

Unsupervised Single Image Dehazing Using Dark Channel Prior Loss

Alona Golts¹, Daniel Freedman, and Michael Elad², *Fellow, IEEE*

Abstract—Single image dehazing is a critical stage in many modern-day autonomous vision applications. Early prior-based methods often involved a time-consuming minimization of a hand-crafted energy function. Recent learning-based approaches utilize the representational power of deep neural networks (DNNs) to learn the underlying transformation between hazy and clear images. Due to inherent limitations in collecting matching clear and hazy images, these methods resort to training on synthetic data, constructed from indoor images and corresponding depth information. This may result in a possible domain shift when treating outdoor scenes. We propose a completely unsupervised method of training via minimization of the well-known, Dark Channel Prior (DCP) energy function. Instead of feeding the network with synthetic data, we solely use real-world outdoor images and tune the network’s parameters by directly minimizing the DCP. Although our “Deep DCP” technique can be regarded as a fast approximator of DCP, it actually improves its results significantly. This suggests an additional regularization obtained via the network and learning process. Experiments show that our method performs on par with large-scale supervised methods.

Index Terms—Energy functions, deep neural networks, unsupervised learning, single image dehazing, dark channel prior.

I. INTRODUCTION

HAZE is an atmospheric phenomenon where small particles, called aerosols, obstruct the clarity of an outdoor scene and lead to poor contrast and loss of detail. The existence of haze affects an image in two aspects. It attenuates the scene radiance with correspondence to an object’s distance from the camera. Moreover, it introduces an additional ambient light component, called the *airlight*, which causes a “veiling effect” over the clear image. The formation of a hazy image is often described as a linear per-pixel combination of the clear scene radiance and the airlight; the effect of each component is controlled by the transmission map. To recover the scene radiance image, one has to solve a system of $3N$ linear equations with $4N + 3$ unknowns (where N is the number of image pixels).

In order to handle the under-constrained haze creation model, many researchers suggested hand-crafted image priors,

Manuscript received March 7, 2019; revised September 18, 2019; accepted November 1, 2019. Date of publication November 12, 2019; date of current version January 23, 2020. This work was supported by the Israel Science Foundation (ISF) under Grant 335/18. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chandra Sekhar Seelamantula. (*Corresponding author: Alona Golts.*)

A. Golts and M. Elad are with the Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: salonaz@cs.technion.ac.il; elad@cs.technion.ac.il).

D. Freedman is with Google Research, Haifa, Israel (e-mail: danielfreedman@google.com).

Digital Object Identifier 10.1109/TIP.2019.2952032

shedding additional light on the behaviour of hazy versus clean images [1]–[9]. These prior-based methods often formulate the problem of dehazing as an energy minimization task, where obtaining the solution of each image is called “inference”, requiring a non-trivial optimization scheme. With the increasing importance of image dehazing as an initial pre-processing stage in many computer-vision tasks (e.g., object detection, autonomous car navigation), large-scale learning-based techniques have been deployed to solve it [10]–[14]. These methods, however, require thousands of input and output examples.

Since clean and hazy images of the exact same scene and lighting conditions are hard to obtain, learning-based methods commonly resort to synthetic dataset creation. Given a clean image and a corresponding depth map, one can calculate the transmission map and use the haze creation model to obtain hazy images with varying amounts of haze and airlight components. These pairs of hazy and clear images are later fed as inputs and labels in a supervised training of a DNN. Outdoor depth information, however, is incredibly imprecise. For instance, the depth information of the outdoor Make3D [15] and KITTI [16] datasets suffers from over 4 meters of average root-Mean-Square-Error (rMSE), while the rMSE of the indoor NYU2 [17] is only 0.5. Consequently, large-scale methods either use the more reliable indoor depth information [10], [12]–[14], or draw the depth map at random [6], [11]. Either of these practices creates a domain shift when addressing real-world outdoor images.

We propose to leverage the representational power of DNNs, but instead of feeding them with inaccurate synthetic pairs of hazy and clean images, we train them in an unsupervised fashion using real-world hazy images only. We optimize the network weights by minimizing an unsupervised loss function, essentially the Dark Channel Prior (DCP) [3] energy function. Our network can be regarded as a fast feed-forward approximator of the DCP. However, by stopping the optimization early, we get a significant boost in results over the classic DCP. This implies an added regularization, stemming from the network architecture and learning process. Our network, based on the Context Aggregation Network (CAN) architecture [18], is trained end-to-end from scratch without relying on any external data apart from raw hazy images. It provides the predicted transmission maps as output, from which the dehazed image can be easily reconstructed. We perform a comprehensive quantitative evaluation¹ of our method and present state-of-the-art results on *SOTS-outdoor* in the recently

¹Our code is available at: https://github.com/AlonaGolts/Deep_Energy

released RESIDE dataset [19]. We show qualitative results on real-world images, demonstrating that the additional regularization provided by the network reduces common artifacts of prior-based methods, such as over-saturation and high-contrast.

Our “Deep-DCP” method offers the following contributions:

- 1) It provides state-of-the-art results in outdoor single image dehazing, outperforming both prior-based and fully-supervised DNN methods.
- 2) It achieves an impressive ~ 6.5 dB boost in outdoor PSNR over classical DCP, validating an effective regularization.
- 3) It treats the sky successfully where DCP typically fails.
- 4) It is among the first² to perform unsupervised training in single image dehazing, reducing the need in synthetic data.
- 5) It does not require an explicit optimization for each image as DCP, but rather learns the underlying transformation during training, requiring a fast forward-pass during test.
- 6) It offers a generic methodology of unsupervised training with energy functions and can be applied to any successful energy function.

The remainder of this paper is structured as follows: Section II provides a survey of previous prior-based and data-driven approaches for dehazing; Section III describes the DCP, its use as a loss function and our CAN-based architecture; Section IV provides quantitative, qualitative and runtime experimental results; Section V includes discussions and further analysis; finally, Section VI concludes this work.

II. RELATED WORK

A. Prior-Based Approaches

Early attempts at image dehazing have incorporated several images of the same scene, taken at different bad weather conditions [21], or using different polarization filters [22]. Kopf *et al.* [23] later performed dehazing of outdoor images by utilizing existing geo-referenced terrain and urban models including depth, texture and GIS data.

In [2], Tan *et al.* unveiled the haze from a single image by maximizing the local contrast of each patch in the image using a Markov Random Field (MRF) framework. In [1], Fattal *et al.* suggested utilizing the lack of correlation between the transmission and shading in a localized set of pixels, as a prior to resolve the ambiguity between the scene albedo and the airlight. Tarel and Hautiere [4] provided a fast calculation of the “atmospheric veil” using a series of edge-preserving linear filter operations. In [24], Nishino *et al.* exploited the statistical independence between the scene albedo and depth and factorized both quantities into an MRF-based energy function.

In [3], He *et al.* proposed the now widely used DCP and demonstrated that in clear images the darkest pixel in an image patch is close to zero (this, however, does not hold in sky-regions). Using this and the assumption that the transmission map within a small image patch is constant, a coarse map can

be easily derived. They further suggested a computationally costly soft matting operation for smoothing out the transmission and reconstructing the final dehazed image. Follow-up works have improved both the quality and efficiency of DCP. Specifically, in [8], the authors proposed a general boundary constraint for the transmission map for which the DCP is a special case.

