Hybrid Deep Learning Based Atmospheric Visibility From Images

K.Saranya
UG Scholar
Department of CSE
saranyak.cs19@gmail.com

M.Vijayalakshmi
UG Scholar
Department Of CSE
vijayalakshmim4777@gmail.com

Mrs.P.Jenifer
Assistant professor (CSE)
Francis Xavier Engineering College
Jeniferjebavaram@gmail.com

Abstract-Atmospheric visibility may change drastically in a matter of minutes which calls for real-time visibility monitors in air traffic control, pollution monitoring, and accident detection (e.g., fire accident or arson). Estimating atmospheric visibility from a photo has a high potential in real-time and ubiquitous monitoring of smog and air pollution.

This method proposes a deep learning approach for directly estimating relative atmospheric visibility from outdoor photos. Normally Convolutional Neural Network (CNN) architecture is responsible for learning the overall visibility from the whole image. Hence this project in-order to obtain coarse-to-,fine representation, the strength of Recurrent Neural Network (RNN) is combined with CNN. To enrich the training data set, we automatically synthesize additional order constraints by employing a dehaze filter. Empowered by a large-scale training data repository, the CNN-RNN features are data-driven which benefit from the rich feature hierarchies inherent in the RNN-CNN architecture. CNN-RNN features derived from the rich feature hierarchies inherent in the training data, while in the top structure we metabolize a stable ranking learning solver for relative Random Forest (RF).

Keywords-Convolutional neural network,Recurrent neural networkALEX NET,Random Forest

I. Introduction

Smog is a type of severe air pollution, that reduces visibility.Estimating atmospheric visibility from a photo has a high potential in real-time and ubiquitous monitoring of smog and air pollution.In the smog affected image the visibility might not be clearance for the visibility of the image a combination of CNN RNN algorithms are used. Quantitative measures of atmospheric visibility are increase being used for purposes other than navigation. For example, measures of visibility are being used as indirect estimates of air pollution especially where direct measurements are not available. They are used to estimate solar irradiance which is important for determining where to situate solar energy farms and for forecasting the near term energy output of existing farms. And, visibility measurements are central to the United States' Environmental Protection Agency's (EPA) goal for improving visual air quality in the Class I Federal areas which include 156 national parks and wilderness areas.Expanding visibility monitoring is key to the EPA's mandates and the agencies charged with monitoring typically use a combination of three techniques. First, they utilize specialized equipment such as transmission meters, which measure light extinction, and nephelo meters, which measure light scattering. Second, they use Mie scattering theory to calculate visibility based on measurements of air- borne particulates. Visibility cameras are currently used for qualitative purposes only such as providing visual examples of good and bad days. There is significant opportunity to use these images to derive quantitative measures of visibility perhaps not as accurately as specialized equipment but at much lower cost and possibly even by piggy-backing onto existing web-connected cameras.

Haze is formed by the scattering or absorption of light passing through it by droplets of water suspended in the air or by a large number of tiny particles. The image obtained under haze and other weather conditions not only have serious color attenuation, low contrast and saturation, and poor visual effects, but also affect various systems that rely on optical imaging instruments, such as satellite remote sensing systems, aerial photography systems, and outdoor monitoring and target recognition systems. It brings certain difficulties to the research objectives. Therefore, in order to effectively improve the quality of haze-degraded images and reduce the impact of haze and other meteorological conditions on outdoor imaging systems, it is an urgent and realistic need for effective dehazing and sharpness recovery processing of hazy image.

At present, there are two main categories of methods for hazy image processing: image processing based enhancement and physical model-based restoration. The image processing based enhancement method starts from the image itself, and does not consider the specific cause of the image degradation. By enhancing the contrast and brightness of the image,

the visual effect of the image is improved to achieve the purpose of clarity. These method are generally mature and efficient, and the results of processing can also meet the requirement of clarity. However, such methods cannot adapt to different scenes and images. In particular, the image with more changes in scene depth is not effective. More importantly,the method is based on image enhancement, without considering the process of fog quality reduction, and cannot improve image definition very well. It cannot remove the fog to restore the original appearance, and the distortion of the result is more serious. The processed image not only has poor visual effect but also is not conducive to subsequent processing. The physical model-based restoration method analyzes the specific causes of image degradation and establishes a degraded model of fog-degraded images. In recent years, significant advances have been made in image processing methods based on physical models. In the method, the fog image is described as the superposition of scene radiation and scattering effects, and widely used in haze image modelling.

