A TWO-BRANCH NEURAL NETWORK FOR NON-HOMOGENEOUS DEHAZING VIA ENSEMBLE LEARNING
A TWO-BRANCH NEURAL NETWORK FOR
NON-HOMOGENEOUS DEHAZING VIA ENSEMBLE LEARNING

BY
YANKUN YU, B.Eng.

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

© Copyright by Yankun Yu, April 2021
All Rights Reserved
Lay Abstract

Recently, there has been rapid and significant progress on image dehazing. Many deep learning based methods have shown their superb performance in handling homogeneous dehazing problems. However, we observe that even if a carefully designed convolutional neural network (CNN) can perform well on large-scaled dehazing benchmarks, the network usually fails on the non-homogeneous dehazing datasets introduced by NTIRE challenges. There are two main reasons. Firstly, the non-uniformly distributed haze is harder to be removed than the homogeneous haze. Secondly, the NTIRE challenge only provides limited data. To this end, we propose a simple but effective approach for non-homogeneous dehazing via ensemble learning. To be specific, we introduce a two-branch neural network to separately deal with the aforementioned problems and then map their distinct features by a learnable fusion tail. We present extensive experimental results to illustrate the effectiveness of our proposed method.
Abstract

Recently, there has been rapid and significant progress on image dehazing. Many deep learning based methods have shown their superb performance in handling homogeneous dehazing problems. However, we observe that even if a carefully designed convolutional neural network (CNN) can perform well on large-scaled dehazing benchmarks, the network usually fails on the non-homogeneous dehazing datasets introduced by NTIRE challenges. There are two main reasons. Firstly, due to its non-homogeneous nature, the non-uniformly distributed haze is harder to be removed than the homogeneous haze. Secondly, the NTIRE challenge only provides limited data (there are only 25 training pairs in NH-Haze 2021 dataset). Thus, learning the mapping from the domain of hazy images to that of clear ones based on very limited data is extremely hard. To this end, we propose a simple but effective approach for non-homogeneous dehazing via ensemble learning. To be specific, we introduce a two-branch neural network to separately deal with the aforementioned problems and then map their distinct features by a learnable fusion tail. We present extensive experimental results to illustrate the effectiveness of our proposed method. The source code is available at https://github.com/liuh127/Two-branch-dehazing.
To my dear family
Acknowledgements

I would like to express my deepest and most sincere gratitude to all those who have helped me to complete this thesis successfully. First of all, I am very grateful to my supervisor, Professor Jun Chen. He has not only provided initial and continuous support for this dissertation, but also provided selfless and continuous support, patience, and guidance throughout my graduate studies.

Furthermore, I would like to thank Mr. Huan Liu and Mr. Minghan Fu. We participated as a team in the NTIRE2021 non-homogeneous dehazing challenge. I want to say thank you for all your patience and help with paper writing as well as helping me out with all the coding and debugging in our projects.

Last but not least, I want to give my thanks to my parents for their unconditional support and care for me. I would also like to thank my girlfriend for the encouragement she gave me during my master’s studies.
Contents

Lay Abstract iii

Abstract iv

Acknowledgements vi

Notation and Abbreviations xiii

1 Introduction 1

2 Background and Related Work 6

2.1 Single Image Dehazing . . . . . . . . . . . . . . . . . . . . . . . . . 6
2.2 Ensemble Learning . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
2.3 Transfer Learning . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
2.4 Attention Modules . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

3 Proposed Method 17

3.1 Network Structure . . . . . . . . . . . . . . . . . . . . . . . . . . 17
3.2 Loss Functions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
4 Experiments

4.1 Datasets .................................................. 25
4.2 Training Details .......................................... 27
4.3 Ablation Analysis ......................................... 28
4.4 Comparisons with the State-of-the-art ............... 30
4.5 Runtime Analysis ......................................... 40

5 Conclusion .................................................. 42
List of Figures

1.1 Qualitative results of our proposed method on NH-Haze 2021 validation set. Our method is able to produce haze-free image with high perceptual quality. .................................................. 2

2.1 The architecture of GCANet. GCANet used a basic auto-encoder structure. Several smoothed dilated resblocks are inserted between encoder and decoder to aggregate context information without gridding artifacts. (Image originally used in Chen et al. (2019)) ............. 7

2.2 The FFA-Net architecture. After passing one feature extraction part, the hazy input is fed into several Group Structures. The output features of Group Structures are fused through Feature Attention module. Finally the features will be reconstructed to haze-free output. (Image originally used in Qin et al. (2020)) ............................................. 8

2.3 Total framework of our HardGAN. The haze-free images generated by the generator are fed into the discriminator with full-scale haze images to discriminate the generated haze-free images real or fake. (Image originally used in Deng et al. (2020)) .................................................. 9

2.4 Haze-Aware Representation Distillation (HARD) Module. (Image originally used in Deng et al. (2020)) .................................................. 9
4.8 Runtime comparison of different methods.
List of Tables

3.1 The performance of our proposed method using fusion tails with different sizes. The evaluation is conducted on NTIRE2021 validation set. Scores are provided by NTIRE2021 online server. ........................................ 21

4.1 Ablation study results. “TL” denotes transfer learning sub-net. “CDF” denotes current data fitting sub-net. ”√” denotes that the network is loaded with pre-trained weights, while ”-” means no weights loaded. . 30

4.2 Quantitative comparisons over SOTS-indoor and O-HAZE for different methods. The Best results are in bold, and the second best are with underline. .................................................. 32

4.3 Quantitative comparisons over Dense-Haze, NH-Haze and NH-Haze2 for different methods. The Best results are in bold, and the second best are with underline. .................................................. 32

4.4 Parameters of each method. .................................................. 40
## Notation and Abbreviations

| Acronym  | Description                                                      |
|----------|------------------------------------------------------------------|
| AI       | Artificial Intelligence                                         |
| ASM      | Atmosphere Scattering Model                                     |
| CNN      | Convolutional Neural Network                                     |
| GPU      | Graphics Processing Unit                                         |
| MS-SSIM  | Multi-scale Structure similarity                                |
| NH-Haze  | Non-homogeneous Haze                                             |
| NTIRE    | New Trends in Image Restoration and Enhancement workshop and challenges |
| O-Haze   | Outdoor Haze                                                     |
| PSNR     | Peak signal-to-noise ratio                                      |
| ReLU     | Rectified Linear unit                                            |
| RESIDE   | REalistic Single Image DEhazing dataset                          |
| SSIM     | Structural Similarity Index                                      |
Chapter 1

Introduction

Single image dehazing as a low-level vision task has gained widespread attention in recent years. In the natural atmosphere, there are smoke, dust, haze, and other atmospheric phenomena that affect visibility. Pictures taken in these environments are often affected by blurring, color distortion, and low contrast problems. Using these kinds of pictures for classification, image segmentation, and other high-level vision tasks significantly reduces prediction accuracy. Single image dehazing aims to restore a clean output image from a hazy input. To this end, many dehazing methods (He et al., 2010; Li et al., 2017; Cai et al., 2016; Zhang and Patel, 2018; Wu et al., 2020; Qin et al., 2020; Liu et al., 2020; Zhou et al., 2020; Ren et al., 2018) have been proposed.

