Unsupervised Single Image Dehazing Using Dark Channel Prior Loss

https://arxiv.org/abs/1812.07051

Student Name	ID	E-mail
Tony Bishara	316347202	tony.bishara@campus.technion.ac.il
Muhammad Dahamshi	318396926	muhammad.dah@campus.technion.ac.il

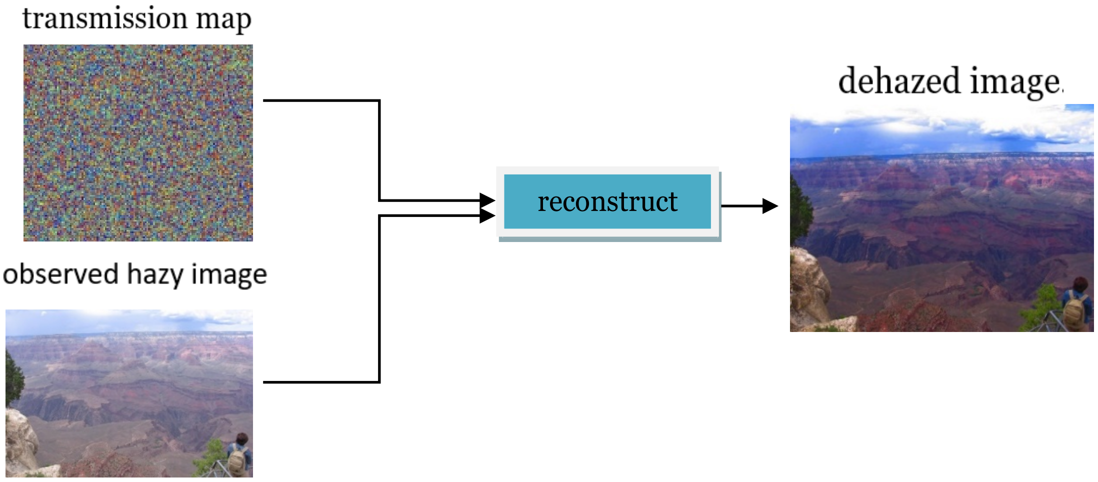
Digital Image Processing - Final Project

1 Introduction

The presence of haze in bad weather will result in poor visibility and contrast lost in images, as a result, a lot of bad impacts will arise on computer vision applications, such as outdoor surveillance, object recognition and tracking, unmanned vehicle systems etc. and make the performance of these vision applications lose authenticity and practical value because all these vision systems assume that the radiance from a scene point to observers is not altered by the airlight and intermediate medium. Therefore, to obtain the haze-free images and recover color, visibility, and details of scene from degraded images, haze removal from images is inevitable.
Recently, image dehazing has become an important and urgent research problem in the field of computer vision, it has been receiving more and more increasing interests. However, since it is depth dependent, some classical image enhancement techniques, such as histogram equalization, histogram specification, cannot receive satisfying effect. Dehazing techniques from the perspective of physical model of degraded images have emerged.

The existence of haze affects an image in two aspects:
1) Attenuation the scene radiance with correspondence to an object's distance from the camera.
2) 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. Given this transmission maps, we can reconstruct the dehazed image.



Hence, we have to learn the underlying transformation between hazy and clear images

2 RELATED WORK

Early attempts at image dehazing have incorporated several images of the same scene, taken at different bad weather conditions, or using different polarization filters. These methods, however, require thousands of input and output examples.

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.

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

3 Problem Setting

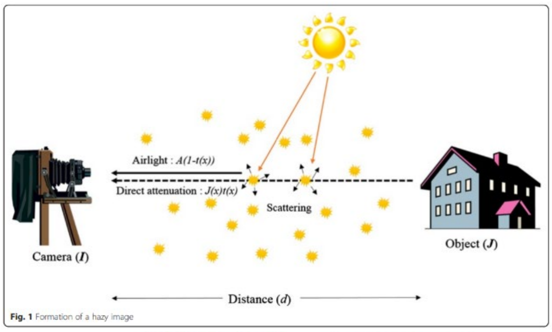
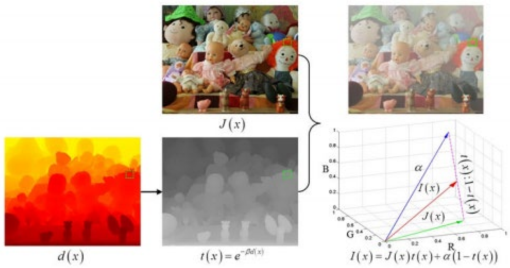


Fig. 1 Formation of a hazy image

Let $I(x) \in R^{Nx3}$ be the image that would have been observed (where N is the number of image pixels). The popular haze formation model in is given as:

$$(1)\ I(x) = t(x)J(x) + \big(1 - t(x)\big)A$$
$$t(x) = e^{-\beta d(x)}$$



$I(x)$ is a convex linear combination of the haze-free scene radiance, J(x), and the atmospheric light constant component, A, called the airlight. The transmission map coefficients, t(x) $\in R^N$ control the relative force of each component, in each pixel in the image, x $\in R^N$. The transmission t(x) is a function of the depth, d(x), of the scene from the observer.

