1.1 Motivation and Contribution:

Most of existing de-raining frameworks suffer from being overfitted to only some specific rain types covered by the synthetic training set, and usually fail to deal with rainy images with more diversified rainy effects. To improve this and construct a more adaptive de-raining model, cascaded de-raining network is formulated such that images with light rains can be processed by fewer blocks while images with heavy rains can be tacked by more blocks in a progressive manner. Despite the improved performance, the effect of each block is unclear and over/under -deraining results are often produced (see Fig.3 of the paper) by existing progressive d-raining models. To find out the reason and get insight on how each block performs de-raining, we propose a series of novel convolutional filter behavior analysis tools (FBAT), which are then used to successfully find out the problems of existing progressive de-raining models. Motivated by our findings, a completely new feature refinement framework is formulated by using the learned rain location information to guide necessary manipulation on rainy regions and avoid over-manipulation on non-rainy regions. By implementing the joint location guidance and feature refinement process with a lightweight two-branch encoder network, an adaptive and explainable de-raining network, AFR-Net, is constructed and demonstrated to be able to (1) deal with different rainy conditions without the artefacts of over/under -deraining, (2) generalize well to unseen examples, and (3) most importantly perform well on real rainy images. What's more, (1) the proposed FBAT can be directly used to explain the behavior of other networks tailored to solve different CV problems, and (2) the proposed AFR-Net has also been successfully used to produce SOTA result on different benchmarks such as the single image denoising task, landmark guided face manipulation task and keypoint guided pose transfer task. The FBAT, AFR-Net, and the additional experimental results will be provided after paper acceptance.

1.2 Results on more complex rainy conditions:

To be consistent with most of existing de-raining models which only report the results on either raindrop removal or rainstreak removal, we also only tested AFR-Net on these two tasks and new SOTA results are obtained on benchmark datasets. For more results on other conditions, the explanations are as follows:

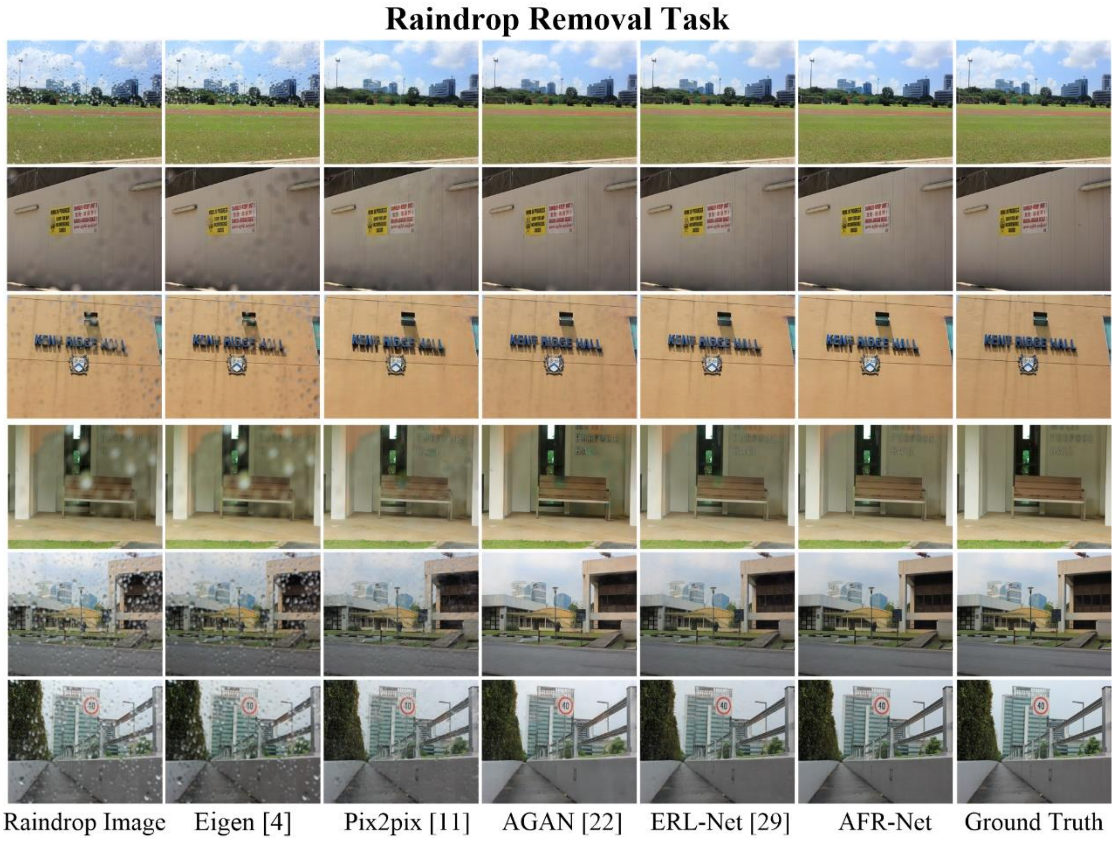
(1) For results on images with veiling effect, we have trained and tested the proposed model on the NYU-Rain dataset, and the AFR-Net with 8 refinement modules achieve new SOTA results with PSNR/SSIM as 21.70/0.861, outperforming the current SOTA by a small margin of 0.14/0.006 on PSNR and SSIM respectively;

