Deep Learning Model for A Driver Assistance System to Increase Visibility on A Foggy Road

Samir Allach*, Mohamed Ben Ahmed, Anouar Abdelhakim Boudhir

LIST Laboratory, FSTT, UAE University, Tangier 90000, Morocco

A B S T R A C T

For many years, a lot of researches have been made to develop Advanced Driver Assistance Systems (ADAS) that are based on integrated systems. The main objective is to help drivers. Hence, keeping them safe under different driving conditions. Visibility for drivers remains the biggest problem faced on the road in an atmosphere of fog. In this paper, we examine a system that can be employed to substantially enhance visibility through using deep neural networks. Researches done recently- which are based on deep learning for eliminating image fog- have made clear that an end-to-end proposed system is such an effective model. However, it becomes a must to extend the idea to end-to-end real-time video deshazing. In this paper, we introduce a model of image dehazing. It is based on Convolutional Neural Networks (CNN) as a basis for developing the video dehazing model. As in addition, we concatenate our model with the faster RCNN for detecting objects on the road in real time. The experimental results on our image datasets shows the performance of our model with regard to Peak Signal to Noise Ratio (PSNR=19.823) and Structural Similarity (SSIM =0.8501). On the dataset of the synthesized videos, our model achieved a performance of PSNR = 21.4032 and SSIM = 0.9354. Moreover, with the concatenation of our dehazing model with Faster R-CNN (regions with convolutional neural networks), our proposed system displays desirable visual quality and a remarkable progress of the object detection achievement on blurred images with mean Average Precision (mAP) equal to 0.933 during the day and 0.804 at night.

1.   Introduction

Advanced Driver Assistance Systems (ADAS) are mainly designed to offer vehicle drivers help, thereby minimizing a potential threat to their safety. The majority of these systems are based on image processing algorithms, such as those allowing detecting of nearby vehicles and pedestrians and the recognition of signs. Because of this, quite a lot of systems are fixed into vehicles. Though such systems are widely employed to draw drivers' attention once a potential threat appears for vehicle drivers, they perform less effectively under certain adverse weather conditions where the vision weakened. This happens most notably in the presence of fog.

Eliminating or reducing the fog of an image captured by an ADAS system seems to be difficult and somehow ill-posed phenomenon. It is worth mentioning here that significant developments that took place in deep learning mainstream paved the way to get considerable results in improving vision degraded

by fog [1]. Therefore, the elimination of fog requires the estimation of the depth map. What is more, previous assumptions are necessary to estimate the depth map for systems using single images as input.

Very lately, a lot of algorithms have been suggested for detecting objects and eliminating fog [1, 2]. It is worth mentioning here that traditional algorithms need two crucial elements: gathered facts on the environment and developed learning as well. In addition, most object detection and fog elimination algorithms are not suited for real-time uses on account of they consist of considerable arithmetic time.

In this way, authors in this area of study have introduced deep learning approaches for degraded vision so that images can be restored and reconstructed [1]-[3]. Yet, these methods cannot directly be applied for removing the fog from the image. The authors F. Hussain, and J. Jeong, came up with an approach, using Deep Neural Network (DNN).They also assumed that an anonymous complex function could model mathematically the fog in an image [4]. Li Chongyi et al. proposed a cascading CNN

*Corresponding Author: Samir Allach, LIST-FSTT Bp 416 Tangier-Morocco,
Tel: +212615571400 & Email: allach.samir@gmail.com

model composed of three components: a component of the shared invisible layers that extract the common features, the work of the global estimate of the atmospheric light and that of the subnet of the average transmission [5]. The researchers Li Boyi and others introduced a model for dehaze of image constructed with a convolutional neurons network (CNN), named "all-in-one dehazing network" and "An All-in-One Network for Dehazing and beyond" (AOD-Net) [1, 6]. It is designed on the basis of a reformulated atmospheric diffusion model. The researchers B.Cai et al. found an end-to-end formable model an excellent alternative for calculating the average transmission, called DehazeNet [2], it takes a hazy image at the input and leaves its transmission matrix, and then image dehazed is recovered by the atmospheric diffusion model. W. Ren et al. used a multi-scale CNN (MSCNN) that created and improved a coarse scale transmission matrix [7].

With regard to the video dehazing, most approaches depend crucially on the phase of improving temporal inconsistencies. The authors Kim et al. suggested inserting a temporal coherence in the cost function, with a clock filter to accelerate the processing [8]. The authors Cai et al. conceived a spatio-temporal optimization for de-hazing video in real time [2].

In recent years, an interest has begun to grow in video modeling using CNN for a huge number of tasks. And here are three examples: super-resolution (SR) [9], blur [10] and classification [11, 12] and style transfer [13]. In [9], the author studied various structure configurations for SR video. In [11, 12], the authors made similar attempts by exploring different connectivity options for video classification. The researchers Liu et al. introduced such adjustable system. In this context, they placed a spatial alignment network between the images [14]. Others introduced an end-to-end CNN to learn how to accumulate information on several hazy images / video [10]. With reference to video style transfer, Chen et al. integrated short- and long-term coherence. This is not to mention they indicated the superiority of multi-image approaches over single-image approaches [13].

