Deep Learning-based Image Bad Weather Removal

Zeyu Gao, Zeyu Li, Chen Liu, Runqiu Shi

Abstract

Many methods have achieved state-of-the-art performance on restoring images degraded by bad weather such as rain, fog and snow. However, they are designed specifically for a type of degradation. Recently, a transformer-based method with a single encoder-decoder pair was proposed to remove all the weather conditions at once. Unfortunately, this model fails to remove heavy rain in real world images. To this end, we propose 3 directions to solve the problem. First, we make modification to the attention block. Second, we introduce a cascaded method to improve the model performance on heavy rain removal task. Besides, we develop a novel model based on UNet and transformer to realize the All-in-One idea. Among all these approaches, the Global-enhanced method and cascaded model achieve significant improvement on heavy rain removal task.

1 Introduction

Traditionally, lots of applications are indoor which are not limited to defect detection, product assembly and scanning barcode in manufacturing, but X-Ray, CT and MRI in healthcare. This application environment is stable since it has relatively steady temperature, illumination, humidity and low atmospheric particles without being affected by weather change. Currently, the most popular computer vision use cases like surveillance systems and autonomous driving are outdoor applications. The weather conditions like rain, fog, and snow degrade the information from an image, which affects the following algorithms' performance in object detection and image segmentation. Therefore, bad weather removal on a single image is a non-trivial.

In the past, there are several methods exploring deraining, dehazing, raindrop removal and desnowing. The RNN-based PReNet(Ren et al., 2019) is a baseline deraining model which is for improving network efficiency. DM2F-Net(Deng et al., 2019) is introduced for Dehazing. This method is CNN-based and the purpose of it is to build a new multi-modal fusion approach that can explore multiple dehazing models to integrate respective strengths and maximize the whole performance. In Attn.GAN(Tao Xu, 2018), its contextual autoencoder structure makes image free from raindrops by introducing multi-scale loss and perceptual loss. The logic behind DesnowNet(Liu et al., 2018) is to restore lost information corrupted by opaque snow particle coverage.

Recently, (Li et al., 2020a) proposed a new deep neural network with All-in-One motivation. It takes all input images corrupted by all bad weather conditions including rain, fog, and snow. Thttps://www.overleaf.com/project/6274ba03d5fed7e2e0cc9c12his method achieves better and comparable performance than task-specific methods. This All-in-One method is a CNN-Based method with multiple encoders. Each encoder corresponds to different adverse weather condition types. This design is clever, but due to the multiple encoders, this network is not computationally efficient. Later on, there is a new transformer-based network coming out to tackle the key limitation of the previous model, called TransWeather(Valanarasu et al., 2021). This method is a single encoder-single decoder network to address all the bad weather conditions. Without using multiple encoders, it introduces the weather type queries in the decoder to make a judgment. But, TransWeather doesn't perform well at night and in heavy rain conditions.

After analysis, we decide to improve TransWeather's performance in heavy rain conditions. First, we modify the Transformer Attention Block by CBMA , Leff, Coordinate Attention and GE Transweather to adapt the model for heavy rain scenarios. Second, the cascade method is introduced. We simply put the dehazing module before the TransWeather and dehaze all inputs. In other

words, we use dehazing as a preprocessing step. Besides, we design a model based on UNet and Weather Type Query form TransWeather.

2 Related Work

2.1 Weather removal Problems

Single-task **image restoration** problems like deraining(Wei et al., 2019; Zhang and Patel, 2018),dehazing(Zhang et al., 2017; Cao et al., 2021b), desnowing(Zhang et al., 2021; Liu et al., 2018) and raindrop(Qian et al., 2018; Quan et al., 2019) removal tasks have been extensively studied in the literature.

Multi-Task weather removal: All-in-One Network (Li et al., 2020a)was proposed to handle multiple weather degradations using a single network. All-in-One designs multiple task-specific encoders, each of which is responsible for handling one particular type of background restoration task. It uses a discriminator to classify the degradation type and only backpropagates the loss to specific encoders. However, All-in-One is still designed in task oriented manner. Valanarasu et al. proposed a single encoder-single decoder transformer network called TransWeahter. Instead of using multiple encoders, they introduce weather type queries in the transformer decoder to learn the task. The model architecture is in Figure 9.

2.2 Vision Transformer

Since the introduction of Vision Transformer (Dosovitskiy et al., 2020) for visual recognition, transformer-based methods have been widely applied to various low-level vision tasks, including semantic segmantation(Strudel et al., 2021; Zheng et al., 2021), super resolution(Yang et al., 2020; Cao et al., 2021a; Kasem et al., 2019), and image restoration(Wang et al., 2021; Gou et al., 2020; Liang et al., 2021; Zamir et al., 2021). Besides, patch-embedding is also from this model; patch-embedding separate image into several sub-image, i.e patches, and extract features from each sub-image and forward them to attention block like eacch word in NLP task.

Additionally, many variations of transformer have been proposed to further improve its performance in computer vision tasks, including Swin-transformer(Liu et al., 2021), Cswin(Dong et al., 2021), PVT(Wang et al., 2022a).

2.3 UNet

Nowadays, UNet (Ronneberger et al., 2015) is a well-known architecture in a lot of low-level vision applications. It has hierarchical feature maps to gain the rich multi-scale contextual features. Additionally, the skip connection mechanism enhance the ability of the network to restore images. Due to the strong adaptive backbone, UNet can be easily applied with different extractive blocks to enhance the performance. Many recent papers(Zamir et al., 2021; Wang et al., 2021; Fan et al., 2022) focusing on image restoration adopt U-Net architecture and achieve the state-of-the-art performance.