The deep network is designed to be two stages: _first, we use the convolutional neural networkto estimate the intermediate transmission map in the first phase; in the second stage, the ratio of foggy image and transmission map is introduced into residual network by using the results of the first stage. The residual network predicts the residual image, that is, the difference betweenthe haze image and the potential clean image, instead of directly outputting the dehazed image. The advantage is that, through the residual learning to characterize the identity mapping or the approximate identity mapping, the effect is better than directly learning. Different from the previous method of using deep learning, the model in this paper applies residual network to image dehazing. We only need to estimate the intermediate transmission map without considering the value of atmospheric light, thereby avoiding the influence of invalid parameters on the network structure.

The recurrent neural network (RNN) which processes inputs sequentially, attending to different locations within the images (or video frames) one at a time, and incrementally combines information from the fixations to build a dynamic internal representation of the scene. Instead of processing an entire image , at each step, the model selects the next location to attend to based on past information and the demands of the task. Both the number of parameters in our model and the amount of computation it performs can control independent of the size of the input image, which is in contrast to convolutional networks whose computational demands scale linear with the number of image pixels. We describe an end-to-end optimization procedure that allows the model to be trained directly a given task and to maximize a performance measure which may depend on the entire sequence of decisions made by the model. This procedure uses back propagation to train the neural-network components and policy gradient to address the non-different abilities due to the control problem.

Computer vision methods have restored clear-day visibility of scenes using neither special radiation sources nor external knowledge about the scene structure or aerosols .These methods rely only on the acquired images, weather conditions to change between image acquisitions. This can take too long to make dehazing practical. They require that the scattering properties will not vary with wavelength. The weather conditions does not need to change, and can thus be applied instantly. The analysis of polarization filtered images has prove to be useful for computer vision. For example, it was used to analyze specularities separate transparent and semi-reflected scenes , classify materials,and segmenting scenes .We note that advances in polarimetric cameras enable acquisition of polarization information in real time.

The degradation of images by fog and mist is a problem. In the literature on atmospheric propagation, distributions of particles such as fog, mist, cloud, and haze are collectively known as atmospheric aerosols.

These effect result in loss of contrast that is characteristic of poor visibility conditions. The distance from the camera to the object increases the level of contrast reduction Many algorithms are available for enhancing image contrast. Of these algorithms, perhaps the best known is histogramequalization. However, these algorithms are designed for images that are stationary in the sense that the image properties are roughly constant across the image.

II. Related work

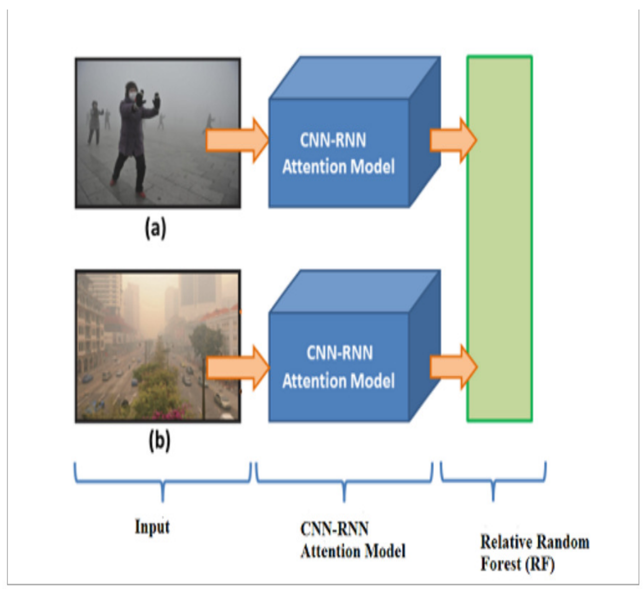
Neighbouring pixels should have similar depths. We show how a stronger prior based on camera geometry can be used to improve the results of any of the single image estimation methods . Our key observation is that weather degradation occurs in outdoor scenes, which means the majority of the images should exhibit the geometry of a camera located above a ground plane. As we will show in Section III, this geometry leads to a simple relationship that objects which appear closer to the top of the image are usually further away. Furthermore, we will show that within the graph cut based α-expansion energy minimization framework, our trend can be implemented as a preference, and does not always have to hold.