Several color-based priors have been suggested as well for boosting dehazing performance [5], [7], [25]. In [5], Fattal used the “color-lines” assumption, stating that pixels in small image patches have a 1D distribution in RGB-space [26]. The offset of these straight lines from the origin in hazy images allow to estimate the transmission map. Berman *et al.* proposed a global approach, called non-local dehazing (NLD) [7]. They observed that a haze-free image contains only several hundreds of distinct colors, clustered as points in RGB-space. In the presence of haze, these color clusters form a “haze-line” where the position of a certain pixel along the line corresponds to its initial radiance color and distance from the camera.

While prior-based methods reveal fine image details, they often suffer from increased saturation and contrast, unrealistic colors and difficulty in handling sky regions. This is due in part to assumptions not suited for all hazy image patches. In addition, each image requires a separate non-trivial optimization and solution which can be prohibitive for real-time applications.

B. Data-Driven Approaches

In [6], a Color Attenuation Prior (CAP) is suggested, mixing hand-crafted observations with a data-driven approach. CAP assumes that the image depth, the amount of haze and the difference between the brightness and saturation are linearly correlated. To find the exact correlation, the authors opt for supervised regression between synthesized hazy patches and their corresponding depth maps. This results in fast inference at test time.

One of the first works to propose single image dehazing using CNNs is [12]. The method, called MSCNN, is trained by feeding a two-stage network with pairs of hazy images and corresponding transmission maps. In DehazeNet [11], Cai *et al.* create a novel CNN architecture (featuring maxout and BReLU layers), inspired by popular prior-based methods [2], [3], [6], [9]. AOD-Net [10] in turn, proposes a joint estimation of both the transmission map and the airlight via a unified representation. Using this representation, one can easily reconstruct the scene radiance directly in an end-to-end forward-pass computation. This helps reduce errors accumulated in the separate calculation of the two quantities.

In the recent Gated Fusion Network (GFN) [13], a dehazed image is produced as a fusion of the white balance, contrast enhanced and gamma corrected images (all derived from the hazy image). The network outputs three confidence maps which determine the effect of each component. To combat halo effects of a single scale encoder-decoder structure, a multi-scale architecture is used where a coarse output is first produced, then added as input to a finer scale network.

²See later discussion on a GAN based alternative [20].

This method provides impressive results on RESIDE’s *SOTS-indoor*, but quadruples the size of the input during training and test, making evaluation inefficient in terms of memory. Finally, in a recent work reported in [14], the authors utilize the pre-trained VGG [27] network as encoder and only train the symmetric decoder via a combination of MSE and perceptual loss.

While learning-based methods achieve impressive results, they are trained in a supervised way, relying on synthetic datasets. Some methods use more accurate *indoor* depth information to create labelled inputs [10], [12]–[14]. This practice, however, directs increasing research effort to optimizing indoor performance, while the predominant need for dehazing is actually outdoors. Other methods use real-world *outdoor* images, but compromise the accuracy of the depth information. For example, [6] draws each pixel in the depth map at random from a (0, 1) uniform distribution and [11] enforces an additional constraint of constant depth within 16×16 patches. These assumptions result in block and halo artifacts in the reconstructed image and require additional post-processing.

In contrast to previous data-driven methods, the work reported in [20] proposes an unsupervised training scheme. The conditional GAN architecture in [20] consists of three generator networks, receiving the hazy image as input and producing the scene radiance, transmission map and airlight channels, and one discriminator network, deciding whether the reconstructed dehazed image is real or fake. While [20] has the same motivation as ours, there are several key differences between the two works: (1) As opposed to GANs, which are known as less stable and harder to train, our network is a relatively simple and easy to train feed-forward model; (2) In practice, [20] does use a synthetic indoor training set to train their GAN. Specifically, during training they use both hazy images, as well as the ground truth dehazed versions; however, they do not require the hazy image and its clear version to be paired/corresponding. Nevertheless, this setup entails the need for a collection of pairs of images for which one has both the hazy and dehazed versions. Our method, on the other hand, uses only real-world hazy images to tune the weights of the network. By doing so, we circumvent the need for a large synthetic dataset and include more complex images that do not necessarily abide to the simplified haze formation model.

III. OUR METHOD

In the following we will describe our method for single image dehazing, including the driving force of our unsupervised loss function, the Dark Channel Prior [3], its implementation as a loss function for training a CNN and the architecture we choose for the task at hand.

A. Haze Model

The popular haze formation model in [28] is given as:

$$\begin{aligned} \mathbf{I}(\mathbf{x}) &= t(\mathbf{x})\mathbf{J}(\mathbf{x}) + (1 - t(\mathbf{x}))\mathbf{A}, \\ t(\mathbf{x}) &= e^{-\beta d(\mathbf{x})}. \end{aligned} \quad (1)$$

According to the above, the observed hazy image, $\mathbf{I}(\mathbf{x}) \in \mathbb{R}^{N \times 3}$, is a convex linear combination of the haze-free scene radiance, $\mathbf{J}(\mathbf{x})$, and the atmospheric light component, \mathbf{A} , called the *airlight*; usually represented as a constant 3-vector in RGB-space, $\mathbf{A} = (A^r, A^g, A^b)$. The transmission map coefficients, $t(\mathbf{x}) \in \mathbb{R}^N$ control the relative force of each component, in each pixel in the image, $\mathbf{x} \in \mathbb{R}^N$. The transmission is a function of the depth, $d(\mathbf{x})$, of the scene from the observer. Our goal in single-image-dehazing is to obtain the haze-free scene radiance, $\mathbf{J}(\mathbf{x})$. To do so, however, one needs to solve a set of $3N$ equations (only $\mathbf{I}(\mathbf{x})$ is given), with $4N + 3$ unknowns ($\mathbf{J}(\mathbf{x}), t(\mathbf{x}), \mathbf{A}$). Thus, additional prior knowledge of the images in question is needed.

B. Dark Channel Prior

The dark channel prior is an image statistical property, indicating that in small patches of haze-free outdoor images, the darkest pixel across all color channels is very dark, and close to zero. The “dark channel” of the image is defined as

$$\mathbf{J}^{dark}(\mathbf{x}) = \min_{c \in \{r, g, b\}} \left(\min_{\mathbf{y} \in \Omega(\mathbf{x})} (\mathbf{J}^c(\mathbf{y})) \right), \quad (2)$$

where $\Omega(\mathbf{x})$ is a small patch, centered around \mathbf{x} . This observation is contributed by three factors which appear in outdoor images: (1) shadows – induced by cars, buildings and trees; (2) colorful objects – where one color channel is dominant, and the others are close to zero, e.g., red flowers, green leaves, blue sea; and (3) naturally dark objects – such as tree trunks and rocks.