3 Problem statement

Different weather phenomena degrade images with regard to different physics properties. For example, a heavy rain image is modeled as(Li et al., 2019):

$$\mathbf{I} = \mathbf{T} \odot (\mathbf{B} + \sum_{\mathbf{i}}^{\mathbf{n}} \mathbf{R_i}) + (1 - \mathbf{T}) \odot \mathbf{A} \quad (1)$$

where $\mathbf{I}$ is the degraded image, $\mathbf{T}$ is the transmission map, $\mathbf{A}$ is the global atmospheric light of the scene, $\mathbf{B}$ is the background image and $\mathbf{R_i}$ represents the rain streaks at the i-th layer.

Similarly, rain drop is modelled as (Qian et al., 2018):

$$\mathbf{I} = (1 - \mathbf{M}) \odot \mathbf{B} + \mathbf{R} \quad (2)$$

where $\mathbf{R}$ is the raindrop residual. According to (Liu et al., 2018), snow is generally modeled as:

$$\mathbf{I} = \mathbf{M} \odot \mathbf{S} + \mathbf{M} \odot (\mathbf{1} - \mathbf{z}) \quad (3)$$

where $\mathbf{M}$ is the binary mask.

Multi-task weather removal can be generalized as:

$$\mathbf{B} = \mathbf{D}(\mathbf{E}(\mathbf{I}_p)) \quad (4)$$

where $\mathbf{E}, \mathbf{D}$ corresponds to the encoder and decoder, p denotes the weather type in the image.

For this problem, we choose state-of-the-art model *TransWeather*(Valanarasu et al., 2021) as our backbone. After evaluating TransWeather's performance on real images, we observed that TransWeather can only perform limited removal on real images with heavy rain. Therefore, our goal is solve this problem while maintaining its performance on other weather removal tasks.

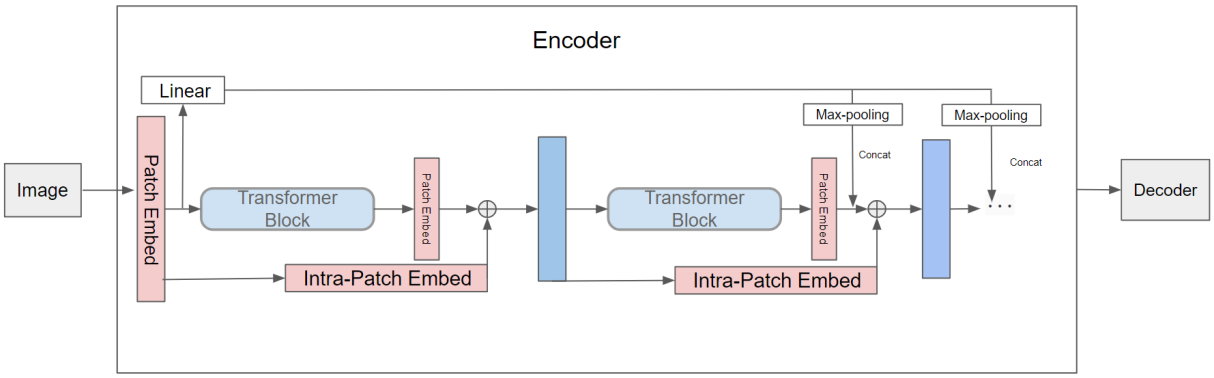

Figure 1: Architecture of GE TransWeather Encoder

4 Proposed Solution

In order not to dramatically increase the complexity of the model, we first try to solve the problem by directly modifying the architecture and subnets of TransWeather. After failures of first trying on LeFF(Wang et al., 2022b) and Coordinate Attention(Hou et al., 2021), we decide to extend our exploration into total three different directions: First, keep proposing methods on modifying TransWeather; Second, Cascade model based on TransWeather; Third, improving base on Unet model by using techniques from TransWeather. The purpose of direction extension is to hopefully have some positive feedback if we failed in all our proposed solutions in TransWeather modification.

4.1 TransWeather Modification

After observation of the real image evaluation, we discover that TransWeather is only capable of removing the rain streaks which are longer and closer to the camera. Based on the understanding of the network structure, we found that the last few layers patch embedding is using 3x3 kernel size which could be too small for the network to capture background information. In most case, a 3x3 kernel only have background that consist of single color and the only outstanding feature is the rain streak, so there will be too much attention paid on the rain streak instead of the background. We proposed 4 modifications on TransWeather to solve this predicted reason.

4.1.1 LeFF

Locally-enhanced Feed-Forward Network(LeFF) was proposed in (Wang et al., 2021) and is used to enhance local attention, we predict that this can help the model gather local information from background. We replace MLP with LeFF in TransWeather's Feed Forward block, and Figure 5 displays this block.

4.1.2 CBAM

Based on prior works (Yi et al., 2021; He et al., 2010), channel information plays an important role in weather removal tasks. This motivates us to investigate the effect of channel-wise attention in this problem. We believe this channel information can help improving performance of refining features, which can hopefully solve the rain removal problem, since the difference of each attention channel works differently for different weather type as showed on TransWeather paper. Therefore, we substitute Multi-head self attention with CBAM(Woo et al., 2018) which contains channel attention block that can gather inter-channel relationship information. Moreover, CBAM also uses spatial attention block which can gather information around the kernel using average-pooling and max-pooling. We predict this spatial attention can help the network focus more on background than the rain streaks. Figure 6 displays the architecture of CBAM.