In this paper, we present a video dehazing system based on an image-dehazing model. We use convolutional neural networks (CNN). Then, we concatenate our model with the faster RCNN to detect objects on the foggy road in real time. Moreover, we deploy our system to validate the quantitative and visual results obtained.

2. Network Design

Our proposed work is made of two important steps. Firstly, we worked hard to develop an end-to-end CNN model [15] that explicitly learns the mapping relationships between the raw images and their associated transmission maps to recover fog-free images. And then we integrated the Faster R-CNN proposed model [16] for objects detection in video. The algorithm presents the general steps of our model for dehazing and detecting objects in a video:

2.1. Atmospheric diffusion model

The model of atmospheric diffusion is the usual description of the hazy image production process, proposed by McCartney [17] and developed later by Nayar and Narasimhan [18]. The model is written as follows [1,2,19]:

$$I(x) = J(x)\,t(x) + A(1 - t(x)) \tag{1}$$

Algorithm: Video dehazing and object detection

Result: video dehazing and object detection

Let's i : counter;

n : number of images after splitting video to images;

V_{in} : input video , M: Resulting of dehazed image( );

Dehz( ): Dehaze Function;

Detect( ): Object Detection Function;

Save( ): Saving resulting image;

Concat( ): Image concatenation;

$V_{split}( )$: Splitting video to images;

$V_{out}( )$: Video generating function;

Image1( ): Output image of Vsplit( );

Image2( ): Output image of Detect( );

begin

   Insert V_{in} ;

   Splitting Video to n images;

      Vsplit(V_{in});

   Dehazing images and detecting objects using Faster R-CNN;

      $i \leftarrow 0$;

      while *(i<=n)*

          M $\leftarrow$ Dehz(Image1 (i)) ;

          Image2 (i) $\leftarrow$ Detect(M) ;

          Save(Image2(i)) ;

          if *i > 0* **then**

             Concat(Image2(i) , Image2(i - 1)) ;

          end

        $i \leftarrow i + 1$;

      end

   V_{out}(image2(i);

end

$$t(x) = e^{-\beta d(x)} \tag{2}$$

where I (x) is an observed hazy image, J (x) is the image that recovered after the estimation of A and t(x), A is the global atmospheric light and t (x) is the transmission map.

In equation (2), β is the atmosphere diffusion coefficient and d (x) is used to refer to the distance between the image and the camera.

Equation (2) indicates that when d (x) moves to infinity, t (x) comes near to zero. With equation (1), we have:

$$A = I(x), d(x) \rightarrow inf \tag{3}$$

In reality, d (x) cannot be infinite, but can be a long distance, which gives a very weak transmission t_0. Then, the global atmospheric light (A) is estimated by the following formula

$$A = \max_{y \in \{x | t(x) \le t_0\}} I(y) \tag{4}$$

2.2. Construction of the proposed CNN model

The proposed model for the image frame consists of an estimation module. This module uses convolutional layers to

estimate transmission map t (x) (see Figure 1), followed by a clean image generation module that consists of a multiplication layer of several element by element and the addition layers for generating the recovery image (see Figure 2).

The estimation module is the essential component, responsible for estimating the depth and level of relative disorder. As Figure 1 shows, we use five convolutional layers by merging filters of various sizes (see Table 1).

After comparing the results obtained by different architectures of CNN (see table 1), we concluded that the CNN model that consists of five convolutional layers with the number of 3x3 filters is the most efficient, with a PSNR of 19.8231 and a SSIM 0.8501.

