WEAKLY SUPERVISED FOG DETECTION

Adrian Galdran[a,*], *Pedro Costa*[a], *Javier Vazquez-Corral*[b], *Aurélio Campilho*[a,c]

[a] INESC TEC Porto
R. Dr. Roberto Frias, 4200
Porto (Portugal)

[b] Universitat Pompeu Fabra
Carrer de Roc Boronat, 138, 08018
Barcelona (Spain)

[c] Faculty of Engineering UP
R. Dr. Roberto Frias, 4200
Porto (Portugal)

ABSTRACT

Image dehazing tries to solve an undesired loss of visibility in outdoor images due to the presence of fog. Recently, machine-learning techniques have shown great dehazing ability. However, in order to be trained, they require training sets with pairs of foggy images and their clean counterparts, or a depth-map. In this paper, we propose to learn the appearance of fog from weakly-labeled data. Specifically, we only require a single label per-image stating if it contains fog or not. Based on the Multiple-Instance Learning framework, we propose a model that can learn from image-level labels to predict if an image contains haze reasoning at a local level. Fog detection performance of the proposed method compares favorably with two popular techniques, and the attention maps generated by the model demonstrate that it effectively learns to disregard sky regions as indicative of the presence of fog, a common pitfall of current image dehazing techniques.

Index Terms— Fog Detection, Image Dehazing, Weakly-Supervised Learning, Multiple-Instance Learning

1. INTRODUCTION

Image dehazing, or fog removal, is the task of improving the quality of images degraded by bad-weather, in order to enhance the performance of further computer vision applications, or just achieve a more pleasant image visualization. Image dehazing has received much attention from the image processing community in the last years. Initial attempts to solve the ill-posed problem of fog removal relied on multiple inputs [1]. More recently, single-image dehazing techniques [2, 3, 4] achieved satisfactory results without any supplementary input. This is accomplished by imposing prior knowledge on statistics of non-degraded natural images, like the Dark Channel Prior [5], or the Color Attenuation Prior. Variants involving non-local information [6], color lines [7], or image fusion [8] have also been proposed.

Recently, machine learning-based image dehazing has reached great levels of performance. In this setting, a non-linear mapping between foggy images and either fog-free counterparts or depth maps is learned from training data, in a realization of a fully-supervised regression problem [9, 10, 11]. If depth maps are regressed, these can then be used to increase contrast in far-away areas of the scene. Unfortunately, it is challenging to collect training data that contains pairs of foggy and fog-free/depth images of exactly the same scene. Hence, current methods are trained on haze-free images on which a layer of synthetic fog is added.

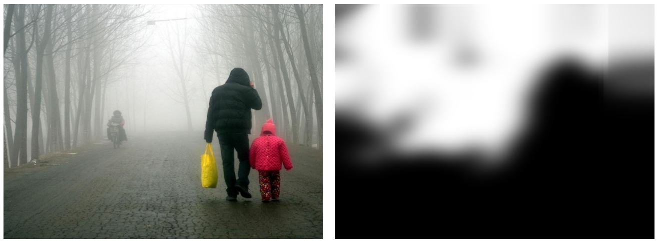

Fig. 1. Left: A hazy image. Right: Attention map produced by our method, indicating which regions of the scene trigger the classifier decision, and showing that the model infers local image properties learning only from image-level labels.

Weakly-supervised learning offers an alternative to the fully-supervised regime on which we only require ground-truth information at the image level. In our case, instead of needing a correspondence between hazy pixels and their haze-free counterparts, we require only a single label per-image, whether the image contains fog. This kind of training data is much easier to collect, and our proposed weakly-supervised method can extract useful local information from images using only this image-level information, as shown in Fig. 1.

In this paper, we introduce a weakly-supervised method for the task of fog detection, *i.e.* predicting the presence of fog within an image. The interest for this task stems from the fact that current dehazing methods, when applied to clean images, can produce artifacts and unpleasant results, as has been recently shown [12]. In these cases, it can be better to avoid processing the image at all. Fog detection can also be useful for comparing different algorithms, or deciding if the fog removal step has been successful, or any relevant parameter should be adjusted to achieve greater dehazing performance.