4.1.3 CA TransWeather

Coordinate Attention(CA) mechanism, as an improvement of CBAM, was proposed in (Hou et al., 2021).The coordinate attention encodes both channel relationships and long-range dependencies with precise positional information. The diagram of the coordinate attention block can be found in the Figure 7. We predict the long-range dependencies can

help TransWeather gather more global information to better focusing on the background not rain streak. We substitute the the Multi-Head attention block in TransWeather with Coordinate Attention to further verify the effectiveness of Coordinate attention in the network.

4.1.4 GE TransWeather

This is our novel method called Global-enhanced TransWeather. Our idea is to keep the global features across the entire encoder. During first trying we pass the entire image input into a fully connect layer before patch embedding, but this lose every inductive bias and fail with very low performance. Second attempt is passing the first layer's embedding features to a fully connected layers and output a global feature. Then concatenate this feature to every following layer's embedding features, due to the feature size difference we also use max-pooling to extract features into certain size. The model structure is Figure 1.

4.2 Cascaded model

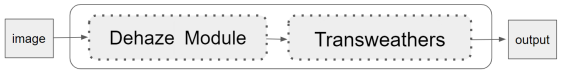

Figure 2: Cascaded model

Our next proposed method is motivated by the observations we had during our experiments. First, we found it is very common that real-world degradation images have a combining effect of fog and other weather types. An ideal All-in-One model then should be able to perform defogging and other weather removal operations simultaneously. Unfortunately, our previous experiments showed that TransWeather didn't dehaze properly on real images that have multiple degradation types. This is probably because TransWeather is trained on synthesized degraded images that do not have many important features the real images have. Second, the dense rain-streak accumulation on real-world pictures also led to fog-like effects that cannot be efficiently removed by TransWeather as shown in Figure 14. The reason is that individual rain streaks are so tiny and dense in these pictures that they are invisible to the rain removal model yet have many similar features to haze. Those observations inspire us to apply a dedicated dehazing module as a preprocessing step followed by the imperfect All-in-One model. In the proposed cascaded model, we put a pre-trained dehazing model, DehazeNet(Bolun Cai and Tao, 2016), before the TransWeather, as shown in Figure 2.

4.3 UNet-based Network

Many recent papers combines Vision Transformer and UNet architecture to solve image restoration tasks like denoising(Fan et al., 2022), deblurring(Zamir et al., 2021; Wang et al., 2021) and deraining(Zamir et al., 2021). Their success motivate us to design an UNet-Transformer based network for multi-task weather removal problem.

4.3.1 Network Architecture

Our main goal is to develop a model that can handle many different weather degradation at once. To achieve this, we borrow the Transformer Decoder Block from TransWeather, which serves as a feature selector according to different weather types. Furthermore, to alleviate the computational bottleneck, we adopt Swin-Transformer(Liu et al., 2021) as our backbone block. The architecture of the proposed model is illustrated in Figure 3.

Overall Pipeline: Given a degraded image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$. Our model first applies a convolution to obtain low-level features embeddings $F_{low} \in \mathbb{R}^{H\times W\times C}$; where C denotes the number of output channels. these features are fed into the a 4-level encoder-decoder and transformed into high-level and multi-scale features $F_{high} \in \mathbb{R}^{H\times W\times C}$. Each level of encoder-decoder contains multiple Swin-Transformer blocks.

Starting from the high resolution input, the encoder gradually reduces the size of feature maps, while expanding channel capacity. At the bottleneck, the decoder takes low resolution features $F_{bottle} \in \mathbb{R}^{\frac{H}{8}\times\frac{W}{8}\times 8C}$. as input and recovers the high resolution feature map. Next, the recovered features are further refined by the Transformer Decoder Block. Finally, we use a 3×3 convolution to generate the clean image $\hat{X} \in \mathbb{R}^{H\times W\times 3}$.

Loss Function: We optimize our model with the regular $L1$ pixel loss for image restoration.

$$\mathcal{L}_{restore} = ||\hat{X} - X||_1$$

5 Experiment Setting

5.1 Datasets

Training on real image is incapable in this topic, since we need both degraded image and ground truth image without weather effect for SSIM and PSNR calculation, and such pairs of images are almost unavailable online. In this case, same as previous work, we only use synthesised images