The authors B. Cai et al. used parallel convolutions with filter sizes varied [2]. W. Ren et al. concatenated the distinctive features of the large- scale network with an intermediate layer of the fine-scale network [7]. Influenced by them, we concatenated in our model the layer "pool1" entities with the layers "conv1" and "conv2". The same thing for "pool2" with those of "conv2" and "conv3"; and "pool3" with those of "conv1", "conv2", "conv3" and "conv4". This model captures the characteristics of images at different scales and at the intermediate connections, compensating for the loss of information during convolutions. Notably, in our proposed model each convolutional layer uses only three filters. Consequently, our model is lighter and performs well in terms of PSNR and SSIM values in comparison with other existing deep methods.
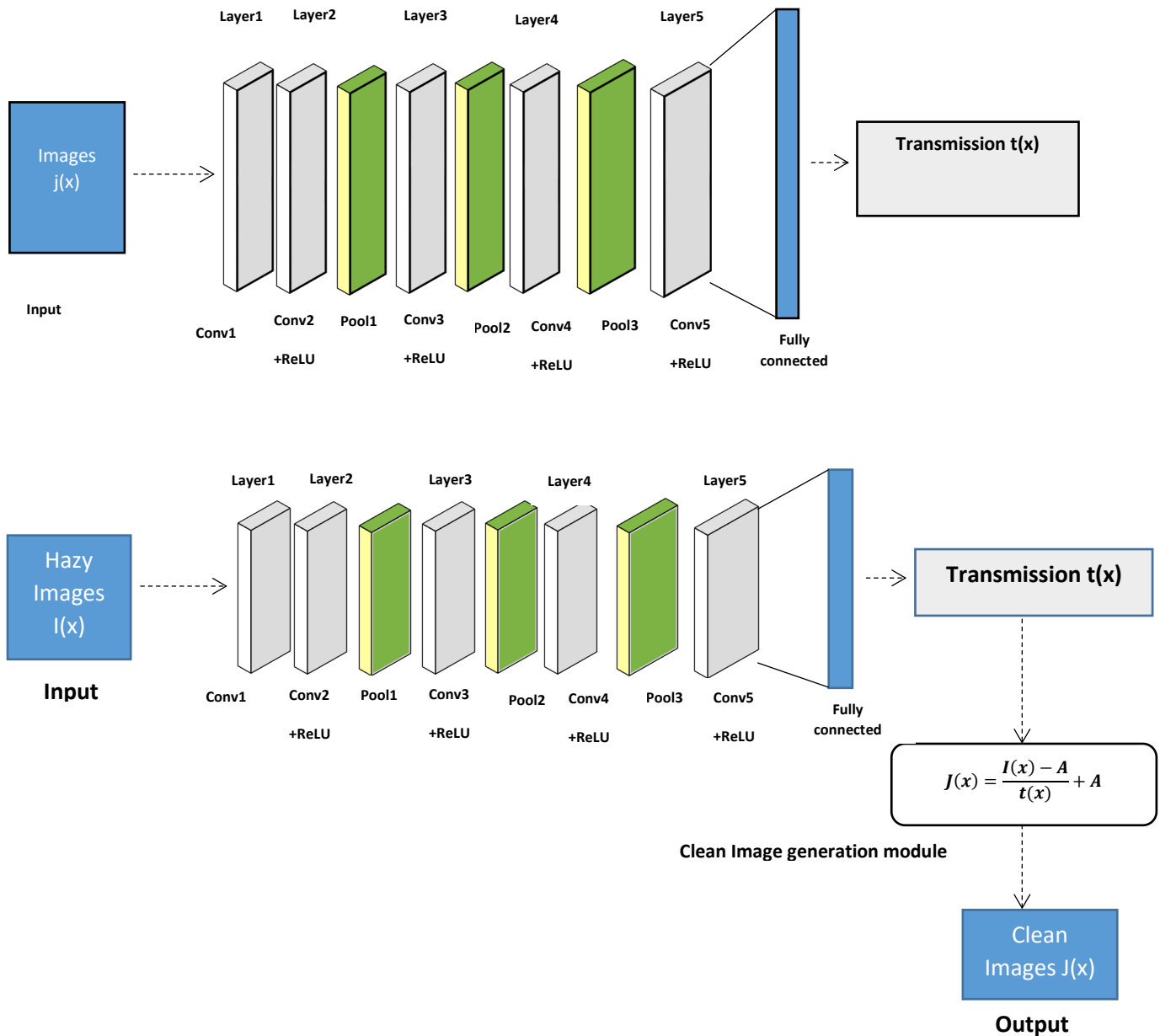

$$J(x) = \frac{I(x) - A}{t(x)} + A$$

Table 1: Average of PSNR and SSIM results for different architectures of CNN model

Number of convolution layers	Number of filters	PSNR	SSIM
2	3 x 3	-	-
	5 x 5	14.0321	0.6801
	7 x 7	14.1321	0.7032
3	3 x 3	15.6706	0.7791
	5 x 5	16.7653	0.7356
	7 x 7	16.0362	0.7452
4	3 x 3	16.8521	0.7937
	5 x 5	17.7653	0.7745
	7 x 7	18.2381	0.8012
5	**3 x 3**	**19.8231**	**0.8501**
	5 x 5	19.6364	0.8125
	7 x 7	19.5364	0.8029
6	3 x 3	18.5381	0.8101
	5 x 5	18.2381	0.7998
	7 x 7	18.3381	0.8111

2.3. Model of Object detection

Faster R-CNN [16] is proposed to detect objects on images accurately. With ResNet, Faster R-CNN achieved a mean average Precision (mAP) of 76.4% on the PASCAL VOC dataset 2007 and 2012 (see Table 2). By combining region and classifier propositions in a large network, it becomes possible to automatically learn good representations of the features for the task.

In this paper, we examine the detection and recognition of objects in the presence of fog with a view to improve high-level vision tasks to combine with the model for dehazing of video. We opted for the Faster R-CNN model as the basic algorithm for detecting robust objects in real time (see Table 2), approved on synthetic and natural blurred images.