A physical model, known as the atmospheric scattering model (ASM) (Middleton, 1952), provides the mapping formula between the hazy images and clean counterparts. The ASM can be formally written as:

\[ I(x) = J(x)t(x) + A(1 - t(x)). \] (1.0.1)
Figure 1.1: Qualitative results of our proposed method on NH-Haze 2021 validation set. Our method is able to produce haze-free image with high perceptual quality.

$I, J$ respectively denote the hazy and haze-free image. $A$ is the global atmosphere light. $t(x) = e^{-\beta d(x)}$ represents the transmission map, where $\beta$ and $d(x)$ are respectively the atmosphere scattering parameter and the scene depth.

Many dehazing methods (He et al., 2010; Berman et al., 2016; Fattal, 2014; Meng et al., 2013) can achieve good performance based on ASM. However, to generate a haze-free image with ASM, dehazing models need to precisely estimate $t$ and $A$. As a result, when dealing with images in complex environments, the inaccurate prediction of $t$ and $A$ usually leads to unsatisfactory dehazing results. Moreover, since $t$ depends on the distance of the scene, the atmospheric scattering model has a strong assumption: the thickness of haze is strongly correlated to the depth of the background scene.
This property forbids the ASM based methods to handle non-homogeneous hazy images. Somewhat surprisingly, recent years have witnessed the tremendous success of deep learning (LeCun et al., 2015) approach in addressing single image dehazing problem (Li et al., 2017; Zhang and Patel, 2018; Cai et al., 2016; Zhou et al., 2020; Qin et al., 2020; Ren et al., 2018). By using deep learning, the weakness of ASM can be avoided to a certain extent and the problem of exploring appropriate hand-crafted features is reduced to that of building a suitable convolutional neural network (CNN). With the availability of powerful CNNs, one can readily train them on large-scale datasets to learn a correct mapping from input hazy images to clear outputs. However, it is costly (and in some cases impossible) to acquire vast quantities of hazy images with their corresponding clean ground truths in the real world.

Recently, NTIRE organized several dehazing challenges and introduced several small-scale real-world datasets, O-Haze (Ancuti et al., 2018b), Dense-Haze (Ancuti et al., 2019a) and NH-Haze (Ancuti et al., 2020a). The challenge datasets have two intrinsic difficulties.

- **Limited training data.** Using limited data, CNN based methods usually cannot gather enough statistical information of haze pattern, which results in a bad performance on dehazing. In other words, training on limited data is more likely to suffer from the over-fitting (Ayinde et al., 2019) problem. This problem is considered undesirable since it severely jeopardizes the generalization of models.

- **Complicated haze pattern.** Most of the previous methods simply assume that haze is homogeneous. Since the non-uniformly distributed haze is more challenging to be removed than the homogeneous haze (haze pattern cannot be simply formulated using ASM), existing methods usually fail to produce satisfactory
results when they work on non-homogeneous dehazing.

To cope with the complex distributions of non-homogeneous images and overfitting problem in small-scale datasets, I here introduce this proposed method. To be specific, the main contribution of this work is a simple two-branch neural network to deal with the aforementioned two issues separately. The first branch, namely transfer learning sub-net, is built upon a ImageNet (Deng et al., 2009) pre-trained Res2Net (Gao et al., 2019). ImageNet pre-training helps significantly alleviate overfitting problems, especially before the large-scaled datasets are available to researchers (He et al., 2019). Besides, a pre-trained network is able to provide robust features for transfer learning (Kornblith et al., 2019). Therefore, the ImageNet pre-trained network is of great significance in solving the limited training data problem. However, only using the ImageNet pre-trained model is not enough to tackle these issues. The pre-training on classification task usually fails to perfectly fit the target task, while a better solution is to find for data specific representations (He et al., 2019). To this end, the other branch are added for fitting on current data, current data fitting sub-net. This branch is trained from scratch and optimized only using the current training data. In favor of the strong mapping capability of residual channel attention network (RCAN) (Zhang et al., 2018), we build the current data fitting sub-net using RCAN. Unlike the original network setting (Zhang et al., 2018) that down-samples the input images at the front of the entire network, our second branch always maintains the original resolution of the inputs and avoids using any down-sampling operations. This adjustment avoids losing fine-detailed features. Finally, in order to aggregate the two varied outputs from our two branches, we design a fusion tail for learning a suitable ensemble strategy.
In summary, the main contributions of our work are as follows:

- We demonstrate the effectiveness of using ImageNet pre-training in the non-homogeneous dehazing challenge (Acuti et al., 2021).

- Towards learning data-specific representations, we propose to build current data fitting sub-net as a complement to the transfer learning sub-net. It can extract more distinctive features on current data distribution.

- We adopt the idea from ensemble learning to design a learnable fusion tail. The fusion tail is simple and effective in fusing the outputs from two branches.

- We present extensive experimental results to show the effectiveness of our two-branch network on both small-scale and large-scale datasets.
Chapter 2

Background and Related Work

2.1 Single Image Dehazing

Single image dehazing methods can be roughly divided into two classes: prior-based methods and learning-based methods.

2.1.1 Prior-based Methods

Prior-based methods estimate the transmission map and the global atmosphere light in ASM (Middleton, 1952). Many prior-based methods (He et al., 2010; Berman et al., 2016; Zhu et al., 2015) showed a good performance in single image dehazing. DCP (He et al., 2010; Zhu et al., 2015) is an outstanding prior-based method. It is based on the assumption that at least in one RGB channel some pixels have low intensity. However, due to the prior mismatch in practice, the prior-based methods are not always robust when encountering complicated scenarios.
2.1.2 Deep Learning Based Methods

With the success of deep convolutional neural networks, deep learning based methods received extensive attention in recent years. DehazeNet (Cai et al., 2016) is the first deep learning based dehazing model. It adopts CNN to estimate the transmission map and then generates dehazed images using the physical scattering formulation. Unlike DehazeNet, AOD-Net (Li et al., 2017) is built to estimate both transmission and atmospheric light in one shot. In addition, many recent methods can recover haze-free images without using the physical scattering model. GFN (Ren et al., 2018) is a gated fusion network, which restores the hazy images with several transformations on the input, such as white balancing and gamma correction. GCANet (Chen et al., 2019) employs smoothed dilated convolution layers to eliminate the gridding artifacts.