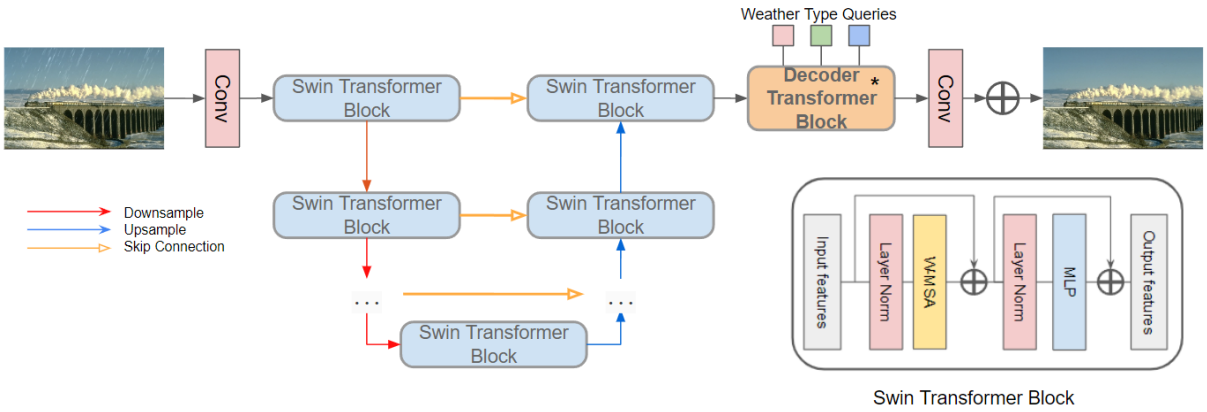

Figure 3: Network Architecture of UNet-based Network

in our training set and further evaluate it on real images. Because of the limitation of our computational resources, both our training and test dataset are sampled from the original dataset used in TransWeather. Following the same distribution, both training and test dataset consist of four weather effects: Rain, Snow, Rain drop and Fog. To further verify our proposed methods's performance on heavy rain removal, we use 64 real rain images collected online to do the evaluation.

5.2 Implementation detail

We train our models on Google Cloud Platform with one NVIDIA K80 GPU under the Pytorch framework for 300 epoch each.

5.3 Baseline

Since we are choosing TransWeather as our backbone, we also choose it as our baseline. To keep the validation of comparison we also retrain TranWeather on this dataset for 300 epoches, same as our proposed methods.

5.4 Metrics

Quantitative: Following previous works, we adopt two commonly used metrics to quantitatively measure the quality of the generated background images in our experiments: PSNR and SSIM.
PSNR denotes the peak signal-to-noise ratio in decibels between two images:

$$PSNR = 20\log_{10}\frac{MAX_I}{\sqrt{MSE}} \tag{5}$$

where the MAX_I denotes the maximum possible value of the ground truth image, and MSE denotes mean squared error between two images. It is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. In our task, signal is the ground truth image, noise is the Weather effect on image.
SSIM measures the perceptual similarity between two images:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

where $C_1 = (k_1 L)^2, C_2 = (k_2 L)^2$ are two variables to stabilize the division with weak denominator; L is the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 12^{\#bits\ per\ pixel} - 1$); $k_1 = 0.01$ and $k_2 = 0.03$. where the μ denotes mean, σ denotes variance, σ_{xy} denotes the covariance of x and y. SSIM focus more on structural information of two images. Structural information is the idea that the pixels have strong interdependencies especially when they are spatially close.
Noramlly higher value of PSNR or SSIM implies higher quality of the generated background image, and they generally have a positive correlation showed in Figure 4. In our experiments, we use average PSNR and average SSIM, which compute the mean of all test cases results.
Qualitative: Additionally, we also perform qualitative evaluation by visually evaluating the restored images of randomly selected test samples and real images. As a complement of the standard qualitative evaluation, we further perform survey evaluation that takes all of our evaluation results to 10 of our friends and asks them to tell which images remove more rain without telling them which image is produced by vanilla model or proposed model.

Methods	PSNR Average	SSIM Average	PSNR Variance	SSIM Variance
Cascaded model	24.43	0.8213	10.34	0.01
UNet-based Network	24.68	0.7787	11.56	0.01
Coordinate Attention	28.12	0.8179	6.92	0.01
LeFF	28.61	0.8251	9.03	0.01
TransWeather	30.10	0.8716	12.52	0.01
GE TransWeather	30.74	0.8806	11.87	0.01
CBAM	30.74	0.8912	13.81	0.01

Table 1: Quantitative results: Sort by PSNR Average

6 Results and Discussion

Tabel 1 is the Quantitative results from each model. From the PSNR variance of each model, it reveals that the dataset is not large enough to produce stable results. And we have not found any relation between PSNR variance with all other values yet. SSIM Variances are all 0.01 because our SSIM are too small to give meaningful feedback, which is also caused by the small dataset.

Tabel 2 is the rounded average Qualitative survey results, the left side of compare ratio denotes the number of proposed method's outputs are better than TransWeather's outputs, and the right side denotes the opposite way. And we do not count the images if survey taker cannot tell the difference, so the sum of these two total number is not 64. We do not include Unet in the survey since it's multi-task results failed in every case when we did the manual comparison by ourselves.

Please note, we only use a subset of original dataset adopted by TransWeather, and incomplete removal performance happens due to the lack of enough training data.

Methods	Compare ratio
LeFF	3/20
CBAM	7/5
Coordinate Attention	10/8
GE TransWeather	8/5
Cascaded model	20/2

Table 2: Qualitative survey result

6.1 TransWeather Modification

6.1.1 LeFF

LeFF does not do well in either Quantitative or Qualitative evaluation.

Takeaway: We implement this since at first we predict enhancing local attention can help the model gather more information from background. However, in contrast, since rain streak is already the most significant local attention in a 3x3 kernel. If we enhance the local attention again, the rain streak will get even more attention, which is the opposite way of our original predict.

6.1.2 CBAM

Quantitative: CBAM has the best average PSNR and average SSIM among all methods, which means it has the best restoration quality, i.e the least blurry restored outputs.

Qualitative: Survey takers also point out that CBAM have clearer restored images. However, even with the best restoration quality CBAM does not improve heavy rain removal compared to TransWeather.

Takeaway: It seems that the channel attention and spatial attention works better in image quality but it suffered same weakness as TransWeahter in heavy rain removal. The spatial attention gathered information around kernel does not push the model to pay more attention on background around the rain streak but helps restoration of clearer background.

6.1.3 CA TransWeather

Quantitative: Although Coordinate Attention improve the performance of the network in many other tasks, including object detection and semantic segmentation. For image restoration task, Coordinate attention fails to improve over CBAM.