[*] adrian.galdran@inesctec.pt

Work financed by FCT - Portuguese Foundation for Science and Technology within project CMUP-ERI/TIC/0028/2014

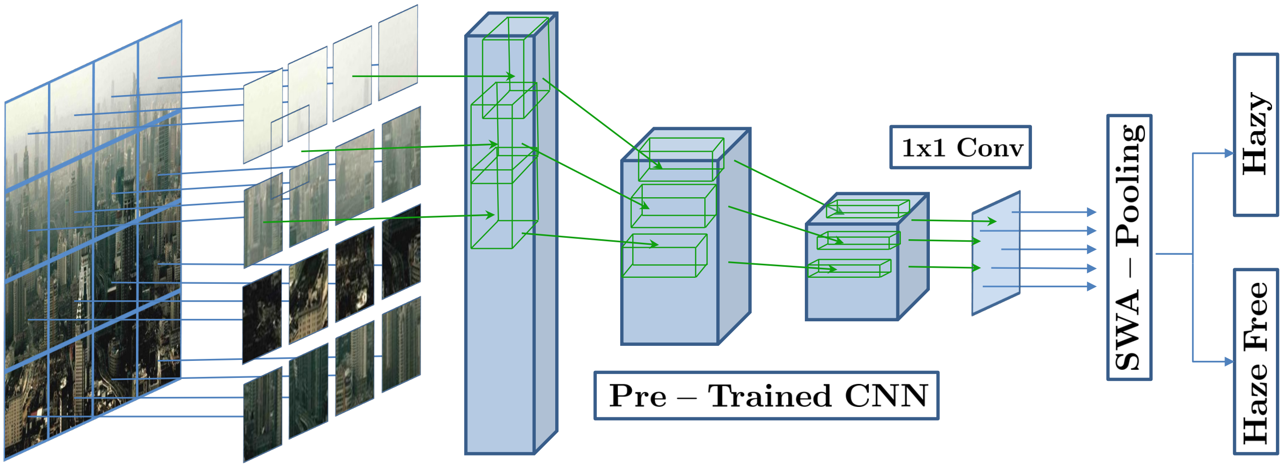

Fig. 2. The proposed model for weakly-supervised fog detection

2. LEARNING TO DETECT FOG FROM WEAKLY-LABELED DATA

2.1. Weakly-Supervised Learning and MIL

In the supervised learning context, the Multiple Instance Learning (MIL) framework consists of a relaxation on training data requirements in order to learn useful low-level information from weakly labeled data, *i.e.* to infer local patch-level information from global image-level labels [13]. In the MIL paradigm, each training example is a collection of instances referred to as bag. During training, the label of the bag is propagated to the label of the instances contained on it, and it has an influence on the classifier. The objective of training is to be capable of inferring information regarding the instances contained in a bag, based solely on the label of each bag.

In this paper, we propose to consider hazy and haze-free images as bags containing instances interpreted as image regions. The task we consider is to classify images into one of these two classes, fog/fog-free, while jointly inferring information related to which image regions are responsible for the image-level label. In this sense, our approach follows an extension of the Standard MIL Assumption [13]. This principle indicates that all the instances inside a negative bag are negative, while a positive bag contains at least one positive instance and an arbitrary number of negative instances. In our case, this can be interpreted as follows: *A haze-free image cannot contain foggy regions in any part of it, while a hazy image will contain some regions (but not all) showing fog.*

Accordingly, we propose a model to predict if an image is haze-free or hazy. To define such a model, we need to specify:

1. How do we represent instances within images?
2. How do we merge, or pool, instance-level predictions into a single image prediction?

We answer below each of these questions.

2.2. Instance-level Representation

Regarding 1), in this work we use a pre-trained Convolutional Neural Network to obtain a representation of each image region. This is the standard choice for feature extraction in current Computer Vision. In our case, we use the GoogleNet V3 architecture [14], initialized with pre-trained weights from ImageNet. However, we do not employ the full architecture end-to-end. Instead, we consider that a CNN fundamentally applies convolutions followed by pooling layers that down-sample an input image iteratively until it reaches a one-dimensional representation, that can be used for making predictions. In our case, we do not arrive until the end of this process, but rather run the forward pass of the CNN until the receptive field of the last layer reaches a pre-specified spatial resolution. Note also that we supply the model one image at a time, instead of splitting it into patches. This is inspired by Fully-Convolutional Networks (FCNs) for image segmentation [15]. Also, once we reach the end of the chopped network, rather than upsampling again the output of this layer as in an FCN, we simply apply a 1×1 convolution to the resulting activation map, followed by a sigmoid function to obtain a prediction on each image patch, as shown in Fig. 2. This enables the generation of predictions for a set of image patches independently. During training, patch labels are treated as a latent variables, since in our setting there is no patch-level information to learn from.