Every patch has uniform reflectance, and that the appearance of the pixels within the patch can be expressed in terms of shading and transmission. The shading and transmission signals are unrelated and used independent as a component analysis to estimate the appearance of each patch. The method works quite well for haze, but has difficulty with scenes involving fog, as the magnitude of the surface reflectance is much smaller than that of the airlight when the fog is suitably thick.From a single weather degraded input image a system is developed for estimating depth.

Motivated by the fact that contrast is minimized in a foggy image, Tan divided the image I into a series of small patches and postulated that the corresponding patch in J should have a higher contrast (where contrast was quantified as the sum of local image gradients). He employed a Markov Random Field to incorporate the prior that neighbouring pixels should have similar transmission values t_i. The method tends to produce over enhanced images in practice

This developed a new method for reversing contrast loss in diffuse daylight illumination conditions is described. In this method, a priori information about scene geometry is exploited together with a physical model for aerosol degradation in order to provide useful improvements in effective visibility. An image processing algorithm is described that is automatic in that the required physical parameters are estimated using only the image data. Two applications are considered, the enhancement of single images and the enhancement of image sequences. This paper is organized as follows. The imaging model is presented in Section II together with background material on the propagation of light in atmospheric aerosols.

Some haze removal techniques make use of multiple images; e.g., images taken under different weather conditions or with different polarizer orientations . Since we are interested in dehazing single images, taken without any special equipment, such methods are not suitable for our needs. There are several works that attempt to remove the effects of haze, fog, etc., from a single image using some form of depth information. For example, Aerial imagery using estimated terrain models. However, their method involves estimating a large number of parameters, and the quality of the reported results is unlikely to satisfy today's digital photography enthusiasts. dehaze single images based on a rough depth approximation provided by the user, or derived from satellite orthophotos. The very latest dehazing methods re able to dehaze single images by making various assumptions about the colors in the scene.



A) Random Forest

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it's simplicity and the fact that it can be used for both classification and regression tasks.One big advantage of random forest is, that it can be used for both Regression and classification problems, which form the current machine learning systems. The random forest in classification is sometimes considered the building block of machine learning.

Random Forest is a flexible and simple machine learning algorithm used for classification and regression tasks. And then aggregates the votes from different decision trees to decide the final class of the test object.The classification is performed in two stages.

Random forest creation
Random forest prediction

B) Random Forest creation

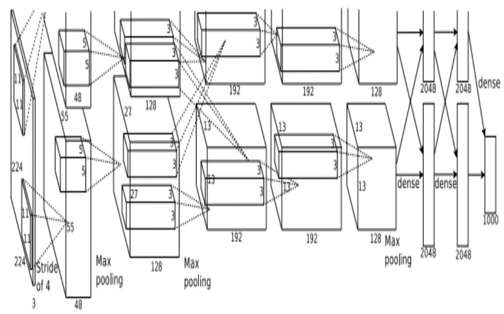
Randomly select **"k"** features from total **"m"** features, where **k << m**
Among the **"k"** features, calculate the node "d" using the best split point.
Split the node into daughter nodes using the best split.
Repeat 1 to 3 steps until **"l"** number of nodes has been reached.
Build forest by repeating steps 1 to 4 for **"n"** number times to create **"n"** number of trees which forms the random forest.