Assuming that \mathbf{A} is known and the transmission within a small image patch, denoted as $\tilde{t}(\mathbf{x})$, is constant, one can apply a minimum operation across channels and pixels in the haze formation equation in (1) (effectively zeroing $\mathbf{J}^c(\mathbf{y})$) and get a prediction for the transmittance [3]:

$$\tilde{t}(\mathbf{x}) = 1 - \omega \cdot \min_c \left(\min_{\mathbf{y} \in \Omega(\mathbf{x})} \left(\frac{\mathbf{I}^c(\mathbf{y})}{A^c} \right) \right), \quad (3)$$

where $\omega = 0.95$ leaves a small amount of haze for natural-looking results. In sky regions, although the dark channel does not always hold, it is assumed that $\mathbf{I}/\mathbf{A} \rightarrow 1$, thus $\tilde{t}(\mathbf{x}) \rightarrow 0$. The resulting coarse transmission map requires an additional step of refinement.

C. Soft Matting

The haze formation model in (1) is very similar to the composition model in image matting [29], where an output image is a convex linear combination of foreground and background images; controlled by the alpha matte, α . If one replaces the α -matte with the coarse transmission map, $\tilde{t}(\mathbf{x})$, the following energy function suggested in [29] can be used to acquire the refined map, $t(\mathbf{x})$:

$$E(\mathbf{t}, \tilde{\mathbf{t}}) = \mathbf{t}^T \mathbf{L} \mathbf{t} + \lambda (\mathbf{t} - \tilde{\mathbf{t}})^T (\mathbf{t} - \tilde{\mathbf{t}}), \quad (4)$$

where the first term promotes successful image matting, and the second, fidelity to the dark channel solution. The parameter λ , controlling the force between the two, is set

to $\lambda = 10^{-4}$ [3]. The matrix \mathbf{L} is a Laplacian-like matrix, dedicated to image matting and given by [29]:

$$\begin{aligned} \mathbf{L}_{ij} &= \sum_{n|(i,j) \in p_n} (\delta_{ij} - w_{ij}^n), \quad \forall i, j = 1 \dots N \\ w_{ij}^n &= \frac{1}{|p_n|} \left(1 + (\mathbf{I}_i - \boldsymbol{\mu}_n)^T (\boldsymbol{\Sigma}_n + \frac{\varepsilon}{|p_n|} \mathbf{U}_3)^{-1} (\mathbf{I}_j - \boldsymbol{\mu}_n) \right), \end{aligned} \quad (5)$$

where i, j are two pixels within a small patch p_n around pixel n ; $|p_n|$ is the size of the patch and equal to $3 \times 3 = 9$ as suggested in [29]; δ_{ij} is the Kronecker delta; $\boldsymbol{\mu}_n \in \mathbb{R}^3$ and $\boldsymbol{\Sigma}_n \in \mathbb{R}^{3 \times 3}$ are the mean and covariance of the patch; \mathbf{U}_3 is the 3×3 identity matrix; and ε is a smoothing parameter set to $\varepsilon = 10^{-6}$ [29]. In cases where i, j are not within the same 3×3 patch, $\mathbf{L}_{ij} = 0$, making the matrix \mathbf{L} sparse.

D. Implementation as a Loss Function

We rewrite the energy function in equation (4) in a tensor-friendly format by using a known decomposition of Laplacian matrices via their weights, given in (5). Rephrasing the first term in (4) in terms of weights, we have that

$$E_1(\mathbf{t}, \tilde{\mathbf{t}}) = \mathbf{t}^T \mathbf{L} \mathbf{t} = \sum_{n \in N} \sum_{i, j \in p_n} w_{ij}^n (t_i - t_j)^2, \quad (6)$$

where we sum over all overlapping 3×3 patches $n \in N$ in the transmission map \mathbf{t} , adding the contributions of all pixels i, j that reside within the same patch p_n . Since i, j can both be in the range [1..9] ('1' being the first pixel in the patch, and '9' being the last), the maximum number of possible $i, j \in p_n$ combinations is $K = (3^2) \cdot (3^2) = 81$. We can rewrite this sum in vector form as follows:

$$E_1 = \sum_{n \in N} \sum_{i, j \in p_n} w_{ij}^n (t_i - t_j)^2 = \sum_{n=1}^N \sum_{k=1}^K \mathbf{W} \odot (\mathbf{T}^I - \mathbf{T}^J)^2, \quad (7)$$

where \odot is an element-wise multiplication, and $k \in [1..K = 81]$ indexes all possible combinations of pixels i, j within a 3×3 patch. The matrix $\mathbf{W} \in \mathbb{R}^{N \times K}$ is a vectorized version of the scalar weights w_{ij}^n , and $\mathbf{T}^I, \mathbf{T}^J \in \mathbb{R}^{N \times K}$ are repetitions of the refined transmission map \mathbf{t} . Denote $(1, 2, \dots, 9)$ as representing each pixel coordinate in a 3×3 patch. Each row $n \in N$ in the matrix \mathbf{T}^I holds the following repetitions of pixel values within the relevant patch p_n : $(1, \dots, 1, 2, \dots, 2, \dots, 9, \dots, 9)_n \in \mathbb{R}^{81}$. Similarly, row n in the matrix \mathbf{T}^J holds the repetitions: $(1, 2, \dots, 9, 1, 2, \dots, 9, \dots, 1, 2, \dots, 9)_n \in \mathbb{R}^{81}$. The vectorization of the second term in Equation (4) is much simpler, giving rise to the following:

$$E(\mathbf{t}_\theta, \mathbf{I}) = \sum_{n=1}^N \sum_{k=1}^K \mathbf{W}(\mathbf{I}) \odot (\mathbf{T}_\theta^I - \mathbf{T}_\theta^J)^2 + \lambda \sum_{n=1}^N (\mathbf{t}_\theta - \tilde{\mathbf{t}}(\mathbf{I}))^2, \quad (8)$$

where we have added a subscript θ to denote the matrices $\mathbf{t}, \mathbf{T}^I, \mathbf{T}^J$ are a result of our learned network. On the other hand, both the weight matrix, \mathbf{W} , and the coarse transmission map, $\tilde{\mathbf{t}}$, are computed using the input hazy image, \mathbf{I} , thus we have adjusted the input of the energy function accordingly.

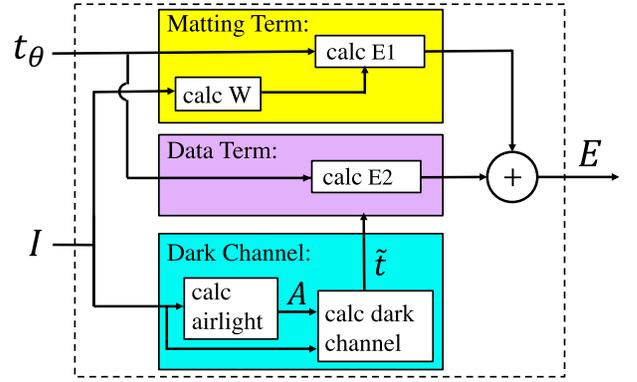


Fig. 1. Our loss module, which receives the prediction of the network, \mathbf{t}_θ , along with the hazy image, \mathbf{I} , and outputs the value of the DCP [3] energy loss.