2.3. A Pooling Mechanism for Fog Detection

Once each patch p_I of an image I has an associated prediction $\mathcal{P}(p_I)$, the next step is to combine them into a single image-level prediction $\mathcal{P}(I)$. This is necessary in order to train with weak labels (hazy/haze-free image), and it can be achieved by a pooling operation . Several possible choices exist [16], each of them modeling different situations:

1. **Max-Pooling**, image-level predictions are obtained as the maximum response across every patch prediction:

$$\mathcal{P}(I) = \max_{i}(\mathcal{P}(p_I^i)),\ i \in \{1, ..., N\}. \tag{1}$$

This corresponds to the standard MIL assumption, and it implies that as long as a hazy patch appears in the scene, the entire image is considered as hazy.

2. **Average-Pooling** corresponds to an image considered as hazy if more than 50% of its patches contain fog:

$$\mathcal{P}(I) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{P}(p_I^i). \tag{2}$$

This may be better suited for our problem, but the fixed percentage of regions needed to consider an image as hazy remains a too rigid boundary decision.

3. **Shifted Average-Pooling** corresponds to stating that an image is predicted as hazy if more than $(100 \cdot k)\%$ of image patches contain fog:

$$\mathcal{P}(I) = k + \frac{1}{N}\sum_{i=1}^{N}\mathcal{P}(p_I^i). \tag{3}$$

Parameter k is learned, giving the model the flexibility to learn from data which is the appropriate proportion of patches in order to consider an image as foggy.

4. **Weighted Average-Pooling**, in this case we can assign a weight to each prediction:

$$\mathcal{P}(I) = \frac{1}{N}\sum_{i=1}^{N}w_i \cdot \mathcal{P}(p_I^i). \tag{4}$$

The weights w_i can be learned during training, and they model the fact that some regions within an image should have more relevance when making the decision of whether the image is hazy or not. This agrees with the idea that fog typically appears in the top part of the scene, and hardly in the bottom. As such, this pooling operation is expected to assign higher weights to regions coming from this part of the training images.

5. **Shifted Weighted Average-Pooling**, a combination of the previous two operations. This pooling type is meant to model the fact that some regions are more representative than others to detect fog, but also that an image should be declared as hazy if more than $k\%$ of image patches contain fog:

$$\mathcal{P}(I) = k + \frac{1}{N}\sum_{i=1}^{N}w_i \cdot \mathcal{P}(p_I^i). \tag{5}$$

In this case, both w_i and k are learned from the data.

According to the above discussion, in this paper we select Shifted Weighted Average-Pooling as the appropriate operation to combine decisions made by the classifier at patch level.

2.4. Model Optimization

The model is trained by standard back-propagation, after resizing every image to 512×512 pixels. We chop the pretrained CNN in such a way that the last layer classifies patches of size 114×114, with an overlap of 106 pixels. This allows the model to see overlapped 61×61 patches from the same image during training. The loss function driving the optimization is binary cross-entropy, since the goal is to classify images into two different class. The weights of the pretrained CNN were fixed, and the Adam algorithm [17] was used with a learning rate of $2e - 4$ to fine-tune the last layer for 30 epochs. Finally, we fine-tuned the entire model using early stopping with a patience of 15 epochs, until the loss in a separate validation set did not decrease anymore. We built the attention maps shown in the paper by upscaling the patch predictions to the same size of the image, taking into account the overlap between patches: values in overlap areas are weighted averages, with weights coming from Gaussians centered in the patches.

3. EXPERIMENTAL RESULTS

In order to train our model, we employed the two sets of 500 real-world hazy and haze-free images provided in [18]. We supplemented this set of images with the NYU subset of the D-Hazy dataset [19], which contains 1449 images on which synthetic fog has been added thanks to the availability of a depth map of the scene.