We modified the Faster R-CNN Caffe source code and used ResNet [20] as the Faster-RCNN convolution backend. We came to the conclusion that using ResNet provides a substantial improvement over other architectures. The time reweighting of the chain gives further improvements.
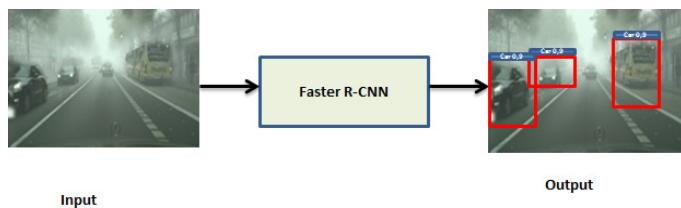

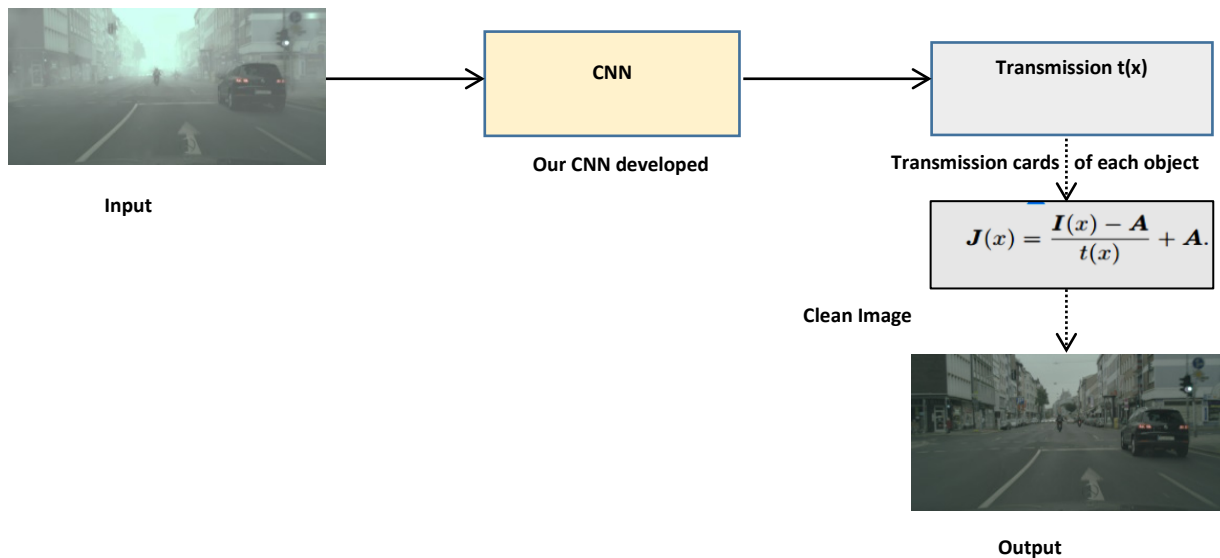Figure 3: Object detection with Faster R-CNN model

Table 2: Comparison of detection methods on the 2007 and 2012 PASCAL VOC dataset

Method	mAP	FPS	Batch size
Fast R-CNN [21]	0.70	0.5	1
YOLO [20]	0.634	45	1
SSD300 [21]	0.743	46	8
Faster R-CNN VGG-16 [21]	0.732	7	1
Faster R-CNN-ResNet [20]	0.764	5	1

Table 2 shows comparison measurements of mAP and the FPS (frames per second) speed with the batch size, using the dataset PASCAL VOC between Fast R-CNN, YOLO, SSD300, Faster R-CNN VGG-16 and R- Faster CNN-ResNetet. Our chosen detection method (Faster R-CNN-ResNetet) surpasses all methods in terms of mAP. Though a set of certain methods may work at higher speeds, they have lower accuracies (mAP). The Faster R-CNN-ResNetet remains the best method in real time to reach more than 76.4% of mAP.

Outside the dehazing part, the temporal coherence should be taken into consideration for object detection so that we can reach satisfactory results. With our proposed model, we can hereby-advancing Faster R- model CNN ideally suited for video [22]. The first two convolutional layers of Faster R-CNN model to an image, was divided into three equivalent sections to insert the earlier, present and upcoming images, in this order. The last three images are concatenated in the next of the second convolutional layer and go through the resting layers for predicting the bounding boxes of the objects for the present image.

3.   Results and discussion

3.1. Dehazing an image

First, we created foggy images synthesized from equations (1) and (2), using the ground truth images with the NYU2 inner depth database.

We put different atmospheric lights (A), selecting them consistently each channel between [0.5, 1.0] and selecting $\beta \in \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$.

The input data is RGB images with a resolution of 640 x 480 and the output was a depth map with a resolution of 320 x 240.

In our model, about 12 epoch are enough for it to converge, and it works well enough after these times. During the learning phase, the weights of our network are initialized in a random way. We used the ReLU function as a more efficient neuron in our specific context. We opted for the function of Mean Square Error loss (MSE), and that it stimulates the PSNR (Peak Signal to Noise Ratio) and the SSIM (Structural Similarity) as well as the visual quality.

Figure 4: Our model of image dehazing

Table 3: Average PSNR and SSIM results on Test1

	ATM	BCCR	NLD	FVR	DCP	DehazeNet	CAP	MSCNN	AOD-Net	**OUR MODEL**
PSNR	14.1475	15.7606	16.7653	16.0362	18.5385	18.9613	19.6364	19.1116	19.6954	**19.8231**
SSIM	0.7141	0.7711	0.7356	0.7452	0.8337	0.7753	0.8374	0.8295	0.8478	**0.8501**

We compared our basic model proposed with several dehazing methods, in particular: Automatic atmospheric light Recovery (ATM) [24], Regularization of the context constrained to limits (BCCR) [25], No-local Image Dehazing (NLD) [26], Fast visibility restoration (FVR) [27], Dark channel priority (DCP) [28], MSCNN [7], DehazeNet [2], Color Attenuation Prior (CAP) [29] and AOD-Net [1]].