![Figure 2.1: The architecture of GCANet. GCANet used a basic auto-encoder structure. Several smoothed dilated resblocks are inserted between encoder and decoder to aggregate context information without gridding artifacts. (Image originally used in Chen et al. (2019))](image)

(Qin et al., 2020) proposed FFA-Net with novel feature attention modules, including pixel attention and channel attention. This work achieved high performance on the RESIDE (Li et al., 2019) dataset.
Figure 2.2: The FFA-Net architecture. After passing one feature extraction part, the hazy input is fed into several Group Structures. The output features of Group Structures are fused through Feature Attention module. Finally the features will be reconstructed to haze-free output. (Image originally used in Qin et al. (2020))

Shao (Shao et al., 2020) proposed a novel domain adaptation framework for dehazing tasks. This method bridges the gap between the real-world and synthetic hazy images. However, most of the previous methods perform poorly when using real-world datasets introduced by NTIRE challenges, owing to the small-scale training set and the complex distribution of the haze. To solve the problems in real-world datasets, some approaches are proposed. Liu (Liu et al., 2020) introduced a Trident Dehazing Network(TDN) in the NTIRE2020 non-homogeneous dehazing challenge (Ancuti et al., 2020b). TDN consists of three branches using ImageNet pre-training, deformable convolution and many off-the-shelf techniques. Despite the remarkable achievements of TDN, the complicated network structure impedes us from carefully analyzing the significance of each technique. This solution won the first place in the NTIRE2020 non-homogeneous dehazing competition. Moreover, HardGAN (Deng et al., 2020) shows a good performance in both RESIDE (Li et al., 2019) and NH-Haze (Ancuti et al., 2020c,a) datasets. This method proposed a novel haze-aware
Figure 2.3: Total framework of our HardGAN. The haze-free images generated by
the generator are fed into the discriminator with full-scale haze images to
discriminate the generated haze-free images real or fake. (Image originally used in
Deng et al. (2020))

Figure 2.4: Haze-Aware Representation Distillation (HARD) Module. (Image
originally used in Deng et al. (2020))

feature distillation (HARD) module. The more the number of stacked convolutional
layers, the higher the possibility of gradient vanishing, thereby significantly limiting the representing power of learned features. HARD is introduced to tackle this problem. However, the performance of most of the previous methods deteriorated significantly when using NH-Haze2 dataset due to the small-scale dataset and the non-homogeneous distribution of the haze.

2.2 Ensemble Learning

Single output models are prone to suffer from the statistical problem, the computational problem, and the representation problem. These problems are caused by limited training data, inability to find global minima, and failure to make a good approximation to the ground-truth label, respectively. Ensemble learning can partially solve these three problems (Dietterich et al., 2002). It is based on the understanding that every model has limitations. Thus, ensemble learning aims to manage the strengths and weaknesses of single models. This management can make the best possible decision (Brown, 2010). The ensemble architecture in our method can be classified as Mixtures of Experts (Jacobs et al., 1991). The principle underlying the architecture is that both sub-nets in our model can focus on particular parts of the input space (Brown, 2010). To generate a haze-free output, our fusion layer acts as a gating network that is responsible for learning the proper combination of the outputs from two branches. Moreover, in our ablation studies, we demonstrate that this ensemble architecture fits well in our method.
2.3 Transfer Learning

Transfer learning is an effective way to solve problems with limited data. It aims to enable the system based on knowledge and skills learned in previous tasks to run on novel tasks (Pan and Yang, 2009). This method starts with learning knowledge from one or more source domains and applies it to a target task. In many cases, transfer learning models can achieve high performance when the source domain is closely related to the target domain. According to the survey (Pan and Yang, 2009), transfer learning can be divided into four categories based on "What to transfer." They are instance-based transfer learning, feature-representation-transfer, parameter-transfer, and relational-knowledge-transfer. Our work belongs to the parameter-transfer, which assumes that source tasks and target tasks share some parameters or distributions. The way we use transfer learning is based on the assumption that the network backbone is generalized and can extract versatile features (Tan et al., 2018).
Therefore, we use the parameters of the Res2Net (Gao et al., 2019) that pre-trained on ImageNet (Deng et al., 2009) as a part of our transfer learning sub-net. In our experiments, the network with pre-trained parameters surpasses the randomly initialized network by a large margin in terms of quantitative evaluation.

2.4 Attention Modules

Attention mechanisms are essential in human perception (Corbetta and Shulman, 2002). To better capture visual structure, humans take advantage of a series of partial glimpses and selectively focus on important pieces. Several attempts have recently been made to integrate attention mechanisms into CNNs to enhance their efficiency in large-scale classification tasks. RAN (Wang et al., 2017) divided their attention modules into trunk branch and mask branch for feature processing and
control gates, respectively. SE-Net (Hu et al., 2018) adopts global average-pooling layers to compute attention for each channel. However, the spatial localization of the object in images is ignored by SE-Net, which only considers the contribution of channel-wise information. In this regard, CBAM (Woo et al., 2018) exploits both pixel attention and channel attention in their CNN model. They use 1 x 1 convolution and

---

Figure 2.7: The architecture of the original Residual module (left) and the SE-ResNet module (right). (Image originally used in Hu et al. (2018))
pooling layers for spatial attention. And the pooling layer is concatenated by global average-pooling and max-pooling. Moreover, the channel attention module utilizes both max-pooling outputs and average-pooling outputs with a shared network. For the channel and spatial attention arrangement, CBAM discovers that generating an attention map sequentially infers a better attention map than doing so in a paralleled fashion. To be specific, the channel-first order outperforms the spatial-first order by a small margin.

Figure 2.8: The architecture of the channel attention module and spatial attention module. The Channel attention module utilizes both max-pooling outputs and average-pooling outputs with a shared network; the spatial sub-module utilizes similar two outputs that are pooled along the channel axis and forward them to a convolution layer. (Image originally used in Woo et al. (2018))

In our work, we adopt feature attention (FA) proposed by (Qin et al., 2020), which combines channel and spatial attention module. The formula of channel attention is

```math
M_{C} = \sigma(\text{Shared MLP}([\text{MaxPool}(F), \text{AvgPool}(F)]))
```
defined as

\[ g_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j), \]  

(2.4.1)

\[ CA_c = \sigma(Conv(ReLU(Conv(g_c)))), \]  

(2.4.2)

where \( g_c \) stands for the result of global pooling, \( X_c(i, j) \) stands for the value of \( c \)-th channel \( X_c \) at position \((i, j)\). The shape of the feature map changes from \( C \times H \times W \) to \( C \times 1 \times 1 \). To get the weights of the different channels, features pass through two convolution layers and sigmoid, ReLu activation function latter. The final output \( F_c^* \) with input \( F_c \) is

\[ F_c^* = CA_c \otimes F_c, \]  

(2.4.3)

where \( \otimes \) stands for element-wise product.