For testing our approach, an independent dataset was built from a completely different source to ensure no overfitting was occurring. We randomly selected a subset of 130 haze-free images from the Places dataset [20], and sampled also randomly a sub-set of similar size from the unannotated hazy images sub-dataset in the RESIDE benchmark [21]. To provide comparison against other state-of-the-art image dehazing that return an estimate of the fog density in the scene, we consider the popular Dark Channel (DC) method [5], as well as the method introduced in [18], which comprises both a metric of fog density (FADE) and a complementary fog removal technique based on it (DEFADE). For the Dark Channel technique, we considered mean intensity value of the estimated inverse transmission map as an indicator of the amount of haze in an image. This consists of a value in $[0, 1]$, being higher when more foggy regions are detected in the image.

3.1. Qualitative Results

The proposed method learns to predict the presence of fog from image-level ground-truth. Nevertheless, it can point towards the regions within each image that led to each prediction by means of attention maps. This has the advantage that those regions are learn from the data, avoiding any manual specification of the aspect of fog. In Fig. 3 we show some examples in foggy and fog-free images of such attention maps.

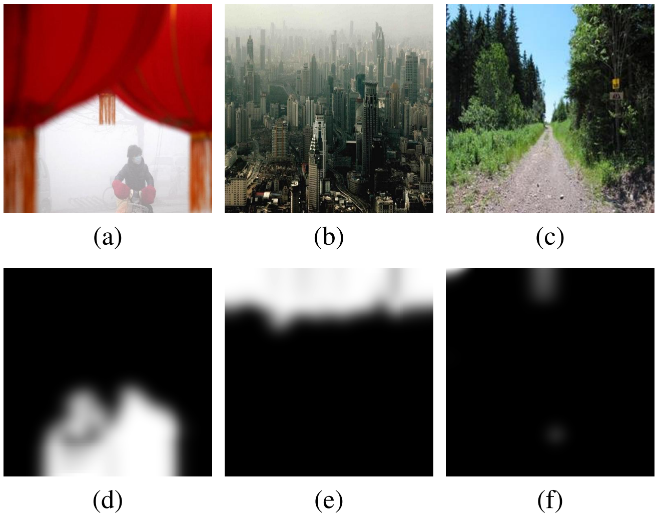

Fig. 3. Top: (a), (b): Foggy images. (c): Fog-free image. Bottom: Corresponding attention maps explaining the regions within each image that triggered the decision of the model.

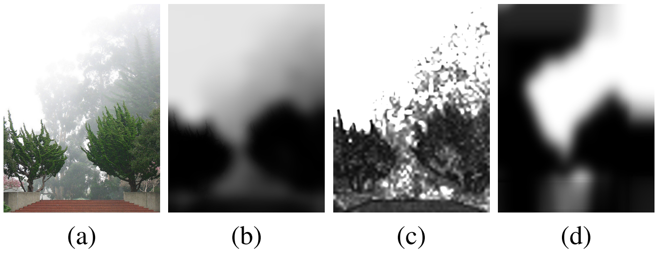

Fig. 4. (a) Foggy image, (b) depth map estimated by the Dark Channel method, (c) fog density map estimated by FADE, (d) attention map generated by the proposed method.

Even if these maps do not contain the full-details of the scene, they are enough to understand the behavior of the model.

To further verify the consistency of these attention maps, we show in Fig. 4 a hazy image, the depth map predicted by the Dark Channel (which is used to guide the enhancement of the image, far-away pixels receive increased contrast), and the fog density map returned by FADE. It can be seen that both methods interpret the sky region as containing fog, due to its bright appearance. This may potentially lead to typical artifacts when dehazing images, since current methods try to extract contrast from a region that has no underlying objects. On the contrary, Fig (4d) shows how the proposed method can learn from the training data that bright regions on the top of the scene with no underlying texture can simply be part of the sky, and not indicative of the presence of fog. In this case, the image is declared as hazy by the method due to the presence of fog on top of the trees in the right-most top region.