Qualitative: The result of the survey is very different from other methods. This model have both high performance on certain cases and low performance on some other cases.

Takeaway: Survey takers' feedback helps us found that this model have strong advantage on dealing with images with both rain and fog, example Figure 8. It seems that its strength of channel relationship and long-range dependency helps this model gather more background information from the image and helps to remove the fog. However, it has disadvantage on removing rain streak, and the problem that

attention focus more on rain streak inside small kernel is not solved by its strength.

6.1.4 GE TransWeather

Quantitative: GE TransWeather has the best PSNR same as CBAM method and the second highest SSIM lower than CBAM, which means it still loses some quality during restoration process.
Qualitative: From survey results, GE TransWeather has a better performance on only a few amount of cases, but for most cases it's results do not have noticeable difference.
Takeaway: Base on previous observation, we only achieve limited improvement on rain removal problem. The reason of this limited improvement could be lacking of enough training set that this model does not show all of its capability. In failure cases, the TransWeather removes rain streak more completely than GE TransWeather, like in Figure 10. In previous example, small rain streaks are removed more in TranWeather which does not fit our predict, and further study needed in this method.
Ablation Study: Since CBAM has a smoother result and GE TransWeather has better rain removal performance but more blurry result, we expect that combining these two methods can overcome the weakness of GE TransWeather. We conduct a ablation study on a model combined GE TransWeather and CBAM. Results of this Ablation Study is in Table 5 and Table 6. Unfortunately it drops the performance in Quantitative evaluation, which results in more failure cases, but has similar success Qualitative cases with GE TransWeather.
Problem: We also sample the performance for updated weights after this model achieves more than 30 SSIM in Figure 13. This comparison reveals an unexpected behavior that with a relative low PSNR the model can still perform better on Qualitative evaluation. We have 3 predicts for this behavior.

- **First:** It is caused by the over-fitting problem since our training and test datasets do not have enough samples.
- **Second:** Encoder converges faster than decoder but encoder keeps update its weight while decoder process is still training, in this case we need self-supervised learning to first learn encoder then learn decoder or an automatic classifier to update encoder's weight if it achieves better Qualitative evaluation.
- **Third:** The way of saving weight could be questionable, we follow same rule as TransWeather that only keep the weight when the model achieves better PSNR during the Training. However, from the potential trade-off we found between Qualitative performance and Quantitative values, a better update rule or loss function that can reveal the weather effect instead of barely PSNR need to implemented.

6.2 Cascaded model

Quantitative: The cascaded model yields relatively poor PSNR and SSIM compared to other methods, because the degraded image goes through more smoothing operations and loses more local details such as brightness and local dominant color. However, by using Google Vision Classification as the extra evaluation metric, we prove that images restored by our cascaded model successfully keep all structural information and actually have a better visual image quality.
Google Vision Classification: In order to evaluate how the cascaded model effects meaningful information of the images, we adopted the pre-trained Image Label Annotation model from Google Cloud Vision API as a sample downstream machine learning task. The average confidence scores for bad weather classes ("Rain", "Fog" or "Haze") are 0.65 and 0.53 on the same real image dataset processed by TransWeather and the cascaded model, respectively. The result shows that the pre-trained classifier is less confident of bad weather happening in the scene when images are processed by the cascaded model than TransWeather. To further justify our method, we also calculated how many mistakes the annotation model made on a dataset of 34 real-world images (10 labels generated per image). On average, the model made 11% less mistakes when the images are processed by the cascaded model than TransWeather, as shown in Table 7 and Figure 14.
Qualitative: In the qualitative survey, the cascaded model gives better results than the TransWeather in most cases, since it is actually a boosted TransWeather which contains an entire TransWeather module. Sometimes this boosting effect may result in an unacceptable distortion so the TransWeather wins in a few cases in the survey.

6.3 UNet-based Network

Since the goal of this direction is to improve Unet using Weather type query from TransWeather, our experiments on this direction is slightly different

from previous two directions. To evaluate the effectiveness of UNet-based Network, we conducted extensive experiments following the same setting as TransWeather. First, we compare this model with TransWeather on single-task removal including deraining, desnowing and raindrop removal. Additionally, we compare this model with TransWeather on our sampled all-in-one dataset. We also evaluate the performance of this model on real heavy rain images.

6.3.1 Results on Single-task removal

Following TransWeather, for single-task removal, the Transformer decoder block is not included in the model for it is used to classify the weather types and select corresponding features. Table 3 shows the quantitative results of our method on single-task weather removal.

On Rain streak removal task, our proposed method achieves relatively good results. However, on snow removal and rain drop removal task, there is a significant gap between our method and TransWeather in terms of PSNR and SSIM.

Dataset	TransWeather	Unet-based
Rain100L	33.67/0.9609	31.07/0.9504
SnowTest100k	33.78/0.9278	24.29/0.7997
RainDrop	34.55/0.9602	23.47/0.7860

Table 3: Quantitative comparison on each removal task based on average PSNR/average SSIM

6.3.2 Results on Multi-task removal

Table 3 shows the quantitative results of our method on multi-task weather removal. It can be noted that our model has a very poor performance compared to TransWeather in terms of PSNR and SSIM.

Figure 11 shows that this method can remove the weather effect to a certain level. However, even if there is a relatively good performance on rain streak removal task when the model is trained on Rain100L dataset, it is still not able to remove the rain streak thoroughly when trained on dataset contained multiple weather type.