Above is the loss function with which we train our network, whose predicted transmission map is parametrized by \mathbf{t}_θ . We tune the parameters, θ , by minimizing the loss function in (8) over the training set of hazy images, $\{\mathbf{I}_m\}_{m=1}^M$:

$$\theta^* = \arg \min_{\theta} \left[\frac{1}{M} \sum_{m=1}^M E(\mathbf{t}_\theta(\mathbf{I}_m), \mathbf{I}_m) \right], \quad (9)$$

where M is the number of images. Note that we do not use the “labels”, i.e., the clear images, at any point, only the original hazy ones. A schematic diagram of the inputs and outputs of our loss module is given in Figure. 1.

E. Computing the Scene Radiance

Once the network has finished training, the transmission map, $t_\theta(\mathbf{x})$, of a new hazy image can be obtained by a forward-pass operation. This is used to recover the scene radiance via the haze formation model in (1):

$$\mathbf{J}(\mathbf{x}) = \frac{\mathbf{I}(\mathbf{x}) - \mathbf{A}}{\max(t_\theta(\mathbf{x}), t_0)} + \mathbf{A}, \quad (10)$$

where t_0 , which discourages division by numbers close to zero, is set to $t_0 = 0.1$ as suggested in [3]. In order to recover the missing airlight component, \mathbf{A} , we follow the method suggested in [3]: we first pick the 0.1% brightest pixels in the dark channel of the hazy image. Then, out of these locations we pick the brightest pixel in the hazy image, \mathbf{I} . That is the final chosen atmospheric light, \mathbf{A} .

F. Architecture

Our fully-convolutional, “Dilated Residual Network”, shown in Figure 2, is inspired by the Context Aggregation Network (CAN) [18], which has shown impressive results in dense-output applications. Similarly to CAN, we keep the resolution of all layers intact and identical to that of the input and output. In order to get an accurate prediction we avoid pooling and upsampling, and instead increase the receptive field via dilated convolutions with exponentially increasing dilation factors. Contrary to [18], between each dilated convolution we add another two regular convolution layers to create a richer nonlinear representation.

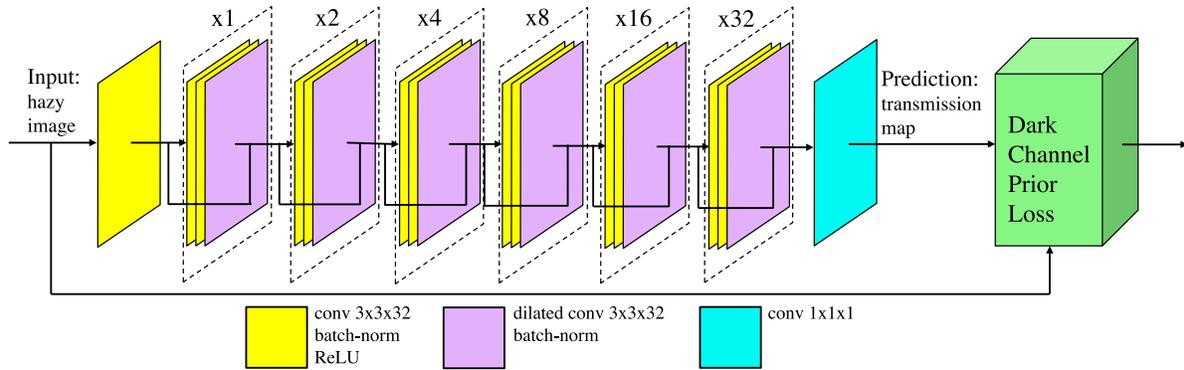


Fig. 2. System architecture. Our fully-convolutional network receives real-world hazy images. Apart from the input and output layers, our network is a cascade of dilated residual blocks (dilation written above each block), which gradually increase the receptive field. The network’s predicted transmission and the input image, are fed to the unsupervised, DCP loss.

Our network is thus built as a cascade of 6 dilated residual blocks; each made up of two regular convolutions, followed by a single dilated convolution. The dilation factors increase by a power of two from one block to the next. The filter size and width of all convolution layers (apart from the output) is $3 \times 3 \times 32$. All regular convolutions are followed by batch normalization [30] and ReLU nonlinearity [31], and all dilated ones are followed by batch-norm only. The final layer is a linear transformation to the output dimension of the transmission map $1 \times 1 \times 1$. To improve gradient flow and propagate finer details to the output, we incorporate additional Resnet-style [32] skip connections between the input and output of each block. The skip connection is a simple addition of the input to the output of each block.

IV. EXPERIMENTAL RESULTS

A. Dataset

In order to train and evaluate the performance of our network, we use the recent large-scale RESIDE (REAListic Single Image DEhazing) dataset [19]. RESIDE’s training set, called “ITS”, includes 13,990 synthetic indoor images, created from the NYU2 [17] and Middlebury stereo datasets [33]. The test set includes both indoor and outdoor sections, called “SOTS-indoor” and “SOTS-outdoor”,³ each containing 500 synthetic images. A smaller test set of 20 outdoor images, called “HSTS”, is also suggested. HSTS has a mix of 10 synthetic images (where ground truth is known) and 10 real-world images. All synthetic hazy images are created by first collecting ground-truth clean images with their corresponding depth maps and applying the haze formation model with different configurations of the \mathbf{A} , β parameters in (1). The beta version of RESIDE provides an additional collection of 4,322 real-world images, mined from the web, called “RTTS”. Instead of using the synthetic indoor database of ITS (or its variations based on NYU2 and Middlebury), as in [10], [12]–[14], we train our network on the real-world images of RTTS. For the evaluation of PSNR (Peak Signal to

Noise Ratio) and SSIM (Structural Similarity) criteria during training, we use as validation a subset of 500 images from RESIDE beta’s “OTS” synthetic outdoor training set. The 500 images are selected at random from Part-I of OTS, where we make sure that no images from the SOTS-outdoor and HSTS test sets are selected for validation.

B. Implementation Details

To enrich the RTTS training set, we perform data augmentation. The first augmentation is simply resizing the original hazy images to size 128×128 using bilinear interpolation. The second, third and fourth augmentations are performed randomly. Each image can be flipped horizontally or kept as is; randomly cropped to 256×256 or 512×512 , and rotated at 0, 45, 90, or 135 degrees. If rotated, only the valid center of the image is taken. All augmented images are then resized to 128×128 . The final number of training images is therefore: $4322 \times 4 = 17,288$.

The parameters of our loss function are taken exactly (no additional tuning) as suggested in [3], [29]: $\lambda = 10^{-4}$, $\omega = 0.95$, $t_0 = 0.1$, $\varepsilon = 10^{-6}$, DCP patch size: 15×15 , and soft matting patch size: 3×3 . We use the Adam optimizer [34] with batch size of 24; initial learning rate of $l_r = 3 \cdot 10^{-4}$, and exponential decay with factor 0.96 every 3 epochs. The network weights are initialized using random initialization with zero mean and variance of 0.1. Our method is implemented in TensorFlow on a GTX Titan-X Nvidia GPU. Training time to get the optimal solution (about 30 epochs) takes 8 hours. For outdoor results we stop the training at epoch 27, whereas for SOTS-indoor, we keep training until we reach 30 epochs. Our stopping criterion is explained further in section V-B.