Similar to channel attention, the formula of pixel spatial attention is

\[ PA = \sigma(Conv(ReLU(Conv(F_c^*)))), \]  

(2.4.4)

\[ F^* = PA \otimes F_c^*, \]  

(2.4.5)
where $F^*_c$ denotes the output of channel attention module, $F^*$ denotes the output of pixel attention module.

Figure 2.10: Feature Attention Module (Image originally used in Qin et al. (2020))
Chapter 3

Proposed Method

This section introduces the details of our two-branch neural network for non-homogeneous dehazing. Firstly, we specify the details of the transfer learning sub-net and current data fitting sub-net. Then, we provide the loss functions that are employed in the training stage.

3.1 Network Structure

As shown in Figure 3.1, our method consists of two sub-nets, the transfer learning sub-net and the current data fitting sub-net. Each sub-net is used for a specific purpose: transfer learning sub-net extracts robust global representations from input images with pre-trained weights, current data fitting sub-net aims to work on the current data and perform well on the specific training image domain. The fusion layer takes the concatenated feature maps of these two sub-nets and outputs haze-free images.
3.1.1 Transfer Learning Sub-net.

Transfer learning sub-net is an encoder-decoder network. Inspired by (Wu et al., 2020), we use Res2Net as the encoder due to its excellent performance on the classification tasks. To be specific, we only adopt the front part of Res2Net with 16 times down-sampling and discard using the fully connected layer. Moreover, we adopt feature attention Figure 2.10 as the attention module, and PixelShuffle (Shi et al., 2016) as the up-sampling module in this sub-net. Before the features enter fusion tail, they will go through an enhancing module proposed by (Qu et al., 2019). At the training stage, the encoder module loads the ImageNet pre-trained parameters. Models with these pre-trained parameters can better extract robust features than those with randomly initialized parameters. However, as is pointed in (He et al., 2019), despite the effectiveness of using ImageNet pre-training in small-scale datasets, a better solution
is to train the neural network on a large-scale dataset directly. Because accessing an identically distributed large-scale dataset is impossible during the NTIRE challenge, we alternatively consider taking benefits from specific data representation in small-scale challenge data to the utmost extent. Therefore, we build the current data fitting sub-net to achieve this object. In addition, due to the nature of Res2Net, we can not apply a full-resolution skip connection from encoder to decoder. It may cause
the network to lose some of the information that is important for restoring the image details. This motivates us to construct the second branch.

### 3.1.2 Current Data Fitting Sub-net.

Our current data fitting branch is based on residual channel attention block (Zhang et al., 2018) shown in Figure 3.2. The block contains convolutional layers and channel attention modules. Owing to residual designing and long skip connections, the network is less likely to suffer from gradient vanishing problems (Pascual et al., 2013). Besides, channel attention highlights salient features to enhance the model’s fitting ability on current data. Moreover, in order to preserve fine-detailed features, this sub-net avoids employing down-sampling and up-sampling operations. The fine-detailed features can be regarded as complements to that from transfer learning sub-net. Since current data fitting sub-network is trained from scratch and built with full-resolution purpose, it would fit on the current data and extract more specified features. Nevertheless, the powerful fitting ability also makes it easier to encounter over-fitting problems (see in Section 4.3). Thus, the role of transfer learning branch cannot be neglected.

### 3.1.3 Fusion Tail.

Our fusion tail as the ensemble method produces the final outputs. Specifically, the fusion tail takes the concatenation of features from two branches and then learns to map the features to clear images. The fusion tail is built with a convolutional layer followed by a hyperbolic tangent activation function. The reason we only adopt a single convolution operation is that the two branches have already provided sufficient
and distinct features for restoring clear images. In contrast, we find that a heavy fusion tail can potentially jeopardize the overall generalization ability of our method and results in performance degradation. To further illustrate the effectiveness of our fusion tail, we empirically study the effect of adopting fusion tails with different sizes in Table 3.1. It can be observed that with the increasing of fusion tail depth, the performance of our method degrades accordingly.

### 3.2 Loss Functions

Our loss function consists of four different components. Each one is used for a specific purpose.

**Smooth L1 Loss.** We apply the smooth L1 loss to ensure the predicted images are close to clean images. It is a robust L1 loss (Girshick, 2015) that is proved to be better than L2 loss in many image restoration tasks (Zhao et al., 2016):

\[
L_{l1} = \frac{1}{N} \sum_{i}^{N} smooth_{L1}(y_i - f_{\theta}(x_i)),
\]

(3.2.1)
where $y_i$ and $x_i$ denote respectively ground truth and hazy image at pixel $i$, $f_\theta(\cdot)$ denotes our network parameterized by $\theta$, $N$ is the total number of pixels.

**MS-SSIM Loss.** In order to let the network learn to produce visually pleasing results, we adopt Multi-scale Structure similarity (MS-SSIM) as our second loss function. Let $O$ and $G$ denote two windows of common size centered at pixel $i$ in the dehazed image and the clear image, respectively. We apply a Gaussian filter to $O$ and $G$, and compute the resulting means $\mu_O$, $\mu_G$, standard deviations $\sigma_O$, $\sigma_G$, and covariance $\sigma_{OG}$. Let

\[
l(i) = \frac{2\mu_O\mu_G + C_1}{\mu_O^2 + \mu_G^2 + C_1}, \tag{3.2.3}
\]

\[
c(i) = \frac{2\sigma_O\sigma_G + C_2}{\sigma_O^2 + \sigma_G^2 + C_2}, \tag{3.2.4}
\]

\[
s(i) = \frac{\sigma_{OG} + C_3}{\sigma_O\sigma_G + C_3}, \tag{3.2.5}
\]

where $C_1$, $C_2$ and $C_3$ are given by

\[
C_1 = (K_1 L)^2, C_2 = (K_2 L)^2, C_3 = 2C_2/2. \tag{3.2.6}
\]

The SSIM for pixel $i$ is defined as

\[
\text{SSIM}(i) = \frac{2\mu_O\mu_G + C_1}{\mu_O^2 + \mu_G^2 + C_1} \cdot \frac{2\sigma_O\sigma_G + C_2}{\sigma_O^2 + \sigma_G^2 + C_2} = l(i) \cdot cs(i), \tag{3.2.7}
\]
where $C_1$, $C_2$ are two variables to stabilize the division with weak denominator, $K_1L << 1$ and $K_2L << 1$ are two scalar constants, $l(\cdot)$ denotes the luminance, and $cs(\cdot)$ refer to contrast and structure measures.