C)Random forest prediction

The procedure to perform prediction using the trained random forest algorithm is given below:
Takes the test features and use the rules of each randomly created decision tree to predict the outcome and stores the predicted outcome (target).
Calculate the votes for each predicted target.
Consider the high voted predicted target as the final prediction from the random forest algorithm.

III.AlexNetArchitecture

It give the solution for image classification problem where the input is an image of one of 1000 different classes and vector of 1000 numbers is the output . The ith element of the output vector is interpreted as the probability of the input image belonging to the ith class. Thus, the sum of all elements of the output vector is 1. AlexNet was much larger than that of previously used CNNs for computer vision tasks (e.g. Yann LeCun'sLeNet paper in 1998). It has 60 million parameters and 650,000 neurons and took 5 to 6 days to train on two GTX 580 3GB GPUs. Today there are much complex CNNs that run on faster GPUs in very efficiently even on very large datasets. But in 2012, this was huge.



AlexNet consists of **5 Convolutional Layers** and **3 Fully Connected Layers**.

AlexNet consists of five Convolutional Layers and three Fully Connected Layers. Multiple Convolutional Kernels (a.k.a filters) take out an interesting features in the image. There are usually many kernels of the same size, in a single convolutional layer. For example, the first Convolutional Layer of AlexNet contains 96 kernels of size 11x11x3. Usually the width and height of the kernel are same and the depth is the same as the number of channels. The first two Convolutional layers are followed by the Overlapping Max Pooling layers that describe next. directly connected layers are third, fourth and fifth convolutional layers . The fifth convolutional layer will come next by an Overlapping Max Pooling layer, the output of this gives a series of two fully connected layers. The second connected layer feeds into softmax classifier with 1000 class labels. ReLU nonlinearity is applied after the convolution and fully connected layers. The ReLU nonlinearity of the first and second convolution layers are followed by local normalization before doing pooling. But researchers later didn't find normalization very useful. So we will not go in detail over that.

A) Overlapping Max Pooling

Max Pooling layers are usually used to downsample the width and height of the tensors, keeping the depth same. Overlapping Max Pool layers are same as that of Max Pool layers, except the adjacent windows over which the max is computed overlap each other. The authors used pooling windows of size 3×3 with a stride of 2 between the adjacent windows. This overlapping nature of pooling help to reduce the top-1 error rate by 0.4% and top-5 error rate by 0.3% respectively when compared to using non-overlapping pooling windows of size 2×2 with a stride of 2 that would give same output dimensions.

U Nonlinearity

The important feature of AlexNet is the use of ReLU Nonlinearity. Tanh or sigmoid activation functions used for the usual way to train the neural network model. AlexNet showed that using ReLU nonlinearity, deep CNNs could trained much faster than using saturating activation functions like tanh or sigmoid. The figure below from the paper shows that using ReLUs(solid curve), AlexNet could achieve a 25% training error rate six times faster than an equivalent network using tanh(dotted curve).

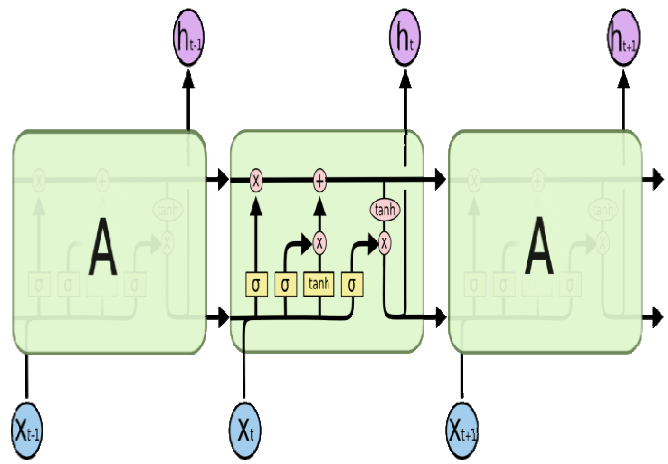
C)Dropout

In dropout, with a probability of 0.5 a neuron is dropped it does not contribute to either forward or backward propagation. So every input goes through a different network architecture . So the learnt weight parameters are more robust and do not get overfitted easily. During testing, there is no dropout and during testing the whole network is used, but output is scaled by a factor of 0.5 to account for the missed neurons . Dropout increases the number of iterations needed to converge by a factor of 2, but AlexNet would overfit substantially,without dropout