Our synthesized hazy images come along with ground truth images, paving the way for us assess PSNR and SSIM and check out if the results stay accurate.

As our model is optimized from start to finish in case of MSE loss, one should not be amazed to see its PSNR performance superior to other methods. Further, even if SSIM is not directly called the optimization criteria, our model obtains SSIM advantages superior to the other models compared.

It is well known that SSIM more accurately reflects human perception, as SSIM measures apart from pixel level errors. We become faithful to our model with such a consistent SSIM improvements we achieved.

Table 3 shows the promising performance of our model compared to the others, in terms of PSNR and SSIM.

Our method has an advantage greater than 0.2 dB in the PSNR and 0.03 in the SSIM when we compare it to approach AOD-Net.

3.2. Video sequences dehazing

First of all, we brought into existence a dataset of foggy synthetic video from on equation(1), employing 20 chosen videos in the TUM RGB-D dataset [30] that consists of different video sequences. Depth information is refined by means of the filling model of Silber man et al. [31]. After that, we broke our dataset into a learning set made up of 12 videos with 120,000 images and a set of non-overlapping tests called Test2, consisting of 8 short video sequences containing a total of 450 images. Finally, we collected a set of natural hazy video sequences to validate and evaluate the performance of our model.

During the training of our model, we adopted the loss of the mean squared error (MSE), which is aligned with the SSIM, the PSNR and the visual quality.

Table 4: PSNR average and SSIM results on Test2.

	STMRF	EVD-Net	OUR MODEL
PSNR	18.9956	20.9908	21.4032
SSIM	0.8707	0.9087	0.9354

Concerning video-based methods, we compared our model with EVD-Net [19] and STMRF [32] methods (not based on CNN). Then, we got a higher advantage over the two previous approaches. Furthermore, we observed that our advanced system performs quite remarkably showing a difference of 0.5 dB in the PSNR and 0.03 in the SSIM.

3.3. Object detection model

We reformed our adaptive Faster-RCNN model on a set of training data provided by Foggy Cityscapes [33] and on a set of synthesized personal data. Foggy Cityscapes is a foggy synthetic data set. It causes fog to operate on real scenes. The images are rendered utilizing the Cityscapes images and depth maps [8]. It contains 2,975 images in the learning set and 500 images in the validation set. In this experiment, we reported our findings on categories such as person, car, truck, bus, motorcycle and bicycle. We opted for the average score of the channel separately for each class based on validation performance.

On the validation set, we achieve a mAP of 0.777 (see Table 5 for a full breakdown between classes).

Table 5: Average Precision (AP) and Mean Average Precision (mAP) between original Faster R-CNN and adaptive Faster R-CNN with ResNet.

Metrics	Original Faster R-CNN	Adaptive Faster R-CNN with ResNet
Person (AP)	0.731	0.779
Car (AP)	0.809	0.812
Truck (AP)	0.750	0.803
Bus (AP)	0.692	0.792
Motorbike (AP)	0.663	0.723
Bike (AP)	0.621	0.756
mAP	0.711	0.777

3.4. The proposed model Architecture for object detection in hazy video

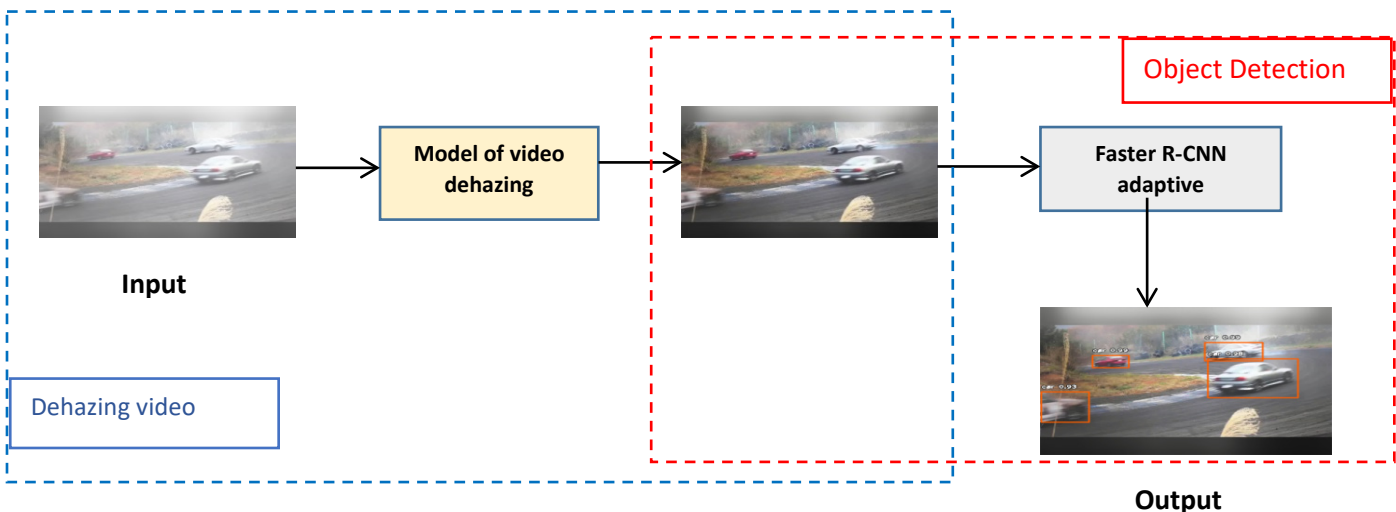The merging of our model for video dehazing and the adapted Faster R-CNN model (Figure 5)has given birth to our general model, which naturally displays an interesting tree structure that is locally linked and is subjected to joint and more crucial optimization.