C. Quantitative Evaluation

We evaluate the performance of our method on the SOTS-indoor, SOTS-outdoor and HSTS test sets. These test sets are created synthetically, therefore featuring both the clean images and their hazy versions. We measure the quality of our solution in terms of the PSNR and SSIM metrics. We obtain the original code and compare our results to the following prior-based approaches: DCP [3],⁴ BCCR [8] and NLD [7], and the following data-driven methods: CAP [6],

³Although in the latest published paper of RESIDE, SOTS-outdoor is not officially featured, the selection of 500 specific outdoor images are still available (as well as in earlier Arxiv versions) in RESIDE’s website in: https://www.dropbox.com/s/y6jupfvitv0dx5w/SOTS.zip?dl=0&file_subpath=%2F5FOTS

⁴Implementation in <https://github.com/sjtrny/Dark-Channel-Haze-Removal>

TABLE I

QUANTITATIVE PSNR/SSIM RESULTS OF OUR APPROACH (HIGHER IS BETTER). FOR BOTH SOTS-OUTDOOR AND HSTS WE REPORT THE RESULT OF EPOCH 27, WHEREAS IN SOTS-INDOOR WE REPORT THE RESULT OF EPOCH 30

	DCP [3]	BCCR [8]	NLD [7]	CAP [6]	MSCNN [12]	DehazeNet [11]	AOD-Net [10]	GFN [13]	Ours
<i>HSTS</i>	17.22/0.798	15.09/0.738	17.62/0.792	21.54/0.867	18.29/0.841	24.49/0.915	21.58/0.922	22.94/0.874	24.44/ 0.933
<i>SOTS-outdoor</i>	17.56/0.822	15.49/0.781	18.07/0.802	22.30/0.914	19.56/0.863	22.72/0.858	21.34/0.924	21.49/0.838	24.08/0.933
<i>SOTS-indoor</i>	20.15/0.872	16.88/0.791	17.29/0.749	19.05/0.836	17.11/0.805	21.14/0.847	19.38/0.849	22.32/0.880	19.25/0.832

TABLE II

QUANTITATIVE RESULTS ON THE MIDDLEBURY PART OF D-HAZY DATASET [36]

Method	DCP [3]	CAP [6]	MSCNN [12]	DehazeNet [11]	GAN-Unsup [20]	Ours (30)	Ours (33)
PSNR/SSIM	12.02/0.690	14.16/0.762	13.55/0.736	13.59/0.750	14.95/0.774	14.01/0.826	14.58/ 0.827

MSCNN [12], DehazeNet [11], AOD-Net [10] and GFN [13]. In case the dehazed images are out of the range $[0, 1]$, we normalize them to $[0, 1]$ only if it improves PSNR and SSIM values.

The numeric results⁵ are given in Table I. We get the highest PSNR and SSIM scores among all other methods in the larger SOTS-outdoor, and the highest SSIM in the smaller HSTS. Our method, represented by a rich neural network and trained to accommodate numerous images, obtains better results compared to prior-based methods. Specifically, compared to DCP, our method strives to approximate the solution of the same energy function, but we stop it before reaching an absolute minimum in order to get further regularization. This is particularly noticed in outdoor images where DCP often over-saturates the sky.

With regard to data-driven approaches, our high score is attributed to the fact that we train on *real-world outdoor* images, whereas competing methods [10], [12], [13] concentrate on *synthetic indoor* images and suffer from a certain domain shift when addressing outdoor data. In addition, the synthetic hazy and clean pairs are created from coarse depth data for which training creates a negative bias towards data-driven approaches. An example of an indoor training image in ITS is given in Figure 5. Notice the rough misplaced edges in the transmission map which later translate to inaccurate hazy images. Indeed, our closest competitor in terms of outdoor results is DehazeNet [11]. Recall that this method is trained on a large variety of clean image patches of outdoor scenes, making it more robust compared to methods trained on ITS.

We include the results of our method on SOTS-indoor in which it performs favourably, but gets a lower score compared to other data-driven methods and even DCP. This is expected since we train on outdoor images, creating a tradeoff between indoor and outdoor performance. As for DCP, it behaves more agreeably on indoor images which coincide better with the haze formation model and do not include sky regions.

D. Comparison to Unsupervised GAN [20]

We further compare our method with the unsupervised GAN-based method, introduced in [20]. We provide the PSNR and SSIM results of our method, along with those of DCP [3], CAP [6], MSCNN [12], DehazeNet [11] and [20], which we

⁵Our results slightly differ from [19]. We use the original DCP [3], whereas they use the faster version [35] with worse quality.

refer to as “GAN-Unsup”. We adopt the D-HAZY dataset [36] as suggested in [20], and report our results on the Middlebury set of images. The D-HAZY dataset consists of synthetic indoor images, thus similarly to RESIDE’s SOTS-indoor, we provide the results of our method trained for 30 epochs. In addition, since the set of images contain relatively thicker haze, we also add the results of our model, trained for 33 epochs. As can be seen from Table II, for 33 epochs of training we provide the second best PSNR result and the best SSIM result. As opposed to [20], which was both trained and tested on indoor synthetic images, our model was fed with only real-world outdoor images, demonstrating its impressive generalization power.

E. Qualitative Results

We present qualitative results on HSTS in Figure 3. In the top part of Figure 3, it can be seen that our method maintains the true colors of the original image, whereas DCP [3], BCCR [8] and NLD [7] tend to produce exaggerated sky regions. Our results are similar to those produced by CAP [6], however slightly closer to the true colors exhibited in the ground truth image. In the bottom half of Figure 3 we provide a comparison to deep-learning based methods. In most images we maintain the true contrast and colors, whereas MSCNN [12] and GFN [13] provide more contrast-enhanced images. At times, we slightly change the color of the sky, which is to be expected since our method is unsupervised and does not witness the clear images at any stage. In Figure 4, one can see a real-world image comparison of our results with both prior-based and data-driven methods. We display the output of our network after 27 epochs (the optimal results for the OTS validation set we use) and after 30 epochs, where the produced images are more similar to DCP (see discussion on sec. V-A). One can see that after 27 epochs we do not remove all of the haze, perhaps indicating that the outdoor images in RESIDE are less hazy than real-world hazy images. For 30 epochs, our result is more saturated and of higher contrast.

F. Runtime Comparison

Apart from improving the overall PSNR and SSIM performance of DCP, we hereby show that we are as efficient as fast implementations of DCP. Our inference procedure consists of two parts: a forward-pass over the trained network to obtain the predicted transmission map (performed in TensorFlow), and reconstruction using Equation 10 (performed in Numpy).