D)Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network (RNN) architecture that remember values over arbitrary intervals. LSTM is well-suited to predict, process and classify time series given time lags of unknown duration.

The structure of RNN is very similar to hidden Markov model. The main difference is with how parameters are constructed and calculated . One of the advantage with LSTM is insensitivity to gap length. RNN and HMM rely on the hidden state before emission sequence. If we predict the sequence after 1,000 intervals instead of 10, the model forgot the starting point . LSTM REMEMBERS.



The long-term memory is called as cell state. The looping arrows indicate recursive nature of the cell. This allows information from previous intervals to be stored with in the LSTM cell. Cell state is modified by the forget gate below the cell state and also adjust by the input modulation gate. From this equation, the previous cell state forgets by multiply with the forget gate and adds new information through the output of the input gates.

The remember vector is called the forget gate. The output of the forget gate tells the cell state which information has to be get by multiplying 0 to a position in the matrix. If the output of the forget gate is 1, the information is kept in the cell state.

The save vector is usually called the input gate. These gates determine what information should enter the cell state long-term memory. The important parts are the activation functions for each gates. The input gate is a sigmoid function that have a range of [0,1]. Because the equation of the cell state is a summation between the previous cell state, sigmoid function alone will only add memory and not be able to remove or forget memory. If it can only add a float number between [0,1], that number will never be zero / turned-off / forget. This is because the input modulation gate has an tanh activation function. Tanh has a range of [-1, 1] that allows the cell state to forget memory.

The focus vector is called as output gate. The working memory is called as hidden state.

IV.Convolutional Layer

Convolutional Neural Network is a type of Feedforward Neural Network.A CNN is a neural network that processes sequentially the input through layers that apply transformations to their inputs.

The main task of the convolutional layer is to detect local conjunctions from the previous layer and map the appearance to a feature map. As a result of convolution in neuronal networks, the image is split into perceptrons, creating local receptive fields and finally compressing the perceptrons in feature maps of size m2 × m3. Thus, this map stores the information where the feature occurs in the image and how well it corresponds to the filter. Each filter is trained in regard to the position in the volume. In each layer, there is a bank of m1 filters. The filters applied in one stage is equivalent to the depth of the volume of output feature maps. The filter detects a feature at every location on the input.

. If both pixels are white (a value of 1) then 1 * 1 = 1. If both are black, then (-1) * (-1) = 1. Every matching pixel gives results as 1. Similarly, any mismatch is a -1. All the pixels in a feature match adding them up and dividing by the total number of pixels gives a result 1. Similarly, if none of the pixels in a feature match the image patch, then the answer is a -1. o complete our convolution, we repeat this process, lining up the feature with every possible image patch. We take the answer from each convolution and new two-dimensional array is formed from it, based on where in the image each patch is located. The original image is a filtered version of map of matches. It's a map of where in the image the feature is found. Values close to 1 show strong matches, values close to -1 show strongmatches for the photographic negative of our feature, and values near zero show no match of any sort.

A) Pooling Layer

Another power tool that CNNs use is called pooling. Pooling is a method to take large images and shrink them down while preserving the most important information in them. The math behind pooling is second-grade level at most. It consists of stepping a small window from the given image and taking the maximum value from the window at each step. In practice, a window steps of 2 pixels are work well.