On the validation set of our proposed model, we achieved a mAP of 0.929 (see Table 6).

Table 6: Average Precision (AP) and mean Average Precision (mAP) results with adaptive Faster R-CNN before and after the dehazing operation.

Metrics	Adaptive Faster R-CNN with ResNet	
	Before the dehazing	After the dehazing
Person (AP)	0.779	0.902
Car (AP)	0.812	0.911
Truck (AP)	0.803	0.909
Bus (AP)	0.792	0.970
Motorbike (AP)	0.723	0.981
Bike (AP)	0.756	0.901
mAP	0.777	0.929

4. Deployment of our system

As an embedded platform, we used a Raspberry Pi 3 Model B +, for our experiments we made predictions on new video sequences using the raspberry Pi with camera designated v2.1, based on the Sony IMX219 CMOS type sensor with a resolution of 8 MP (3280 × 2464 pixels) and we downloaded our pre-trained model on our pi. We needed Raspbian Stretch 9, because TensorFlow 1.9 officially supports Raspberry Pi if you use Raspbian 9.
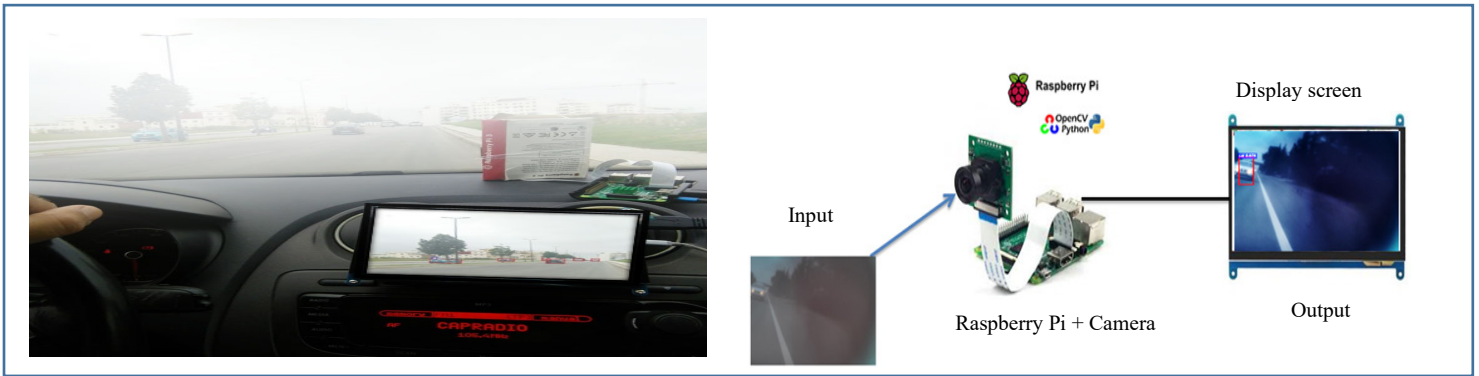

Figure 5: The proposed general model

Figure 6: structure of system deployment

Table 7: Results of mAP for 5 states: Heavy haze (day), Medium haze (day), Light haze (day), Medium haze (night), Light haze (night), before and after dehazing.

	During the day			At night	
	Heavy haze	Medium haze	light haze	Medium haze	light haze
Faster R-CNN (adaptive)	mAP = 0.501	mAP = 0.663	mAP = 0.815	mAP = 0.531	mAP = 0.678
Dehazing + Faster R-CNN (adaptive)	mAP = 0.773	mAP = 0.891	mAP = 0.933	mAP = 0.723	mAP = 0.804

Table 8: the visual results of our model on images taken by the raspberry camera

	During the day			At night	
	Heavy haze	Medium haze	Light haze	Medium haze	Light haze
Hazy image					
Objet detection in hazy images					
Objet detection in Image dehazing					

We performed tests on images taken during the day and at night using the Faster R-CNN model adaptive at first, and then the proposed general model (image dehazing + Faster R-CNN adaptive). Table 7 and 8 show the results obtained.

4.1. The Quantitative results

For quantitative results, we calculated the mean Average Precision (mAP) across all images, using both the adaptive Faster R-CNN model and the dehazing model concatenated with the adaptive Faster R-CNN. (See Table 7). During the day, the heavy haze degrades mAP to around 0.27. Adding the dehazing model, mAP is improved by 0.12 for the detection of objects under a light veil conditions, 0.23 in medium fog and 0.27 in thick fog. During the night, with the dehazing model the mAP improves by 0.28 for detecting objects in light haze conditions and 0.21 in the middle fog (see table 7).