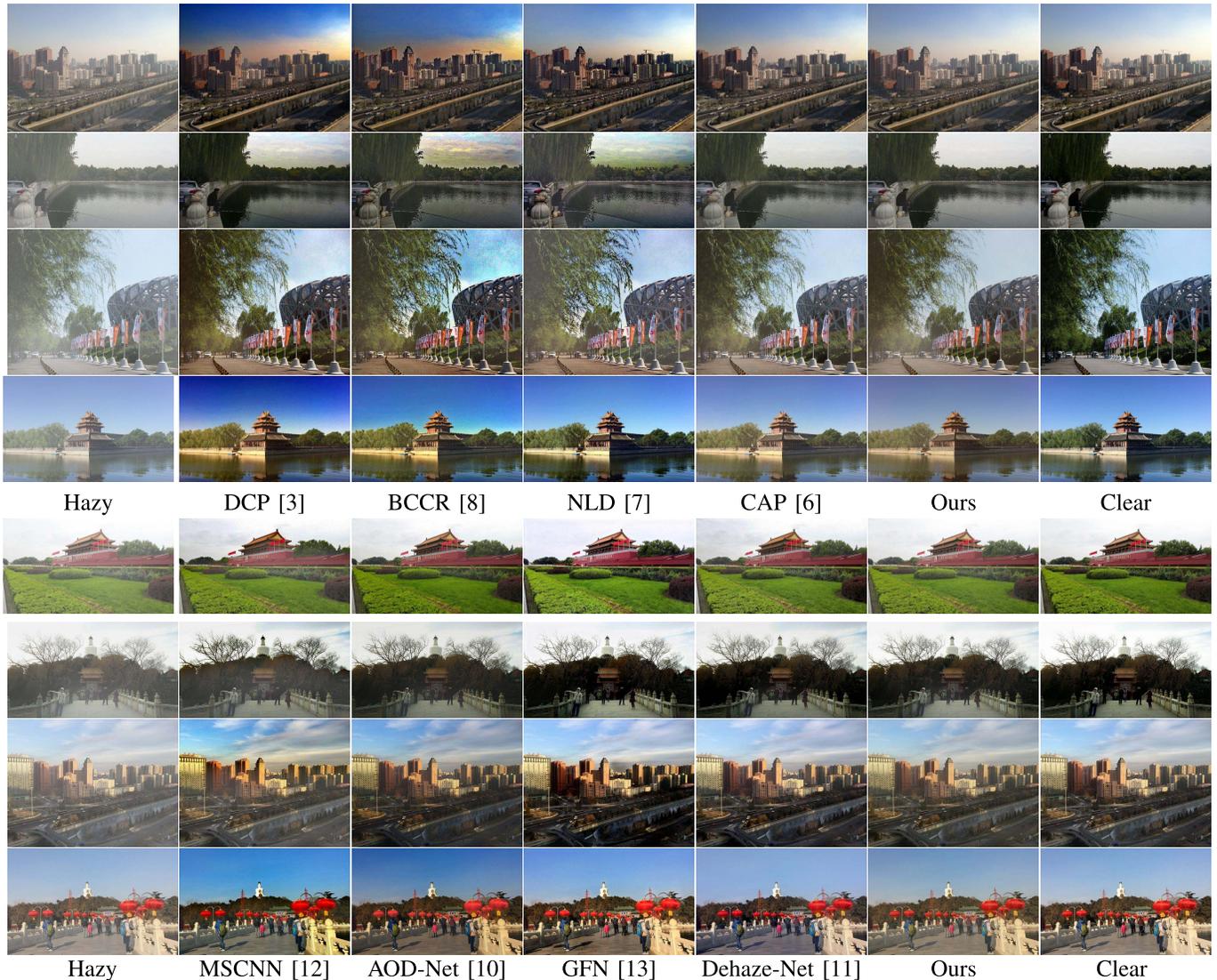


Fig. 3. Qualitative results on RESIDE’s HSTS. Upper half: comparison to prior-based methods; bottom half: comparison to deep-learning-based methods.

We compare ourselves to a Matlab implementation⁶ of soft matting DCP [3], denoted as “slow-DCP”, and guided image filter DCP [35], denoted as “fast-DCP”. Note that fast-DCP is an approximation of slow-DCP and though being very efficient, achieves inferior results. Although Matlab is more efficient than Numpy and TensorFlow, we do get the benefit of using the GPU. Thus for fair comparison, we include both GPU and Intel(R) i7-5930k 3.5GHz CPU runtimes of our solution. In Table III, we report the average runtimes (lower is better) over the 500 images in SOTS-outdoor which feature varying widths of ~ 500 pixels. Compared to the explicit optimization of slow-DCP, our feed-forward inference is much faster ($\times 77$ in GPU and $\times 27$ in CPU). We perform slightly faster than fast-DCP ($\times 3.8$ in GPU and $\times 1.3$ in CPU), and we supply results of much better visual quality which translate to a ~ 9.5 dB increase in PSNR. Additional speed-up of our method can be performed by joint estimation of \mathbf{t} , \mathbf{A} during training, but we leave this for future work. To conclude,

⁶Code given in: <https://github.com/sjtrny/Dark-Channel-Haze-Removal>. We use the regular guided image filter implementation, without subsampling.

as efficient as the DCP explicit solution may be, it will lack the additional regularization obtained by our approach.

G. Limitations of Our Method

Figure 6 shows some failure cases of our approach. Although we train our network on a variety of real-world hazy images, our method performs poorly on images with very high degrees of haze. This is often the case for other data-driven methods, such as [6], [11], [12]. In our case, the residual haze is mainly due to the choice of stopping point that currently optimizes PSNR performance on RESIDE’s OTS benchmark, biased towards images with mild haze. An alternative strategy would be choosing the stopping point so as to optimize the performance on a multitude of benchmarks, including indoor, outdoor, mild and heavy haze images. If the final goal of the dehazing algorithm is enhancement and display to a human observer, one can instead optimize a no-reference perceptual quality score as NIQE [37] (see discussion in V-D).

Additionally, as can be seen in Figure 3, our method sometimes distorts the overall color of the sky (see rows 4,7,8). This shift in color may be due to inaccurate estimation