6.3.3 Results on Real heavy rain images

As Figure 12 demonstrates, this method suffers the same disadvantage as the TransWeather. When there is high intensity of rain presented in the images, our method is only able to perform limit removal of rain streaks.

6.3.4 Limitations

After further analysis of Unet-based TransWeather, we conclude it mainly suffers from 4 limitations.
Model Complexity: Even though we adopt Swin-Transformer block in this method to reduce the computational complexity, the whole network is rather complex compared to other multi-task weather removal methods.

Method	GMACs
All-in-One(Li et al., 2020b)	12.26
TransWeather(Valanarasu et al., 2021)	6.12
UNet-based Network	33.08

Table 4: **Comparison of GMACs(fixed-point multiplyaccumulate operations performed per second)** UNet-based Network is much more computational complex compared to recent methods.

Throughout the project, the model takes a huge amount of time for training, and this is the bottleneck when we try to tune the model.
Feature Selection: Essentially, current methods for multi-task weather removal involve soft classification of the weather types using attention mechanism, then restore original image based on these weighted features. As the results demonstrate, the technique used in TransWeather does not work well in our method. Similar technique is also proposed in (Wang et al., 2022b) which is also a UNet-Transformer based model, however, the detailed implementation is not illustrated in the paper.
Upsampling Technique: As shown in Figure 11, this model is not able to reconstruct the details in the original image, might because of the upsampling technique. In this model, we use bilinear interpolation to upsample the images. However, bilinear interpolation is deterministic which might cause the loss of structural information. We need to explore different upsampling techniques like transpose convolution.
Backbone Selection: We choose UNet as backbone of this direction because current UNet-Transformer models achieves state-of-the-art performance on image denoising and image deblurring. However, the results show that UNet-Transformer based models might not be a good choice for multi-task weather removal tasks. The relation between each weather removal task, deblurring and denoising still need to be further examined.

7 Conclusion and Future work

Compared to TransWeather, each of our methods has its own strength, weakness and unsolved new problems as we showed above. If we only consider the improvement of PSNR and SSIM, we do improve TransWeather in a limited scope. However, we believe the downstream task on real images are the most important part, where we do not achieve a remarkable improvement. Thus, we do not achieve our goal to completely improve the TransWeather, and further research is needed.

As aforementioned, the datasets used by all the weather removal methods is synthesised, which introduces unavoidable artifacts in the images. To this end, using deep learning techniques to add realistic weather effects to the image still remains unexplored. We believe this can be of great importance for removing weather effect in real-world images.

By the way, contribution is in the last page.

References

Kui Jia Chunmei Qing Bolun Cai, Xiangmin Xu and Dacheng Tao. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198.

Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. 2021a. Video super-resolution transformer. *arXiv preprint arXiv:2106.06847*.

Zhiwei Cao, Yong Qin, Limin Jia, Zhengyu Xie, Qinghong Liu, Xiaoping Ma, and Chongchong Yu. 2021b. Haze removal of railway monitoring images using multi-scale residual network. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7460–7473.

Zijun Deng, Lei Zhu, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Qing Zhang, Jing Qin, and Pheng-Ann Heng. 2019. Deep multi-model fusion for single-image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2453–2462.

Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. 2021. Cswin transformer: A general vision transformer backbone with cross-shaped windows. *arXiv preprint arXiv:2107.00652*.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu. 2022. Sunet: Swin transformer unet for image denoising. *arXiv preprint arXiv:2202.14009*.

Yuanbiao Gou, Boyun Li, Zitao Liu, Songfan Yang, and Xi Peng. 2020. Clearer: Multi-scale neural architecture search for image restoration. *Advances in Neural Information Processing Systems*, 33:17129–17140.

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

Qibin Hou, Daquan Zhou, and Jiashi Feng. 2021. Coordinate attention for efficient mobile network design. In *CVPR*.

Hossam M Kasem, Kwok-Wai Hung, and Jianmin Jiang. 2019. Spatial transformer generative adversarial network for robust image super-resolution. *IEEE Access*, 7:182993–183009.

Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1633–1642.

Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. 2020a. All in one bad weather removal using architectural search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3172–3182.

Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. 2020b. All in one bad weather removal using architectural search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3175–3185.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844.

Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. 2018. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022.

Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491.

Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. 2019. Deep learning for seeing through window with raindrops. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2463–2471.

Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Robin Strudel, Ricardo Garcia, Ivan Laptev, and Cordelia Schmid. 2021. Segmenter: Transformer for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7262–7272.

Qiuyuan Huang Han Zhang Zhe Gan Xiaolei Huang Xiaodong He Tao Xu, Pengchuan Zhang. 2018. Attngan: Fine-grained text to image generation with attentional generative adversarial networks.

Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. 2021. Transweather: Transformer-based restoration of images degraded by adverse weather conditions.

Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. 2022a. Pvt v2: Improved baselines with pyramid vision transformer. *Computational Visual Media*, pages 1–10.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. 2021. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint arXiv:2106.03106*.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. 2022b. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. 2019. Semi-supervised transfer learning for image rain removal. In *The IEEE Conference on Computer Vision and Pattern Recognition*.

Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19.

Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. 2020. Learning texture transformer network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5791–5800.

Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tieyong Zeng. 2021. Structure-preserving deraining with residue channel prior guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4238–4247.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2021. Restormer: Efficient transformer for high-resolution image restoration. *arXiv preprint arXiv:2111.09881*.

He Zhang and Vishal M Patel. 2018. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*.