4.2. The results visualized

For the visualized results, Table 8 makes a visual comparison of the results when the object is detected on hazy images. And here we are illustrating five cases:

- **During the day:**

  - Detection of objects in the Heavy Haze image with adaptive Faster R-CNN before and after dehazing.

  - Detection of objects in the Medium haze image with adaptive Faster R-CNN before and after dehazing.

  - Detection of objects in the Light haze image with adaptive Faster R-CNN before and after dehazing.

- **At night:**

  - Detection of objects in the Medium haze image with adaptive Faster R-CNN before and after dehazing.

  - Detection of objects in the Light haze image with adaptive Faster R-CNN before and after dehazing.

Experiments we conducted lead us to this conclusion: as soon as the haze becomes heavier at night, detecting objects becomes less dependable. Most of all, under whichever fog conditions- light or medium or high- our highly advanced system can regularly ameliorate the process of detecting objects.

5. Conclusion

In this paper, we present an intelligent system with a view to enhance visibility quality in atmospheric fog, allowing dehaze and object detection. The proposed system is based on the deep learning model using a light CNN model of five convolutional layers works well in terms of PSNR and SSIM compared to other existing deep methods.

We modified the Faster R-CNN model, using ResNet as its convolution backend. We concluded that using ResNet offers a substantial improvement over other architectures. Further, we concatenated the CNN video dehazing model with the faster RCNN as a robust model compared to the others. Our objective was detecting objects on the road in real time with significant performance in terms of mAP.

Based on the qualitative and visual results, our system demonstrated both efficiency and superiority over other existing systems in terms of PSNR, SSIM, mAP and visual quality.

It should be noted that, though the proposed system was much more applicable in enhancing visibility for road drivers when there is fog, there were some limitations. First of all, we did not test our proposed system in case of a rainy foggy environment. Secondly, our system is not applicable in case of very dense fog. Lastly, it is not also applicable when detecting so many objects. Consequently, in our future work we shall definitely spare no effort to enhance the performance of our system by an automated Chabot. This Chabot interprets the captured image and announces audio warnings in the event of critical situations, a project that mainly aims at developing a system in the situation where the presence of fog and rain come at the same time.

References

[1] B.Li, X.Peng, Z.Wang, J.Xu, D. Feng, "Aod-net: All-in-one dehazing network." In Proceedings of the IEEE International Conference on Computer Vision, 4770-4778, 2017. DOI: 10.1109/ICCV.2017.511

[2] B.Cai, X.Xu, K.Jia, C.Qing, D. Tao, "Dehazenet: An end-to-end system for single image haze removal." IEEE Transactions on Image Processing, 25(11), 5187-5198, 2016. DOI: 10.1109/TIP.2016.2598681

[3] C.Dong, C. C.Loy, K.He, X.Tang, "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307, 2015. DOI: 10.1109/TPAMI.2015.2439281

[4] F.Hussain, J. Jeong, "Visibility enhancement of scene images degraded by foggy weather conditions with deep neural networks." Journal of Sensors, 2016. https://doi.org/10.1155/2016/3894832.

[5] C.Li, J.Guo, F.Porikli, H.Fu, Y.Pang, "A cascaded convolutional neural network for single image dehazing." IEEE Access, 6, 24877-24887,2018. DOI: 10.1109/ACCESS.2018.2818882

[6] B.Li, X.Peng, Z.Wang, J.Xu, D.Feng, "An all-in-one network for dehazing and beyond." arXiv preprint, 2017. https://arxiv.org/abs/1707.06543

[7] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. H. "Single image dehazing via multi-scale convolutional neural networks." In 2016Springer/Cham, European conference on computer vision, 154-169,2016. https://doi.org/10.1007/978-3-319-46475-6_10

[8] J. H.Kim, W. D.Jang, J. Y.Sim, C. S. Kim, "Optimized contrast enhancement for real-time image and video dehazing." Journal of Visual Communication and Image Representation, 24(3), 410-425, 2013. http://dx.doi.org/10.1016/j.jvcir.2013.02.004

[9] A.Kappeler, S.Yoo, Q.Dai, A. K. Katsaggelos, "Video super-resolution with convolutional neural networks." IEEE Transactions on Computational Imaging, 2(2), 109-122, 2016. DOI: 10.1109/TCI.2016.2532323

[10] S.Su, M.Delbracio, J.Wang, G.Sapiro, W.Heidrich, O.Wang, "Deep video deblurring for hand-held cameras." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1279-1288, 2017. DOI: 10.1109/CVPR.2017.33

[11] A.Karpathy, G.Toderici, S.Shetty, T.Leung, R.Sukthankar, L. Fei-Fei, "Large-scale video classification with convolutional neural networks." In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 1725-1732, 2014. https://doi.org/10.1109/CVPR.2014.223

[12] H.Shen, S.Han, M.Philipose, A.Krishnamurthy, "Fast video classification via adaptive cascading of deep models." In Proceedings of the IEEE conference on computer vision and pattern recognition, 3646-3654, 2017. DOI: 10.1109/CVPR.2017.236