The MS-SSIM loss is shown in Eq. 3.2.8 where $\alpha$ and $\beta_j$ are default parameters, $M$ denotes the total number of scales:

$$L_{\text{MS-SSIM}}(i) = 1 - l_M^\alpha(i) \cdot \prod_{j=1}^{M} [cs_j(i)]^{\beta_j}. \quad (3.2.8)$$

**Perceptual Loss.** Unlike the MS-SSIM loss that focuses mainly on the structural similarity, we adopt perceptual loss (Zhu *et al.*, 2017) to provide additional supervision in high-level feature space. It has been acknowledged that training with perceptual loss allows the model to better reconstruct fine details. The loss network $\phi$ is VGG-16 (Simonyan and Zisserman, 2014) that is pre-trained on ImageNet. Then we use the features from the last layer of each of the first three stages (Conv1-2, Conv2-2 and Conv3-3). The loss function is described as

$$L_{\text{perc}} = \frac{1}{N} \sum_j \frac{1}{C_jH_jW_j} ||\phi_j(f_\theta(x)) - \phi_j(y)||_2^2, \quad (3.2.9)$$

where $x$ and $y$ are hazy inputs and ground truth images, respectively, $f_\theta(x)$ is the dehazed images, $\phi_j(\cdot)$ denotes the feature map with size $C_j \times H_j \times W_j$. The feature reconstruction loss is the $L_2$ loss, $N$ is the number of features that used in perceptual loss function.

**Adversarial Loss.** Adversarial loss is proved to be effective in helping restore photo-realistic images (Ledig *et al.*, 2017). Especially for the small-scaled dataset, the pixel-wise loss function usually fails to provide sufficient supervision signals to
train a network for recovering photo-realistic details. Therefore, we finally implement the adversarial loss with the discriminator in (Zhu et al., 2017). The loss function is described as

\[ L_{adv} = \sum_{n=1}^{N} -\log D(f_{\theta}(x)), \quad (3.2.10) \]

where \( D(\cdot) \) denotes discriminator. The probability that the dehazed image \( f_{\theta}(x) \) is a ground truth image is shown as \( D(f_{\theta}(x)) \).

The total loss function is defined as:

\[ L = \gamma_1 L_{l1} + \gamma_2 L_{MS-SSIM} + \gamma_3 L_{perc} + \gamma_4 L_{adv}, \quad (3.2.11) \]

where \( \gamma_1, \gamma_2, \gamma_3, \) and \( \gamma_4 \) are the hyperparameters to balance between different losses.
Chapter 4

Experiments

In this section, we start with the description of datasets, training details, and evaluation metrics. Then we conduct ablation studies to clarify the effects of different modules in our method. Finally, we compare our method with other state-of-the-art dehazing algorithms quantitatively and qualitatively.

4.1 Datasets

We choose both real-world datasets and synthetic datasets to evaluate our network. For real-world datasets, we adopt the O-Haze (Ancuti et al., 2018b) from NTIRE2018 Dehazing Challenge (Ancuti et al., 2018a), Dense-Haze (Ancuti et al., 2019a,b) in NTIRE2019 Dehazing Challenge (Ancuti et al., 2019), NH-Haze (Ancuti et al., 2020c,a) used in NTIRE2020 Dehazing Challenge (Ancuti et al., 2020d), and NH-Haze 2 in NTIRE2021 Dehazing challenge(Ancuti et al., 2021). For synthetic datasets, we choose the Indoor Training set of RESIDE (Li et al., 2019).
4.1.1 Real-world datasets

To verify the performance of our method on real-world datasets, we evaluate the proposed one on NTIRE challenge datasets, O-Haze, Dense-Haze and NH-Haze. O-Haze contains 35 pairs of outdoor hazy images and ground truth images for training. Dense-Haze, NH-Haze and NH-Haze 2 respectively contain 45 dense hazy images, 45 non-homogeneous hazy images, 25 non-homogeneous hazy images and their paired ground truths for training. Each of these datasets has 5 image pairs for validation and 5 pairs for testing. In our experiments, we conduct our evaluation based on the official train, val and test split for O-Haze, Dense-Haze and NH-Haze. For NH-Haze 2, since the ground truth images for the validation set and testing set have not yet been released, we choose the first 20 official training pairs as our training data, and the rest 5 image pairs are used for evaluation. It should be noted that we conduct experiments on these datasets separately and do not use extra data to boost performance.

4.1.2 Synthetic Dataset

RESIDE is a benchmark for single image dehazing, which contains large-scale training and testing images in indoor and outdoor scenarios. The Indoor Training Set (ITS) and the indoor Synthetic Objective Testing Set (SOTS) are used in our experiments. The Indoor Training Set (ITS) of RESIDE consists of 13990 image pairs. The atmospheric scattering model is used to generate hazy images. The range of atmospheric lights is between 0.7 and 1.0 for each channel. And scattering coefficient ranges from 0.6 to 1.8.
4.1.3 NH-Haze and NH-Haze2 Datasets

NH-Haze (Ancuti et al., 2020c,a) represented the first realistic image dehazing dataset with non-homogeneous hazy and haze-free (ground-truth) paired images. The non-homogeneous haze was generated using a professional haze generator that imitates the real conditions of haze scenes. NH-Haze was the dataset of NTIRE2020 image dehazing challenge (Ancuti et al., 2020d). NH-Haze2 is the extended version of the former NH-Haze dataset (Ancuti et al., 2021). The NH-Haze2 consists of 35 hazy images and their corresponding ground truth (haze-free) images of the same scene. NH-Haze2 contains real outdoor scenes with non-homogeneous haze generated using a professional haze setup. The NTIRE2021 image dehazing challenge was built on NH-Haze2.