He Zhang, Vishwanath Sindagi, and Vishal Patel. 2017. Joint transmission map estimation and dehazing using deep networks. *IEEE Transactions on Circuits and Systems for Video Technology*, PP.

Kaihao Zhang, Rongqing Li, Yanjiang Yu, Wenhan Luo, and Changsheng Li. 2021. Deep dense multi-scale network for snow removal using semantic and depth priors. *IEEE Transactions on Image Processing*, 30:7419–7431.

Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. 2021. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890.

A Appendices

We tried to format this appendix section, but the figures and tables position is automatically generated.

A.1 Figures

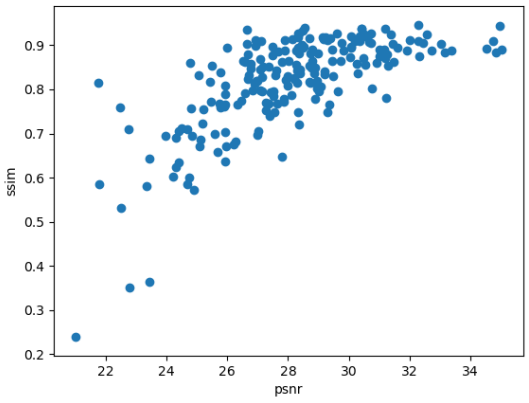

Figure 4: Correlation between PSNR and SSIM

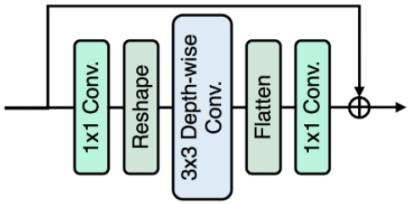

Figure 5: Architecture of LeFF

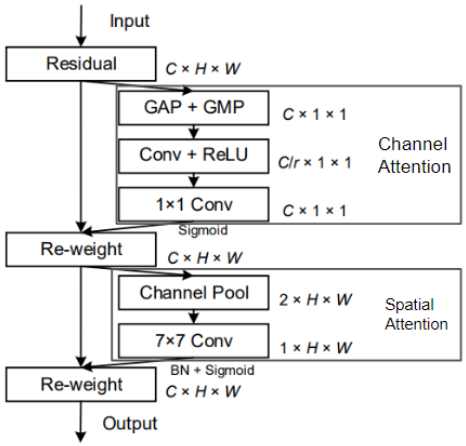

Figure 6: Architecture of CBAM

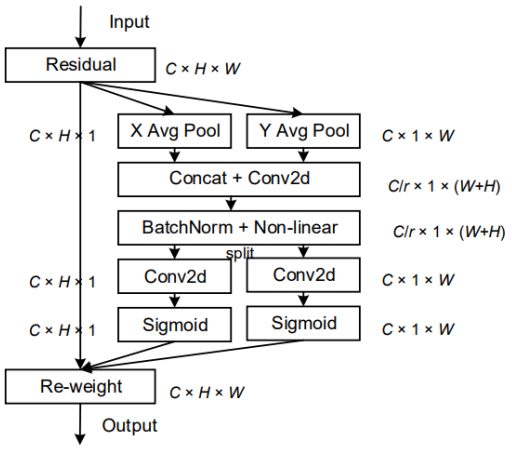

Figure 7: Architecture of Coordinate Attention

Figure 8: Left: TransWeather; Right: CA TransWeathe;

Quantitative results	GE	GE + CBAM
PSNR Average	29.26	30.10
SSIM Average	0.8716	0.8335
PSNR Variance	12.52	8.33
SSIM Variance	0.01	0.01

Table 5: Quantitative results of GE ablation study

Methods	Compare ratio
GE TransWeather	8/4
GE TransWeather + CBAM	7/8

Table 6: Qualitative survey of GE ablation study

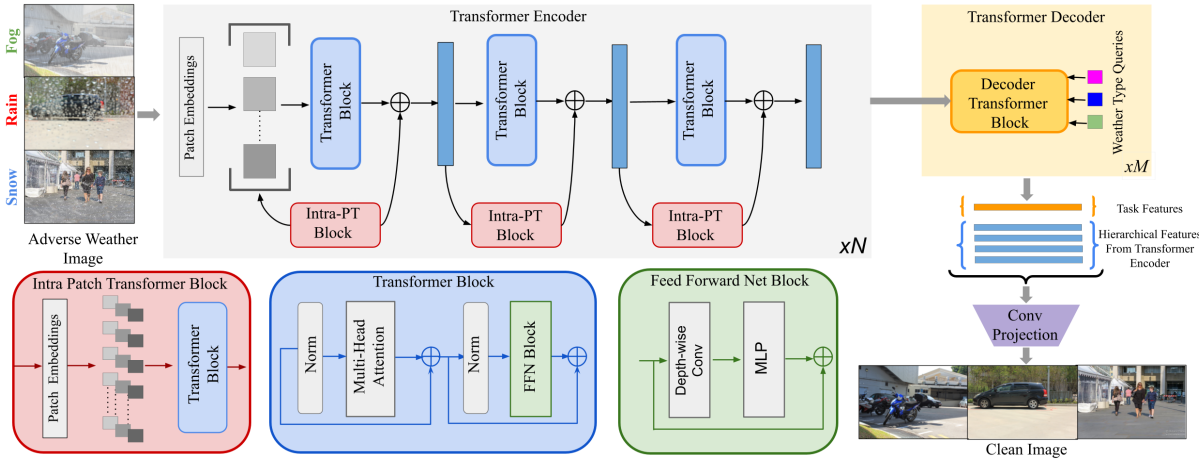

Figure 9: Network Architecture of Transweather: detailed explanation is on our project survey report