[13] D.Chen, J.Liao, L.Yuan, N.Yu, G.Hua, "Coherent online video style transfer." In Proceedings of the IEEE International Conference on Computer Vision, 1105-1114, 2017. DOI: 10.1109/ICCV.2017.126

[14] D.Liu, Z.Wang, Y.Fan, X.Liu, Z.Wang, S.Chang, T. Huang, "Robust video super-resolution with learned temporal dynamics." In Proceedings of the IEEE International Conference on Computer Vision, 2507-2515, 2017. DOI: 10.1109/ICCV.2017.274

[15] S.Albawi, T. A.Mohammed, S.Al-Zawi, "Understanding of a convolutional neural network." In 2017IEEE International Conference on Engineering and Technology (ICET), 1-6, 2017. DOI: 10.1109/ICEngTechnol.2017.8308186

[16] Y.Xu, G.Yu, Y.Wang, X.Wu, Y. Ma, "Car detection from low-altitude UAV imagery with the faster R-CNN." Journal of Advanced Transportation, 2017. https://doi.org/10.1155/2017/2823617

[17] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles." New York, John Wiley and Sons, Inc., 421, 1976. https://doi.org/10.1063/1.3037551

[18] S. G.Narasimhan, S. K. Nayar, "Chromatic framework for vision in bad weather." In 2000 Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, 598-605, 2000. DOI: 10.1109/CVPR.2000.855874

[19] B.Li, X.Peng, Z. Wang, J.Xu, D. Feng, "End-to-end united video dehazing and detection." In 2018 AAAI Thirty-Second Conference on Artificial Intelligence, 2018. https://arxiv.org/abs/1709.03919

[20] K.He, X.Zhang, S.Ren, J. Sun, "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778, 2016. DOI: 10.1109/CVPR.2016.90

[21] W.Liu, D.Anguelov, D.Erhan, C.Szegedy, S.Reed, C. Y.Fu, A. C. Berg, "Ssd: Single shot multibox detector." In 2016 Springer/Cham, European conference on computer vision, 21-37, 2016. https://doi.org/10.1007/978-3-319-46448-0_2

[22] S.Ren, K.He, R.Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems, 91-99, 2015. DOI: 10.1109/TPAMI.2016.2577031

[23] Z.Li, P.Tan, R. T.Tan, D.Zou, S.Zhiying Zhou, L. F. Cheong, "Simultaneous video defogging and stereo reconstruction." In Proceedings of the IEEE conference on computer vision and pattern recognition, 4988-4997, 2015. DOI: 10.1109/CVPR.2015.7299133

[24] M.Sulami, I.Glatzer, R.Fattal, M.Werman, "Automatic recovery of the atmospheric light in hazy images." In 2014 IEEE International Conference on Computational Photography (ICCP), 1-11, 2014. DOI: 10.1109/ICCPHOT.2014.6831817

[25] G.Meng, Y.Wang, J.Duan, S.Xiang, C.Pan, "Efficient image dehazing with boundary constraint and contextual regularization." In Proceedings of the IEEE international conference on computer, 617-624, 2013. DOI:10.1109/ICCV.2013.82

[26] D.Berman, T.Treibitz, S.Avidan, "Air-light estimation using haze-lines." In 2017 IEEE International Conference on Computational Photography (ICCP), 1-9, 2017. DOI: 10.1109/ICCPHOT.2017.7951489

[27] J. P. Tarel, N. Hautiere, "Fast visibility restoration from a single color or gray level image." In 2009 IEEE 12th International Conference on Computer Vision, 2201–2208, 2009. DOI: 10.1109/ICCV.2009.5459251

[28] K.He, J.Sun, X. Tang, "Single image haze removal using dark channel prior." IEEE transactions on pattern analysis and machine intelligence, 33(12), 2341-2353, 2010. DOI: 10.1109/TPAMI.2010.168

[29] Q.Zhu, J.Mai, L.Shao, "A fast single image haze removal algorithm using color attenuation prior." IEEE transactions on image processing, 24(11), 3522-3533, 2015. DOI: 10.1109/TIP.2015.2446191

[30] J.Sturm, N.Engelhard, F.Endres, W.Burgard, D.Cremers, "A benchmark for the evaluation of RGB-D SLAM systems." In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 573-580, 2012. DOI: 10.1109/IROS.2012.6385773

[31] F. Mal, S. Karaman "Sparse-to-dense: Depth prediction from sparse depth samples and a single image." In 2018 IEEE International Conference on Robotics and Automation (ICRA), 1-8, 2018. DOI: 10.1109/ICRA.2018.8460184

[32] B.Cai, X.Xu, D. Tao, "Real-time video dehazing based on spatio-temporal mrf." In 2016 Springer/Cham, Pacific Rim Conference on Multimedia, 315-325,2016. DOI: 10.1007/978-3-319-48896-7_31

[33] C.Sakaridis, D.Dai, L. Van Gool, "Semantic foggy scene understanding with synthetic data." International Journal of Computer Vision, 126(9), 973-992, 2018. https://doi.org/10.1007/s11263-018-1072-8