4.2 Training Details

We augment the training set with 90, 180, 270 degrees of random rotation, horizontal flip, and vertical flip. The input images are randomly cropped to a size of $256 \times 256$. We adopt Adam optimizer with default $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate is 0.0001. The hyper-parameters of loss functions, $\gamma_1$, $\gamma_2$, $\gamma_3$, $\gamma_4$, are 1.0, 0.5, 0.01, 0.0005, respectively. Our method is implemented using the Pytorch library (Paszke et al., 2017). All experiments are conducted on Nvidia V100 GPUs. For quantitative evaluation, we adopt the Peak Signal to Noise Ratio (PSNR) metric and the Structural Similarity Index (SSIM) metric. These two metrics are often used to evaluate image quality in low-level vision tasks.
4.3 Ablation Analysis

To intentionally analyze and evaluate the effectiveness of each architecture component, we conduct ablation studies by considering the combination of three factors: ImageNet pre-trained weights, transfer learning sub-net, and current data fitting sub-net. The ablation experiments are shown as following:

- TL without pre-trained weights: only use the transfer learning sub-net with randomly initialized parameters.
- TL with pre-trained weights: only use transfer learning sub-net with ImageNet pre-trained weights.
- CDF: only use current data fitting sub-net.
- TL + CDF: use both current data fitting sub-net and transfer learning sub-net without ImageNet pre-trained weights.
- Ours: use both transfer learning sub-net with ImageNet pre-trained weights and current data fitting sub-net.

In detail, we use NH-Haze2 as the training set and testing set that respectively formed by first 20 images and the rest 5 images. The quantitative results for ablation studies are shown in 4.1. From the table, we can observe that the ImageNet pre-trained weights can significantly improve the PSNR and SSIM of the transfer learning net. Although CDF is trained from scratch, it still outperforms the TL without pre-training and approaches TL with pre-training in terms of PSNR.

On the other hand, since the current data fitting sub-net is built with full-resolution purpose, it would fit the current data and perform well on the specific
training image domain. Thus, for the SSIM metric, the result of current data fitting sub-net is 0.062 higher than that of transfer learning sub-net with pre-trained weights.

All the two-branch models (TL+CDF and ours) perform better than those only with one branch. This indicates the effectiveness of our two-branch design. It is no surprise that our method with all the components performs best. The PSNR and SSIM of our full model achieve 21.66 db and 0.843. The scores indicate that every factor we consider plays an essential role in the network performance, especially the
ImageNet pre-training.

| Methods       | ImageNet Pre-training | PSNR  | SSIM  |
|---------------|-----------------------|-------|-------|
| TL            | -                     | 17.92 | 0.555 |
| TL √          |                       | 20.78 | 0.753 |
| CDF           | -                     | 19.83 | 0.815 |
| TL + CDF      | -                     | 20.07 | 0.831 |
| Ours √        |                       | 21.66 | 0.843 |

Table 4.1: Ablation study results. “TL” denotes transfer learning sub-net. “CDF” denotes current data fitting sub-net. ”√” denotes that the network is loaded with pre-trained weights, while ”-“ means no weights loaded.

4.4 Comparisons with the State-of-the-art

In this section, we show the comparisons between our method and the state-of-the-art. The comparison is conducted on both synthetic dataset (RESIDE (Li et al., 2019)) and real-world datasets (O-Haze (Ancuti et al., 2018b), Dense-Haze (Ancuti et al., 2019a,b), NH-Haze (Ancuti et al., 2020c,a), and NH-Haze2). In our experiments, we select five state-of-the-art methods, including DCP (He et al., 2010), AOD-Net (Li et al., 2017), GCANet (Chen et al., 2019), FFA (Qin et al., 2020), and TDN (Liu et al., 2020). We also train a TDN without pre-training to verify the effect of ImageNet pre-training on this method. To be specific, we train all models on real-world datasets (O-Haze, Dense-haze, NH-Haze, NH-Haze2). On RESIDE dataset, we train AOD, TDN without ImageNet pre-trained model and ours from scratch. For the other methods, we use the results from their paper and released code. The comparison results are shown in Table 4.2 and Table 4.3. Our method outperforms other methods by a significant margin.
Figure 4.2: Qualitative evaluation on RESIDE(ITS).
# 4.4.1 Results on RESIDE Dataset

The qualitative evaluation on RESIDE is shown in Figure 4.2. We can observe that DCP and AODNet cause severe color distortions owing to the inaccurate estimation of the ASM. The rest of the deep learning models all have PSNR above 30dB, and it is difficult to distinguish the difference between them by human eyes. The RESIDE dataset has a large-scale training dataset and a relatively uniform distribution of haze, which is not difficult for deep neural networks to generate haze-free images.

## Table 4.2: Quantitative comparisons over SOTS-indoor and O-HAZE for different methods. The Best results are in **bold**, and the second best are with underline.

| Methods            | RESIDE(ITS) | NTIRE18(O-Haze) |
|--------------------|-------------|-----------------|
|                    | PSNR | SSIM | PSNR | SSIM |
| DCP                | 16.62 | 0.817 | 12.92 | 0.505 |
| AOD                | 19.06 | 0.850 | 17.69 | 0.616 |
| GCANet             | 30.23 | 0.975 | 19.50 | 0.660 |
| FFA                | 36.39 | 0.988 | 22.12 | 0.768 |
| TDN                | 34.59 | 0.975 | 23.53 | 0.754 |
| TDN(No pre-trained)| 30.52 | 0.960 | 21.67 | 0.721 |
| Our                | **37.61** | **0.991** | **25.54** | **0.783** |

## Table 4.3: Quantitative comparisons over Dense-Haze, NH-Haze and NH-Haze2 for different methods. The Best results are in **bold**, and the second best are with underline.

| Methods                | NTIRE19 | NTIRE20 | NTIRE21 |
|------------------------|---------|---------|---------|
|                        | PSNR   | SSIM   | PSNR   | SSIM   | PSNR   | SSIM   |
| DCP                    | 10.85  | 0.404  | 12.29  | 0.411  | 11.30  | 0.605  |
| AOD                    | 13.30  | 0.469  | 13.44  | 0.413  | 13.22  | 0.613  |
| GCANet                 | 12.42  | 0.478  | 17.58  | 0.594  | 18.76  | 0.768  |
| FFA                    | 16.26  | 0.545  | 18.51  | 0.637  | 20.40  | 0.806  |
| TDN                    | 15.29  | 0.511  | 20.51  | 0.671  | 20.31  | 0.763  |
| TDN(No pre-trained)    | 15.10  | 0.495  | 17.29  | 0.616  | 17.73  | 0.696  |
| Our                    | **16.36** | **0.582** | **21.44** | **0.704** | **21.66** | **0.843** |
Figure 4.3: Qualitative evaluation on O-Haze.
4.4.2 Results on O-Haze Dataset

The results are shown in Figure 4.3. It can be seen that the color distortion of DCP is very serious and the output images turn blue. The haze in the images generated by AODNet is not completely removed. The GCA images are greenish compared to the ground truth images, and some of the haze is not removed. FFA and TDN without pre-training restore visually pleasing images, where the images produced by TDN without pre-training suffer from over-enhancement. The PSNR of both TDN and our method exceeds 23 dB, and both can restore the details of the original image well.