Unfortunately, inverting eqn. 1 is difficult. To do so, however, one needs to solve a set of 3N equations (only I(x) is given), with 4N + 3 unknowns (J(x); t(x);A). Thus, additional prior knowledge of the images in question is needed.

4 The Assumptions and the Model

The authors propose a completely unsupervised method of training via minimization of the well-known, Dark Channel Prior (DCP) energy function. Assumed 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.

The authors' hypothesis is that the space of natural images as well as the space of natural refractive distortions is structured enough that a neural network can learn a reasonable mapping between hazed input images and dehazed output images.

Inspired by this, the authors train a network to predict the transmission map such that "Deep DCP" technique can be regarded as a fast approximator of DCP. improves its results significantly. This suggests an additional regularization obtained via the network and learning process.

4.1 Training objectives

Dark Channel Prior:
To solve the single image dehazing problem. propose a very simple but effective dark channel prior, which is a statistical assumption based on their observations of plenty of haze-free outdoor images. According to the dark channel prior, in most of the non-sky local image areas, at least one color channel will have pixels with very low intensity, in other words, the minimum intensity in such local patch should have a very low value, even close to zero and three main factors attribute the low intensity in the dark channel, that is, shadows, colorful objects and dark objects.

For any image, $J(x)$, the dark channel is defined as follows:

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

Here, $J^c$is a color channel of J, and $\Omega(x)$ is a small patch, centered around x. According to the dark channel prior, in most of non-sky local image areas, we have, The “dark channel” of the image is defined as:

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

With this statistical prior as theory constraints, He et al. suggest a 15*15 patch size to get dark channel with the assumption that depth is always same in a local patch, that is, the transmission $\tilde{t}(x)$ constant in the window, then the following eq. can be deduced:

$$\tilde{t}(x) = 1 - \omega \cdot \min_{c \in \{r,g,b\}} \left(\min_{y \in \Omega(x)} \left(\frac{I^c(y)}{A} \right) \right)$$

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 $I/A \to 1$, thus $\tilde{t}(x) \to 0$. The resulting coarse transmission map requires an additional step of refinement.

However, the assumption will fail when there are abrupt depth jumps in the patch, halo effects will appear and the bigger the patch size, the more serious the effects. To solve that we use the soft-matting operation as described below:

Soft Matting:
The soft matting method works very well for halo removal, however, this operation will have expensive computational cost, therefore, lots of improved algorithms appear recently aiming at solving this problem. In which, propose a transmission map gradient prior law based on dark channel prior.

In fact, assuming constant transmission in a local patch Ω(x) is inappropriate. The transmission obtained is only a coarse estimation, the recovered image using the coarse transmission contains severe halo artifacts, so it is necessary to refine the transmission map and capture the depth changes at object edges. the closed-form matting framework is applied to suppress the blocky artifacts in the coarse transmission map, the following energy function suggested in the "A closed form solution to natural image matting" paper can be used to acquire the refined map:

$$E(t;\tilde{t}) = t^T Lt + \lambda(t - \tilde{t})T\ (t - \tilde{t})$$

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. The matrix L is a Laplacian-like matrix, dedicated to image matting and given by:

$$L_{ij} = \sum_{n|(i;j)\in p_n} \left(\delta_{ij} - \omega_{ij}^n\right) , \forall i,j = 1, \dots, N$$

$$\omega_{ij}^n = \frac{1}{|p_n|}(1 + (I_i - \mu_n)^T \left(\Sigma_n + \frac{\epsilon}{|p_n|}U_3\right)^{-1} (I_i - \mu_n)$$

where i; j are two pixels within a small patch pn around pixel n; $|p_n|$ is the size of the patch and equal to 3 x 3 = 9; μ_n, Σ_n are the mean and covariance of the patch; U3 is the identitly matrix; and ϵ is a fixed smoothing parameter.