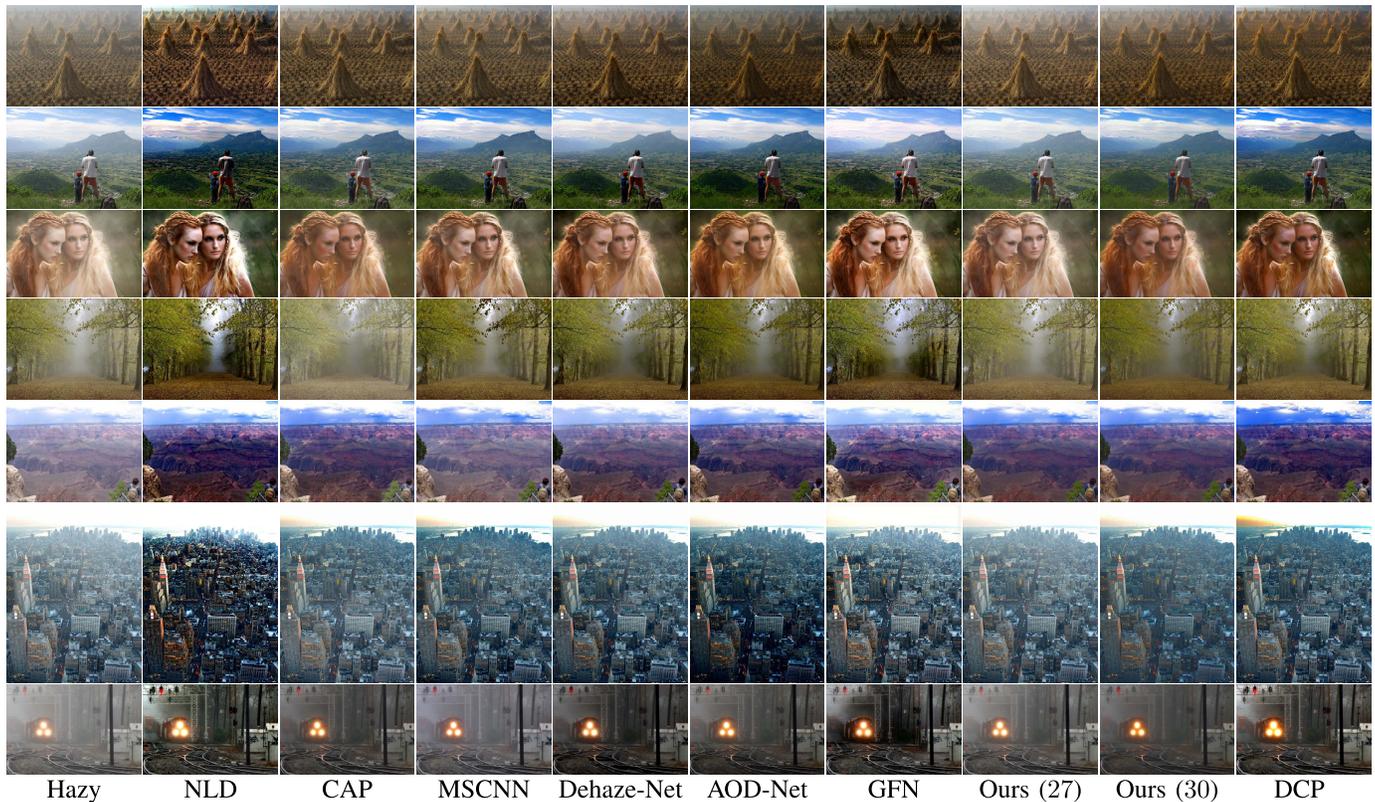


Fig. 4. Qualitative results of single image dehazing on real-world images. The numbers in parenthesis are the number of epochs spent training our network.



Fig. 5. Example of a synthetic image and its coarse transmission map from RESIDE’s ITS training dataset [19].

TABLE III

AVERAGE RUNTIME AND PERFORMANCE OF SOTS-OUTDOOR

	slow-DCP [3]	fast-DCP [35]	ours-CPU	ours-GPU
PSNR/SSIM	17.56/0.822	14.62/0.752	24.08/0.933	24.08/0.933
runtime[sec]	21.67	1.08	0.81	0.28

of the atmospheric airlight component. Instead of using a constant airlight vector, derived from the dark channel prior itself as suggested in [3], we can model the contribution of the atmosphere as a separate spatially-varying channel, and estimate it using the DNN, exactly as performed for the transmission map. This direction, however, is currently left for future work.

V. DISCUSSION

A. Proximity to Dark Channel Prior

During training, our network strives to approximate the DCP energy function. Since it optimizes the loss for the entire corpus of images, it may output different results from DCP [3]. While DCP operates on one image at a time, our network

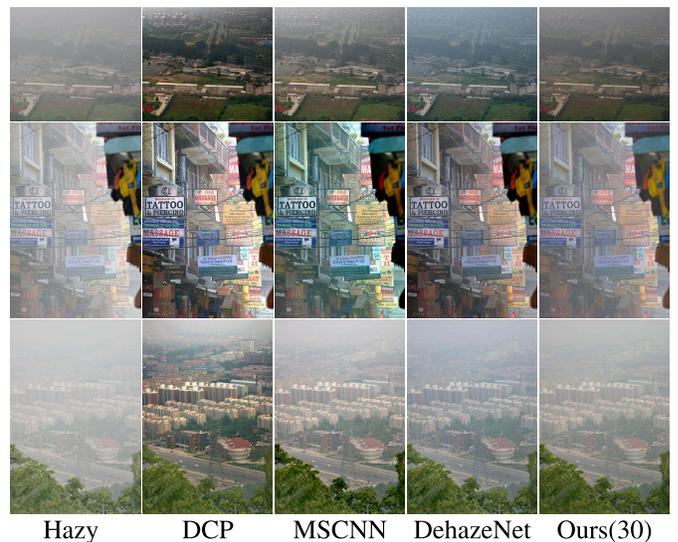


Fig. 6. Example failure cases of our approach.

learns a more “universal solution”, suited for multiple images. In addition, as the epochs evolve and the loss value decreases, we reach closer and closer to DCP, as can be seen in the three rightmost columns in Figure 4. At earlier epochs the output images still contain a large amount of haze, whereas further on, most of the haze is lifted, but the colors appear more saturated, even non-realistic. We search for a middle-ground where most of the haze is removed and one can see the details, but the colors and contrast remain realistic and physically valid. The benefit of stopping before reaching a deeper optimum

of DCP is especially noticeable in sky regions where DCP would output an exaggerated and amped-up version of the sky, whereas we produce a more natural color. In our case this “sweet-spot” is reached after 27 epochs over the training data. Nonetheless, we can keep training for a few more epochs to get more vivid results which may be more pleasant to the human eye.

B. Unsupervised Training Regime

Although our training is completely unsupervised, we do need a stopping criterion since reaching the minimum of the loss function is not always beneficial in terms of the visual and quantitative results. To do so, we evaluate the average loss value, PSNR and SSIM of a small supervised set of 500 images from OTS (not part of SOTS-outdoor or HSTS). A typical behaviour of the results is a decrease of the average loss; an increase in performance in PSNR and SSIM; reaching a maximum, and then, a decrease of these criteria. We choose the epoch/model that gave the best performance on the validation set from OTS. The learning parameters are chosen using a similar technique. We do note that choosing the training stopping point using the PSNR and SSIM metrics, does require a certain degree of supervision. This supervision, however, determines a single scalar (the number of training epochs) – a small measure as compared to the hundreds of thousands of learned parameters of our DNN. A completely unsupervised alternative may be suggested, optimizing a no-reference perceptual quality score as NIQE [37], requiring no supervision whatsoever, but naturally leading to higher distortion. A future direction of our work includes experimenting with a combination of both distortion-based and no-reference perceptual metrics.

C. Treatment of Sky Areas

The original DCP in [3] has been shown to produce artifacts such as noise amplification and exaggerated contrast in sky areas. The weighted Guided Image Filter (WGIF), proposed in [38], models the dehazing application as an image enhancement problem, and shows the amplification of DCP in sky areas is $\times 9$. The authors of [38], [39] suggest tackling this issue by adding a constant sky compensation term to the initial transmission map, provided by DCP. They target the sky areas exclusively by observing the histogram of the hazy image and adding the constant term according to the amount of haze.

Although our method is based on the DCP energy function, we perform early stopping mid-training, before reaching the absolute minimum of the loss function. As a result, we do not suffer as much from noise amplification in sky areas, as can be seen in rows 1, 2 in Figure 3 and rows 2, 5, 6 in Figure 4. This contributes greatly to our high PSNR values in outdoor images, compared to DCP. This benefit, however, is a side-effect and we do not address the sky regions directly. An interesting modification of our scheme is adding the sky compensation term straight into the energy module, but we leave this important direction for future work.