4.4.3 Results on Dense-Haze Dataset

As shown in Figure 4.4, Dense-haze is a difficult dataset for deep learning models. The distribution of dense-haze does not correlate with the depth of the image, resulting in the ASM model-based methods (DCP and AOD) not being able to dehaze effectively at all. The performance of GCANet on dense-haze is unsatisfactory. Its outputs still contain hazy areas and suffer from serious color distortion. Although FFA and TDN are able to generate better results than above methods, we can still find obvious visual problems, such as low brightness and blurry borders. Our method also struggles in removing dense-haze.

4.4.4 Results on NH-Haze Dataset

Non-homogeneous haze in Figure 4.5 and Figure 4.6 also has a large impact on the ASM model-based methods. These methods (DCP and AOD) produce images in which not only the haze is not reduced but also suffer from severe color distortion. GCA
Figure 4.4: Qualitative evaluation on Dense-Haze.
|       | Hazy | DCP | AOD | GCA | FFA | TDN | TDN(NP) | Ours | GT |
|-------|------|-----|-----|-----|-----|-----|---------|------|----|
| Example Image 1 | ![Hazy Image](image1) | ![DCP Image](image2) | ![AOD Image](image3) | ![GCA Image](image4) | ![FFA Image](image5) | ![TDN Image](image6) | ![TDN(NP) Image](image7) | ![Ours Image](image8) | ![GT Image](image9) |
| Example Image 2 | ![Hazy Image](image1) | ![DCP Image](image2) | ![AOD Image](image3) | ![GCA Image](image4) | ![FFA Image](image5) | ![TDN Image](image6) | ![TDN(NP) Image](image7) | ![Ours Image](image8) | ![GT Image](image9) |
| Example Image 3 | ![Hazy Image](image1) | ![DCP Image](image2) | ![AOD Image](image3) | ![GCA Image](image4) | ![FFA Image](image5) | ![TDN Image](image6) | ![TDN(NP) Image](image7) | ![Ours Image](image8) | ![GT Image](image9) |

Figure 4.5: Qualitative evaluation on NH-Haze.
and FFA can remove most of the haze, but incomplete removal exists in areas with high haze concentration. TDN and our method are significantly better than the aforementioned networks in removing non-homogeneous haze.

### 4.4.5 Results on NH-Haze2 Dataset

The results on this dataset are generally similar to those of NH-Haze. It is worth mentioning that FFA outperformed TDN in both PSNR and SSIM. For the comparison between TDN and TDN without pre-trained weights, we can conclude that the ImageNet pre-training significantly improves the performance when using small-scale NH-Haze and NH-Haze2 datasets. This further proves that the importance of using ImageNet pre-training as initialization when working on small-scale datasets.

### 4.4.6 Results on NTIRE2021 Testing Set

During the NTIRE2021 dehazing challenge (Ancuti et al., 2021), we use NH-Haze 2020 as extra data. To eliminate the distribution shift between NTIRE2020 and NTIRE2021, we further employ gamma correction on NTIRE2020 with a hyperparameter of 0.65. Our results in NTIRE2021 test dataset are shown in Figure 4.7. The haze distribution in the original hazy images is non-homogeneous. Some areas have been turned purely white by the haze, while other parts can display the color of the grass field. In our dehazed images, we can see that in the first, second and last images, the haze is removed clearly. Although there is a small haze area remaining in the third and fourth images, the color and the outline of the leaves are clearly displayed. In the challenge report of NTIRE2021 (Ancuti et al., 2021), our dehazed results achieve 21.0183 and 0.8370 in PSNR and SSIM, respectively. This method
Figure 4.6: Qualitative evaluation on NH-Haze2.
Figure 4.7: Test results of NTIRE2021 challenge. The first column shows the original hazy images and the second column illustrates dehazing results using our proposed method.
ranked third and won the runner up award in the challenge

| Methods     | Parameters |
|-------------|------------|
| AOD         | 1.7K       |
| GCA         | 0.7M       |
| FFA         | 4M         |
| TDN         | 46M        |
| TL          | 49M        |
| CDF         | 1M         |
| Ours(TL + CDF) | 50M    |

Table 4.4: Parameters of each method.

### 4.5 Runtime Analysis

**Complexity Analysis.** We run this experiment on a same NVIDIA V100 GPU. Table 4.4 shows the total number of parameters and Figure 4.8 shows the runtime per image (size 1600 × 1200) of above methods. Although the transfer learning(TL) sub-net contains a large number of parameters, its processing speed is quite fast. The reason is that the down-sampling operations shrink the size of features in the middle layer, which speeds up the computation. In conclusion, except for AOD-Net, the runtime of our method is faster than that of other state-of-the-art methods even though we have the most number of parameters.
Figure 4.8: Runtime comparison of different methods.
Chapter 5

Conclusion

In this thesis, we have proposed a two-branch neural network for non-homogeneous dehazing via ensemble learning and proved its strong power in various dehazing tasks. To generate images with fine detail and color fidelity, we stack the features from transfer learning sub-net and current data fitting sub-net and then map them to haze-free images by fusion tail. Our method has a significant advantage in small-scale datasets. It surpasses many state-of-the-art methods in both real-world and synthetic datasets. Besides, we demonstrate the effectiveness of the ImageNet pre-trained model.
Bibliography

Ancuti, C., Ancuti, C. O., Timofte, R., Van Gool, L., Zhang, L., Yang, M., Patel, V. M., Zhang, H., Sindagi, V. A., Zhao, R., Ma, X., Qin, Y., Jia, L., Friedel, K., Ki, S., Sim, H., Choi, J., Kim, S., Seo, S., Kim, S., Kim, M., Mondal, R., Santra, S., Chanda, B., Liu, J., Mei, K., Li, J., Luyao, Fang, F., Jiang, A., Qu, X., Liu, T., Wang, P., Sun, B., Deng, J., Zhao, Y., Hong, M., Huang, J., Chen, Y., Chen, E., Yu, X., Wu, T., Genc, A., Engin, D., Ekenel, H. K., Liu, W., Tong, T., Li, G., Gao, Q., Li, Z., Tang, D., Chen, Y., Huo, Z., Alvarez-Gila, A., Galdran, A., Bria, A., Vazquez-Corral, J., Bertalmo, M., Demir, H. S., Adil, O. F., Phung, H. X., Jin, X., Chen, J., Shan, C., and Chen, Z. (2018a). Ntire 2018 challenge on image dehazing: Methods and results. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 1004–100410.