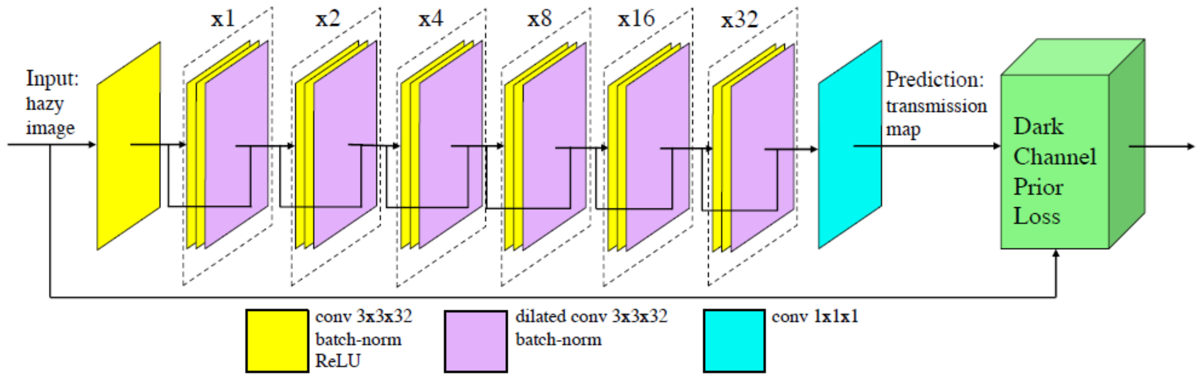
Implementation as a Loss Function:

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

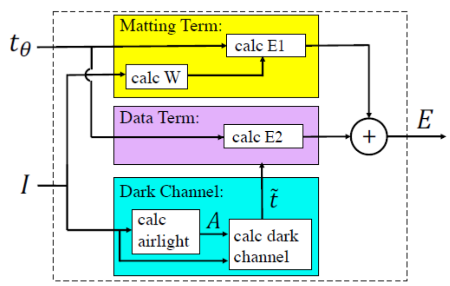
$$\theta^* = \underset{\theta}{\text{argmin}} \left(\frac{1}{M}\sum_{m=1}^{M} E(t_\theta; \tilde{t}(I_m)\right)$$

5 The Architecture

The fully-convolutional, “Dilated Residual Network”, shown in Figure below, is inspired by the Context Aggregation Network (CAN), 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. between each dilated convolution we add another two regular convolution layers to create a richer nonlinear representation.



The loss module, which receives the prediction of the network, t_θ, along with the hazy image, I, and outputs the value of the DCP energy loss.



6 Authors' Results

The test set includes both indoor and outdoor sections, called "SOTS-indoor" and "SOTS-outdoor". each containing 500 synthetic images.

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
HSTS	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**
SOTS-outdoor	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**
SOTS-indoor	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: Average runtime and performance of SOTS-outdoor

	slow-DCP [3]	fast-DCP [34]	ours-CPU	ours-GPU
PSNR/SSIM	17.56/0.822	14.62/0.752	**24.07**/**0.933**	**24.07**/**0.933**
runtime[sec]	21.67	1.08	1.71	**0.67**

As can be seen, integrating additional sophisticated elements into the loss function does improve the results; the feed-forward inference is much faster (x30in GPU and x12 in CPU).

7 Review

We credit this project for the following achievements:

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 boost in outdoor PSNR over classical DCP, validating an effective regularization.

3) It treats the sky successfully where DCP typically fails.

4) It is the first 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.

8 Our Contribution

This section will consists of a research based on Single Image Haze Removal Using Dark Channel Prior paper.

8.1 Estimating the Atmospheric Light

We have been assuming that the atmospheric light A is known. In this section, we propose a method to estimate A. In the previous works, the color of the most haze-opaque region is used as A or as A's initial guess. However, little attention has been paid to the detection of the "most haze-opaque" region.

The brightest pixels in the hazy image are considered to be the most haze-opaque. This is true only when the weather is overcast and the sunlight can be ignored. In this case, the atmospheric light is the only illumination source of the scene. So, the scene radiance of each color channel is given by:

$$J(x) = R(x)A$$

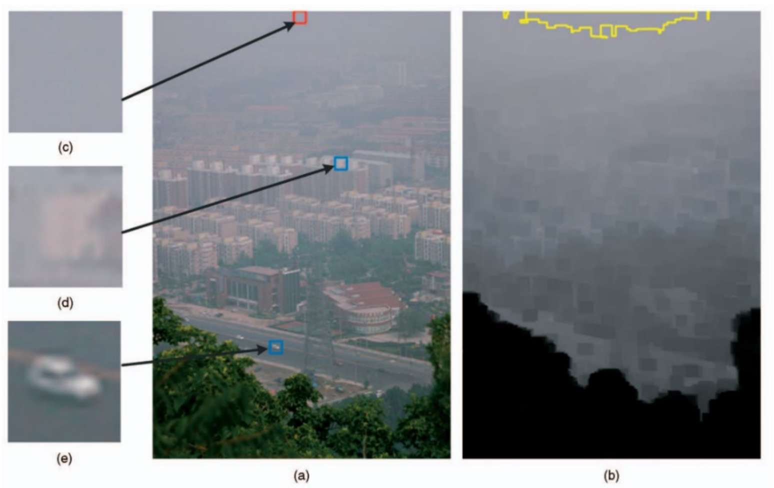
where $R \leq 1$ is the reflectance of the scene points. The haze imaging equation (1) can be written as:

$$I(x) = t(x)R(x)A + \big(1 - t(x)\big)A$$
$$\rightarrow I(x) \leq \mathrm{A}$$

When pixels at infinite distance ($t \approx 0$) exist in the image, the brightest I is the most haze-opaque and it approximately equals A. Unfortunately, in practice we can rarely ignore the sunlight. Considering the sunlight S, we modify the above eq by:

$$J(x) = R(x)(S + A)$$

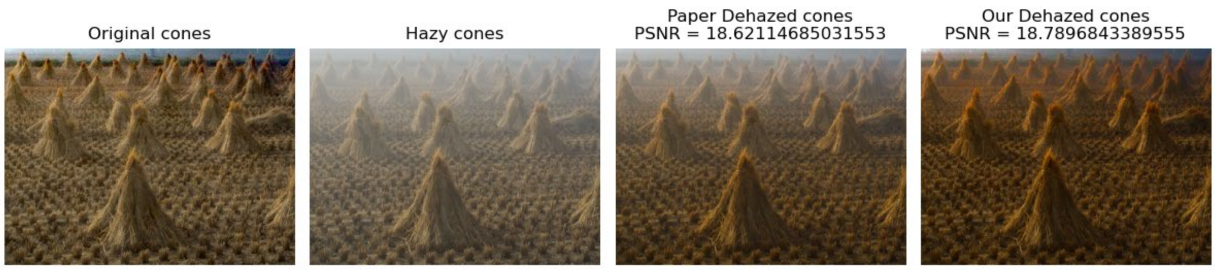
$$I(x) = R(x)St(x) + R(x)At(x) + \big(1 - t(x)\big)A$$



Estimating the atmospheric light. (a) Input image. (b) Dark channel and the most haze-opaque region. (c) The patch from where our method automatically obtains the atmospheric light. (d), (e) Two patches that contain pixels brighter than the atmospheric light.

8.1 Our Contribution Results

We tried to Estimate the Atmospheric Light. For this we downloaded (partially) relevant pictures of ImageNet, cleaned it up a bit and ran the best network of the article on it. On figure below one can find comparison of pictures in terms of PSNR. In middle is the result of the paper model and in the right our improvement. We can see that the PSNR value increases.



9 Limitations

While working on paper we paid attention to the next facts:

9.1 The haze model the paper does not always hold in sky regions, we suggest integrating a hybrid image defogging approach, for example: https://arxiv.org/abs/2007.06492

9.2 Atmospheric light estimation: the atmospheric light is reliably estimated from the dark channel, especially when the dark channel is obtained using a large local patch. Therefore, if the local patch size used in dark channel construction is not large enough, it is recommended to use an additional dark channel with a larger local patch size only for atmospheric light estimation. The use of local entropy is also found to be effective in enhancing the estimation accuracy because atmospheric light estimation from bright objects can be prevented.

9.3 Transmission map estimation: the under-estimation problem of the transmission map is addressed. The conventional gain and offset control methods are examined, but an adaptive correction scheme is found to be necessary for precise estimation of the transmission map, which is missing in the paper.

9.4 Transmission map refinement: the performance of transmission map refinement is improved when a hazy image is used as a guidance image. The soft matting method shows the best transmission map estimation accuracy, and the guided and cross-bilateral filters show the second-best accuracy. The Gaussian and guided filters perform best in terms of the computational complexity, but the guided filter is most memory-inefficient among the five investigated refinement schemes.

9.5 Quality metric for image dehazing: the performance of the image dehazing can be indirectly measured by comparing the ground-truth and estimated transmission maps. The conventional quantitative quality metrics using only the dehazed image are investigated, but they are found to be not trustworthy enough. An advanced or application-specific quality metric needs to be developed.

Code references:
https://github.com/AlonaGolts/Deep_Energy
https://github.com/He-Zhang/image_dehaze
https://github.com/joyeecheung/dark-channel-prior-dehazing

papers references:
https://arxiv.org/abs/1812.07051
https://arxiv.org/abs/2007.06492
http://mmlab.ie.cuhk.edu.hk/archive/2011/Haze.pdf