D. Perception Versus Distortion

An important issue that often arises in image restoration problems, such as single image dehazing, is the perception-distortion tradeoff, as introduced in [40]. On the one hand, one requires a dehazing algorithm to provide low distortion – the reconstructed image should be faithful to the ground truth clear day image. This is enforced by minimizing distortion-based metrics, such as PSNR or SSIM. On the other hand, one wishes the obtained image to be vivid, sharp and full of contrast and detail – perceptually pleasant to a human observer. This can be performed either by maximizing no-reference quality scores, for example NIQE [37], or displaying the reconstructions to a GAN or a human, and deciding whether the image is realistic or fake. The work reported in [40] postulates that perception and distortion are always at the expense of the other.

Indeed, we observe the same phenomenon in single image dehazing. Methods which provide better distortion – higher PSNR values, such as CAP [6], DehazeNet [11] and ourselves, produce reconstructions with less contrast, and even some residual haze. Other methods, such as DCP [3], NLD [7] and BCCR [8], provide more vivid and sharp results. This, however, at the expense of amplified noise, heightened contrast and non-realistic colors, leading to lower PSNR values. This difference in behaviour between data-driven and prior-based methods, may stem from the fact that the former are optimized to provide high PSNR values on a select benchmark.

We provide two possible solutions, one for 27 epochs of training, providing better distortion, though with residual haze, and the other, for 30 epochs of training, producing more vivid and colorful results, closer to DCP. We can easily control this tradeoff by choosing a different stopping point of the network, based on the metric and the dataset we wish to optimize. The final choice of the model depends on the end-user, whether it be a human, requiring better perception, or part of a computer vision pipeline, perhaps requiring lower distortion.

VI. CONCLUSION

We have presented a method of unsupervised training of deep neural networks for the purpose of single image dehazing. Our method relies on the well-known Dark Channel Prior (DCP) [3] and manages to improve it considerably. In addition to providing state-of-the-art performance in outdoor scenarios, our method also eliminates the need for synthetic training sets. While our focus here is DCP, we could have incorporated any other successful energy function, using it as our unsupervised loss. Our future research is focused on finding an even better *combination* of energy functions, or incorporating some amount of supervision to benefit from both worlds.

REFERENCES

- [1] R. Fattal, “Single image dehazing,” *ACM Trans. Graph.*, vol. 27, no. 3, Aug. 2008, Art. no. 72.
- [2] R. T. Tan, “Visibility in bad weather from a single image,” in *Proc. CVPR*, Jun. 2008, pp. 1–8.
- [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] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *Proc. ICCV*, Sep. 2009, pp. 2201–2208.
- [5] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, Dec. 2014, Art. no. 13.
- [6] 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.
- [7] D. Berman, T. Tali, and S. Avidan, "Non-local image dehazing," in *Proc. CVPR*, Jun. 2016, pp. 1674–1682.
- [8] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. ICCV*, Dec. 2013, pp. 617–624.
- [9] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in *Proc. ACCV*, 2010, pp. 501–514.
- [10] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proc. ICCV*, Oct. 2017, pp. 4770–4778.
- [11] 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.
- [12] 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. ECCV*, Sep. 2016, pp. 154–169.
- [13] W. Ren *et al.*, "Gated fusion network for single image dehazing," in *Proc. CVPR*, Jun. 2018, pp. 3253–3261.
- [14] Z. Xu, X. Yang, X. Li, X. Sun, and P. Harbin, "Strong baseline for single image dehazing with deep features and instance normalization," in *Proc. BMVC*, 2018, p. 5.
- [15] A. Saxena, M. Sun, and A. Y. Ng, "Make3D: Learning 3D scene structure from a single still image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 824–840, May 2009.
- [16] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [17] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgb-d images," in *Proc. ECCV*, 2012, pp. 746–760.
- [18] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," Nov. 2015, *arXiv:1511.07122*. [Online]. Available: <https://arxiv.org/abs/1511.07122>
- [19] B. Li *et al.*, "Benchmarking single image dehazing and beyond," Dec. 2017, *arXiv:1712.04143*. [Online]. Available: <https://arxiv.org/abs/1712.04143>
- [20] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Proc. AAAI*, Apr. 2018, pp. 1–8.
- [21] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. CVPR*, Jun. 2000, pp. 598–605.
- [22] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Polarization-based vision through haze," *Appl. Opt.*, vol. 42, no. 3, pp. 511–525, Jan. 2003.
- [23] J. Kopf *et al.*, "Deep photo: Model-based photograph enhancement and viewing," *ACM Trans. Graph.*, vol. 27, no. 5, Dec. 2008, Art. no. 116.
- [24] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, Jul. 2012.
- [25] K. B. Gibson and T. Q. Nguyen, "An analysis of single image defogging methods using a color ellipsoid framework," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, Dec. 2013, Art. no. 37.
- [26] I. Omer and M. Werman, "Color lines: Image specific color representation," in *Proc. CVPR*, Jun. 2004, p. 2.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [28] W. K. Middleton, "Vision through the atmosphere," in *Geophysik II/Geophysics II*. Springer, 1957, pp. 254–287.
- [29] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," in *Proc. CVPR*, Jun. 2006, pp. 61–68.
- [30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.
- [33] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. CVPR*, Jun. 2003, p. 1.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [35] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [36] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. ICIP*, Sep. 2016, pp. 2226–2230.
- [37] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
- [38] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 120–129, Jan. 2015.
- [39] Z. G. Li and J. H. Zheng, "Edge-preserving decomposition-based single image haze removal," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5432–5441, Dec. 2015.
- [40] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.



Alona Golts received the B.Sc. and M.Sc. degrees from the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel, in 2010 and 2015, respectively, where she is currently pursuing the Ph.D. degree with the Department of Computer Science. Her research interests include deep learning, inverse problems, and sparse representations.



Daniel Freedman received the A.B. degree (*Magna Cum Laude*) in physics from Princeton University in 1993, and the Ph.D. degree in engineering sciences from Harvard University, in 2000. From 2000 to 2009, he served as an Assistant Professor and an Associate Professor with the Computer Science Department, Rensselaer Polytechnic Institute (RPI), Troy, NY, USA. In 2007, he became a Fulbright Fellow and a Visiting Professor of applied mathematics and computer science with the Weizmann Institute of Science. He worked in a number of research positions with HP Labs, IBM Research, Microsoft Research, and finally Google Research. In addition to the Fulbright Fellowship, he received the National Science Foundation CAREER Award.



Michael Elad received the B.Sc., M.Sc., and D.Sc. degrees from the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel, in 1986, 1988, and 1997, respectively. Since 2003, he has been a Faculty Member with the Department of Computer Science, Technion–Israel Institute of Technology, where he has also been a Full Professor, since 2010. He works in the field of signal and image processing, specializing in inverse problems, and sparse representations. He was a recipient of numerous teaching awards, the 2008 and 2015 Henri Taub Prizes for Academic Excellence, and the 2010 Hershel-Rich prize for innovation. He is a SIAM Fellow in 2018. Since January 2016, he has been the Editor-in-Chief of the *SIAM Journal on Imaging Sciences*.