Ancuti, C. O., Ancuti, C., Timofte, R., and De Vleeschouwer, C. (2018b). O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 867–8678.

Ancuti, C. O., Ancuti, C., Sbert, M., and Timofte, R. (2019a). Dense haze: A
benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing (ICIP)*, IEEE ICIP 2019.

Ancuti, C. O., Ancuti, C., Timofte, R., Gool, L. V., Zhang, L., and Yang, M.-H. (2019b). Ntire 2019 image dehazing challenge report. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, IEEE CVPR 2019.

Ancuti, C. O., Ancuti, C., Timofte, R., Van Gool, L., Zhang, L., Yang, M., Guo, T., Li, X., Cherukuri, V., Monga, V., Jiang, H., Yang, S., Liu, Y., Qu, X., Wan, P., Park, D., Chun, S. Y., Hong, M., Huang, J., Chen, Y., Chen, S., Wang, B., Michelini, P. N., Liu, H., Zhu, D., Liu, J., Santra, S., Mondal, R., Chanda, B., Morales, P., Klinghoffer, T., Quan, L. M., Kim, Y., Liang, X., Li, R., Pan, J., Tang, J., Purohit, K., Suin, M., Rajagopalan, A. N., Schettini, R., Bianco, S., Piccoli, F., Cusano, C., Celona, L., Hwang, S., Ma, Y. S., Byun, H., Murala, S., Dudhane, A., Aulakh, H., Zheng, T., Zhang, T., Qin, W., Zhou, R., Wang, S., Tarel, J., Wang, C., and Wu, J. (2019). Ntire 2019 image dehazing challenge report. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2241–2253.

Ancuti, C. O., Ancuti, C., and Timofte, R. (2020a). NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, IEEE CVPR 2020.

Ancuti, C. O., Ancuti, C., Vasluianu, F.-A., Timofte, R., et al. (2020b). NTIRE 2020
challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, IEEE CVPR 2020.

Ancuti, C. O., Ancuti, C., Vashianu, F.-A., and Timofte, R. (2020c). Ntire 2020 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Ancuti, C. O., Ancuti, C., Vashianu, F.-A., Timofte, R., Liu, J., Wu, H., Xie, Y., Qu, Y., Ma, L., Huang, Z., Deng, Q., Chao, J.-C., Yang, T.-S., Chen, P.-W., Hsu, P.-M., Liao, T.-Y., Sun, C.-E., Wu, P.-Y., Do, J., Park, J., Kim, M., Metwaly, K., Li, X., Guo, T., Monga, V., Yu, M., Cherukuri, V., Chuang, S.-Y., Lin, T.-N., Lee, D., Chang, J., Wang, Z.-H., Chang, Y.-B., Lin, C.-H., Dong, Y., Zhou, H., Kong, X., Das, S. D., Dutta, S., Zhao, X., Ouyang, B., Estrada, D., Wang, M., Su, T., Chen, S., Sun, B., de Dravo, V. W., Yu, Z., Narang, P., Mehra, A., Raghunath, N., and Mandal, M. (2020d). Ntire 2020 challenge on nonhomogeneous dehazing.

Ancuti, C. O., Ancuti, C., Vashianu, F.-A., Timofte, R., *et al.* (2021). NTIRE 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.

Ayinde, B. O., Inanc, T., and Zurada, J. M. (2019). Regularizing deep neural networks by enhancing diversity in feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), 2650–2661.

Berman, D., Avidan, S., *et al.* (2016). Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682.

Brown, G. (2010). Ensemble learning. *Encyclopedia of machine learning*, 312, 15–19.
Cai, B., Xu, X., Jia, K., Qing, C., and Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing, 25*(11), 5187–5198.

Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., and Hua, G. (2019). Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience, 3*(3), 201–215.

Deng, J., Dong, W., Socher, R., Li, L., Kai Li, and Li Fei-Fei (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Deng, Q., Huang, Z., Tsai, C.-C., and Lin, C.-W. (2020). Hardgan: A haze-aware representation distillation gan for single image dehazing. In *European Conference on Computer Vision*, pages 722–738. Springer.

Dietterich, T. G. *et al.* (2002). Ensemble learning. *The handbook of brain theory and neural networks, 2*, 110–125.

Fattal, R. (2014). Dehazing using color-lines. *ACM transactions on graphics (TOG), 34*(1), 1–14.
Gao, S., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., and Torr, P. H. (2019). Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, **33**(12), 2341–2353.

He, K., Girshick, R., and Dollár, P. (2019). Rethinking imagenet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4918–4927.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Jacobs, R., Jordan, M., Nowlan, S., and Hinton, G. (1991). Adaptive mixtures of local experts, neural computation3 (1991), 79-87. *Google Scholar Google Scholar Digital Library Digital Library*.

Kornblith, S., Shlens, J., and Le, Q. V. (2019). Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, **521**(7553), 436–444.
Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4681–4690.

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. (2017). Aod-net: All-in-one dehazing network. In Proceedings of the IEEE international conference on computer vision, pages 4770–4778.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z. (2019). Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1), 492–505.

Liu, J., Wu, H., Xie, Y., Qu, Y., and Ma, L. (2020). Trident dehazing network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

Meng, G., Wang, Y., Duan, J., Xiang, S., and Pan, C. (2013). Efficient image dehazing with boundary constraint and contextual regularization. In Proceedings of the IEEE international conference on computer vision, pages 617–624.

Middleton, W. E. K. (1952). Vision through the atmosphere. University of Toronto Press.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345–1359.
Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318. PMLR.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch.

Qin, X., Wang, Z., Bai, Y., Xie, X., and Jia, H. (2020). Ffa-net: Feature fusion attention network for single image dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**(07), 11908–11915.

Qu, Y., Chen, Y., Huang, J., and Xie, Y. (2019). Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., and Yang, M.-H. (2018). Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261.

Shao, Y., Li, L., Ren, W., Gao, C., and Sang, N. (2020). Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2808–2817.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883.
Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning.

Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., and Tang, X. (2017). Residual attention network for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19.

Wu, H., Liu, J., Xie, Y., Qu, Y., and Ma, L. (2020). Knowledge transfer dehazing network for nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Zhang, H. and Patel, V. M. (2018). Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301.

Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2016). Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging, 3*(1), 47–57.
Zhou, Z., Shi, Z., Guo, M., Feng, Y., and Zhao, M. (2020). Cggan: A context guided generative adversarial network for single image dehazing.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232.

Zhu, Q., Mai, J., and Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing, 24(11), 3522–3533.