Figure 10: Left: GE TransWeather; Right: TransWeather

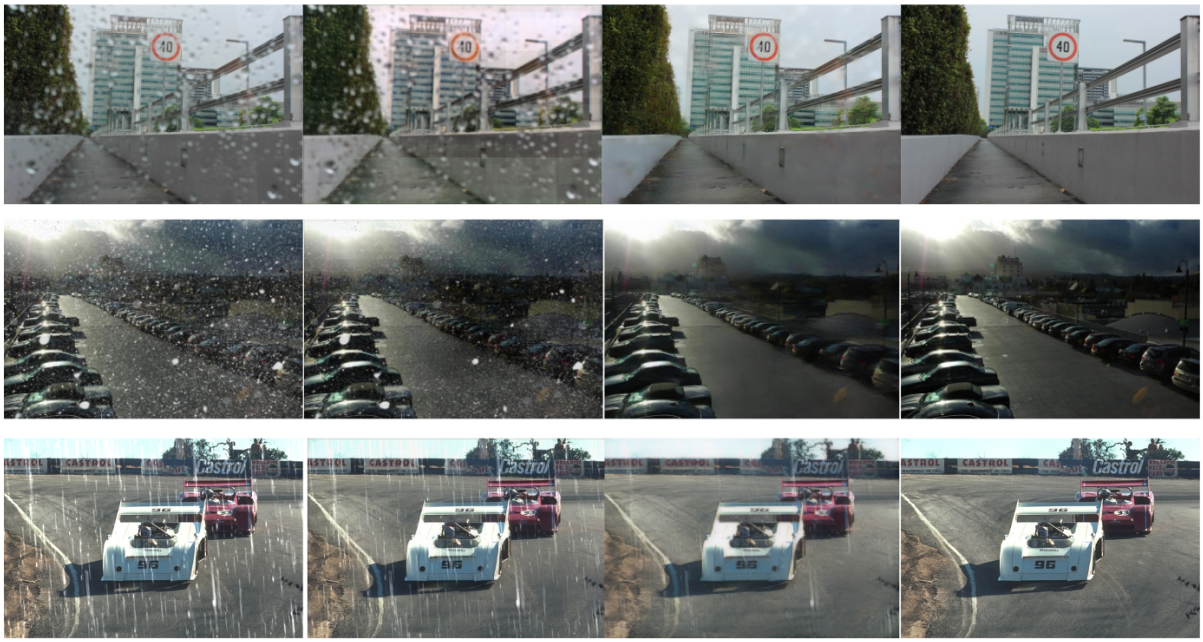

Figure 11: From left to right: Input Images, removal results of Unet-based method, removal results of TransWeather, Groundtruth images.

Figure 12: From left to right: Input Images, removal results of Unet-based, removal results of TransWeather. It can be noted that both methods cannot perform well when there is a high intensity of rain in the images

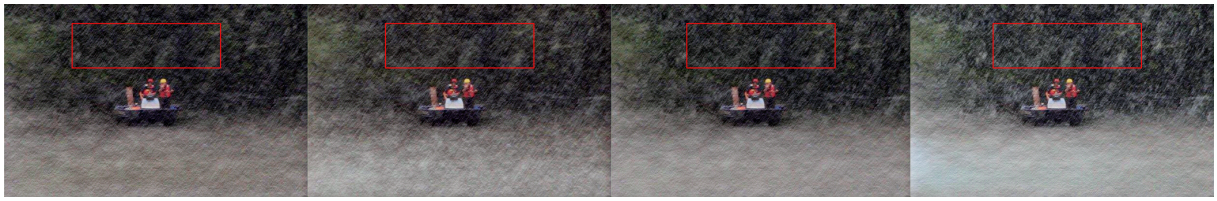

Figure 13: PSNR from left to right: 30.04, 30.10, 30.27, 30.74. Need zoom-in to watch detail

Figure 14: From left to right: Input Images, removal results of TransWeather, removal results of the cascaded model. The foglike dense rain streaks are successfully removed in the cascaded model.

Cascaded model	TransWeather
Water:0.96	Water:0.94
Natural landscape:0.87	Natural landscape:0.84
Asphalt:0.84	Atmospheric phenomenon:0.83
Road Surface:0.82	Grass:0.82
Grass:0.81	Asphalt:0.80
Groundcover:0.78	Automotive tire: 0.80
Landscape:0.75	Wood:0.80
Road:0.70	Woody plant:0.80
Soil:0.68	Groundcover:0.78

Table 7: Downstream Label Annotation results of the images from Figure 14. Format as Label:Score. Automotive tire is obvious wrong in TransWeather, but all labels in Cascaded model still make sense.

A.2 Contribution

Chen Liu: Part of all reports and presentations. Decide approach 2 based on other paper research. Research and implement CBAM. Implement Cascaded Models based on DehazeNet.
Zeyu Li: Part of all reports and presentations. Decide approach 3 based on other paper research. Research and implement Coordinate Attention. Implement Unet based on TransWeather.
Runqiu Shi: Part of all presentations and reports. Major contributor of final presentation. Real image dataset collection. Research on the rain removal in night, raindrop and Inpainting method in weather removal.
Zeyu Gao: Pre-research of deciding Backbone. Part of all reports and presentations. Further Experiments of training different datasets on Transweather. Decide approach 1 directions based on other paper research. Research and implement LeFF. Implement two ways of Global-enhanced Transweather.