After this layer an image has about a quarter as many pixels as it started with.Each window keeps a maximum value, it preserves the best fits of each feature within the window. This means that it doesn't care so much exactly where the feature fit as long as it fit somewhere within the window. The result of the CNNs can find whether a feature in an image without worrying about where it is. This helps to solve the problem of computers being hyper-literal. The operation of pooling layer is just perform pooling on an image or a collection of images. Theoutput will each have fewer pixels but the numbe4r of images is same. This will help in managing the computational load. An 8 megapixel imageis taken down to a 2 megapixel image makes life a lot easier for everything downstream.

After this layer an image has about a quarter as many pixels as it started with.Each window keeps a maximum value, it preserves the best fits of each feature within the window. This means that it doesn't care so much exactly where the feature fit as long as it fit somewhere within the window. The result of the CNNs can find whether a feature in an image without worrying about where it is. This helps to solve the problem of computers being hyper-literal. The operation of pooling layer is just perform pooling on an image or a collection of images. Theoutput will each have fewer pixels but the numbe4r of images is same. This will help in managing the computational load. An 8 megapixel imageis taken down to a 2 megapixel image makes life a lot easier for everything downstream.

B) Rectified Linear Units (ReLU)

The important player in this process is the Rectified Linear Unit or ReLU. It's math is also very simple—wherever a negative number occurs, swap it out for a 0. It keeps learned values from getting stuck near 0 or blowing up toward infinity so CNN stay mathematically healthy. It's the axle grease of CNNs—not particularly glamorous, but without it they don't get very far. The output of a ReLU layer and the actual size as whatever is put into it is same, just with all the negative values removed.

A) Dense Layer:

A dense layer is a type of regular layer of neurons in a neural network. Each neuron recieves input from all the neurons in the previous layer, thus densely connected.dense layer represents a matrix vector multiplication. (assuming your batch size is 1) The values in the matrix are the trainable parameters which is updated during backpropagation. So you get a m dimensional vector as output. A dense layer thus is used to change the vector dimension. Mathematically speaking, it applies a rotation, scaling, translation transform to your vector.

D) Dropout Layer

A dropout layer is used for regularization where you randomly set some of the dimensions of your input vector to be zero with probability keep_. A dropout layer has no trainable parameters i.e. nothing gets updated during backward pass of back propagation. To ensure that the expected sum of vectors fed to this layer remains the same if no dropout was applied,then the remaining dimensions which are not set to zero are scaled by 1/keep_prob. Dropout is a regularization technique, which aims to prevent overfitting.

Using “dropout", you randomly deactivate certain units (neurons) in a layer with a certain probability p from a Bernoulli distribution (typically 50%, but this yet another hyper parameter to be tuned). So, if you set half of the activations of a layer to zero, the neural network won’t be able to rely on particular activations in a given feed-forward pass during training. As a consequence, the neural network will learn different, redundant representations; the network can’t rely on the particular neurons and the combination (or interaction) of these to be present. Another nice side effect is that training will be faster.

E)Fully connected layers

Fully connected layers take the filtered images and translate them into votes. For Example, we only have to decide between two categories, X and O. Fully connected layers are primary building block of traditional neural networks. Instead of treating inputs as a two-dimensional array, they are treated as a single list identically. Every value gets its own vote on whether the current image is an X or and O. However, the process isn’t entirely democratic. Some values are better than others at knowing when the image is an X, and some are particularly good at knowing when the image is an O. These get larger votes than the others. These votes are expressed as weights, or connection strengths, between each value and each category.

When a new image is presented in CNN, it percolates through the lower layers until it reaches the fully connected layer at the end. Then an election is held. The answer with the most votes wins and declared the category of the input.

V.Random Forest Algorithm

Random forests are a popular ensemble method that can be used to build predictive models for both classification and regression problems. Ensemble methods use multiple learning models to gain better predictive results in random forest, the model creates an entire forest of random uncorrelated decision trees to arrive at the best possible answer. Random Forest is a supervised learning algorithm.

Random Forest has nearly the same hyper parameters as a decision tree or a bagging classifier. Fortunately, you don’t have to combine a decision tree with a bagging classifier and can just easily use the classifier-class of Random Forest. Instead of searching for the most important feature it searches for the best feature among a random subset of features. This results in a better model.