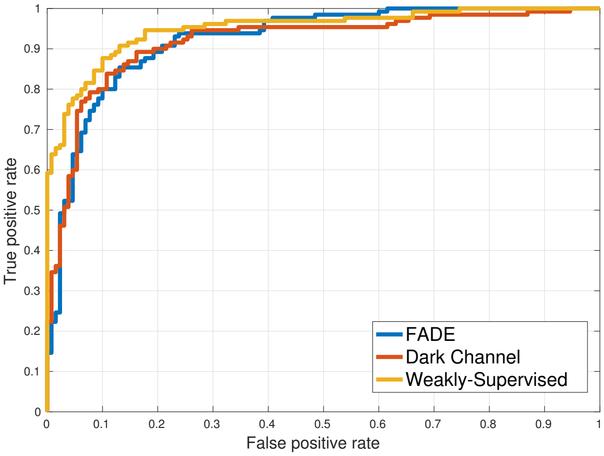

Fig. 5. ROC curve for each considered method

	FADE	Dark Channel	Proposed
Accuracy	0.8615	0.8577	**0.8846**
AUC	0.9183	0.9226	**0.9502**

Table 1. Fog Detection Performance Comparison between our proposed Weakly-Supervised method and FADE, DC.

3.2. Quantitative Results

Using the fog-predictive score derived from the Dark Channel and the FADE metric, we can compare the performance in terms of fog detection of the proposed method against these two techniques. For this, we estimated the probability of an image being considered haze-free/hazy in the independent test set explained above. The Receiver Operating Characteristic (ROC) curve resulting from each method is shown in Fig. 5. We also derived from them standard performance metrics: accuracy and Area under the ROC curve (AUC). Performance for each of the considered approaches is shown in Table 1.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a weakly-supervised method for fog detection that learns to predict haze presence using only image-level labels. The proposed model can detect fog with great accuracy, and it produces attention maps indicating the regions within an image that triggered its decision.

Predicting if an image contains fog is can be useful for improving the adjustment of relevant parameters from existing image dehazing techniques, as well as to compare among them. In the future, the attention maps produced by the method will be optimized to capture more scene details, enabling the development of a no-reference image dehazing quality metric, which will be further extended to haze-like artifacts and degradations typical of eye fundus images.

5. REFERENCES

[1] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, vol. 1, pp. I–325–I–332 vol.1.

[2] Raanan Fattal, "Single Image Dehazing," in *ACM SIGGRAPH 2008 Papers*, New York, NY, USA, 2008, SIGGRAPH '08, pp. 72:1–72:9, ACM.

[3] Codruta Orniana Ancuti and Cosmin Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.

[4] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmo, "Enhanced Variational Image Dehazing," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1519–1546, Jan. 2015.

[5] K. He, J. Sun, and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[6] D. Berman, T. Treibitz, and S. Avidan, "Non-local Image Dehazing," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1674–1682.

[7] Raanan Fattal, "Dehazing Using Color-Lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 13:1–13:14, Dec. 2014.

[8] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmo, "Fusion-Based Variational Image Dehazing," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 151–155, Feb. 2017.

[9] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[10] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang, "Single Image Dehazing via Multi-scale Convolutional Neural Networks," in *Computer Vision ECCV 2016*. Oct. 2016, Lecture Notes in Computer Science, pp. 154–169, Springer, Cham.

[11] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-One Dehazing Network," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 4780–4788.

[12] K. Ma, W. Liu, and Z. Wang, "Perceptual evaluation of single image dehazing algorithms," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept. 2015, pp. 3600–3604.

[13] Jaume Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, Aug. 2013.

[14] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv:1512.00567 [cs]*, Dec. 2015, arXiv: 1512.00567.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[16] P. Costa, A. Campilho, B. Hooi, A. Smailagic, K. Kitani, S. Liu, C. Faloutsos, and A. Galdran, "EyeQual: Accurate, Explainable, Retinal Image Quality Assessment," in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2017, pp. 323–330.

[17] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Dec. 2014, arXiv: 1412.6980.

[18] L. K. Choi, J. You, and A. C. Bovik, "Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[19] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept. 2016, pp. 2226–2230.

[20] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million Image Database for Scene Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2018.

[21] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang, "RESIDE: A Benchmark for Single Image Dehazing," *arXiv:1712.04143 [cs]*, Dec. 2017, arXiv: 1712.04143.