional neural network and leveraged it for image dehazing. In constrast to the existing plug-in methods, the proposed network requires no knowledge of the atmosphere scattering model and avoids the parameter estimation step completely. Experimental results have verified its potential in generating haze-free images and its capability to overcome some of the common pitfalls of state-of-the-art methods, like color darkening and excessive edge sharpening. We have made no attempt to optimize the design of this network, and it is conceivable that various refinement can be made to further improve the dehazing performance.

More importantly, due to its generic architecture, the proposed network could lay the groundwork for future research on general-purposed image restoration techniques. Indeed, we expect that through suitable data augmentation and pre-processing, our network can be trained to handle a wide class of image distortions and corruptions. In this sense, the present work not only suggests a viable approach to image dehazing, but also represents a progressive move towards developing a universal image restoration method.

REFERENCES

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

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

[3] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[4] D. Berman *et al.*, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 1674–1682.

[5] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.*, vol. 48, no. 3, pp. 233–254, 2002.

[6] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[7] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 2995–3000.

[8] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.

[9] B. Li *et al.*, "AOD-Net: all-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 4780–4788.

[10] L. T. Goncalves, J. D. O. Gaya, P. Drews, and S. S. D. C. Botelho, "Deepdive: An end-to-end dehazing method using deep dearning," in *Proc. IEEE 30th SIBGRAPI Conf. Graph., Patterns Images*, 2017, pp. 436–441.

[11] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1–9.

[12] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 1646–1654.

[13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[14] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. Nat. Conf. Artif. Intell.*, 2017, vol. 4, pp. 4278–4284.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.

[17] B. Li *et al.*, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[18] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2024–2039, Oct. 2016.

[19] P. K. Nathan Silberman, D. Hoiem, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 746–760.

[20] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 195–202. [Online]. Available: http://dl.acm.org/citation.cfm?id=1965841.1965865

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[22] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.

[23] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 567–576.