Therefore, in Random Forest, only a random subset of the features is taken into consideration for splitting a node. You can even make trees more random, by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).Advantage of random forest is, that it can be used for both classification and regression problems classification is sometimes considered the building block of machine learning

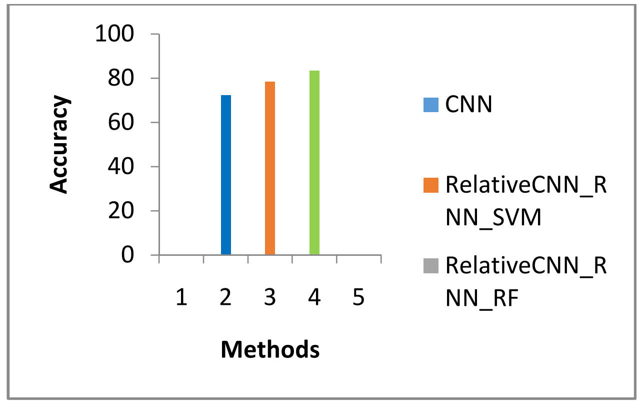
In RF and other ranking learning methods with combined haze and GIST feature, the disadvantage is obvious: handcrafted features are inadequate to describe atmospheric visibility for diversified scenes. The pure CNN feature benefits our relative deep learning framework, but the solution in the top layer is less optimal and weak in modeling ranking. In comparison,our framework is empowered by the data-driven CNN-RNN features derived from the rich feature hierarchies inherent in the training data, while in the top structure we metabolizea stable ranking learning solver for RF .

VI.Conclusion & Future Enhancement

We propose the relative CNN-RNN model, where the complementarysynergy of CNN and RNN modules, with theformer being "local-to-global" and the latter being "global-to local,"is effectively utilized.
This results in good performancein predicting relative visibility for a great variety of situations.
To enrich the training data set, we automatically synthesizeadditional order constraints by employing a dehaze filter.Empowered by a large-scale training data repository, the CNN-RNN
features are data-driven which benefit from the rich feature hierarchies inherent in the RNN-CNN architecture.

The relative CNN-RNN coarse-tofinemodel, where CNN stands for convolutional neural networkand RNN stands for recurrent neural network, exploits the jointpower of RF which has a good ranking representation,and the data-driven deep learning features which are derivedfrom our novel CNN-RNN model. This avoids the problem of handcrafted scene/haze featureswhichfall short of adequately accommodating high varietyof different outdoor scenes. Our relative model can be effectivelyadapted in a small data scenario where absolute visibility data are typically sparsely available. Our frameworkis scalable to include more data which are arguably notdifficult to annotate since human and computer are better inrelative judgment when it comes to visibility measurement.

Methods	Accuracy
CNN	72.36
Relative CNN-RNN-SVM	78.49
Relative CNN-RNN-RF	83.46



References:

[1] Yang You, Cewu Lu,"Relative CNN-RNN: Learning Relative Atmospheric Visibility from Images" ,2018.

[2] P. Carr and R. Hartley, "Improved single image dehazing using geometry," in *Digital Image Computing: Techniques and Applications, 2009.DICTA'09*, 2009.

[3] R. Fattal, "Single image dehazing," in SIGGRAPH, 2008

[4] R. T. Tan, "Visibility in bad weather from a single image," in CVPR, 2008.

[5] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE TIP*, vol. 7, no. 2, pp. 167–179, 1998.

[6]NARASIMHAN, S. G., AND NAYAR, S. K. 2003. Interactive (de)weathering of an image using physical models. IEEE Workshop on Color and Photometric Methods in Computer Vision.

[7]SCHECHNER, Y. Y., NARASIMHAN, S. G., AND NAYAR, S. K. 2003. Polarization-based vision through haze. Applied Optics 42, 3, 511–525.

[7]FATTAL, R. 2008. Single image dehazing. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2008) 27, 3, 73.