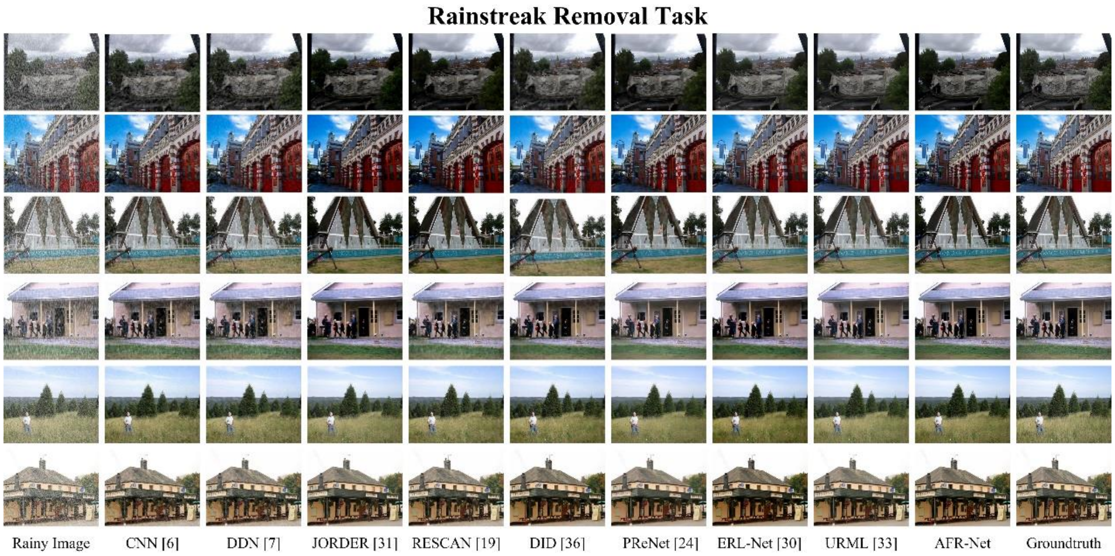
(2) For results on images with combinational effect including both raindrops and rainstreak, we leave this for future work because there is not any datasets covering both effect, and there is no existing de-raining models reporting results on such rainy conditions. We are now using the AGAN-Data to synthesize such effect, and we will use AFR-Net trained on both RS-Data and RD-Data to produce results on our synthesized dataset.

All results and related datasets obtained under the above two scenarios will be added to the paper in a future version, and also released on a GitHub page after paper acceptance.

1.3 Explanation on unconvincing results:

Quantitative results are better than all the other comparative models on both raindrop and rainstreak removal tasks. Superiority of qualitative results cannot be clearly recognized because the images of the paper are small due to space limitation. For better visual comparison, we provided very large de-raining results with clear details in the supplementary material. Besides, we provide more visual examples as in Fig. R1 for better showing the superiority of our de-raining models.





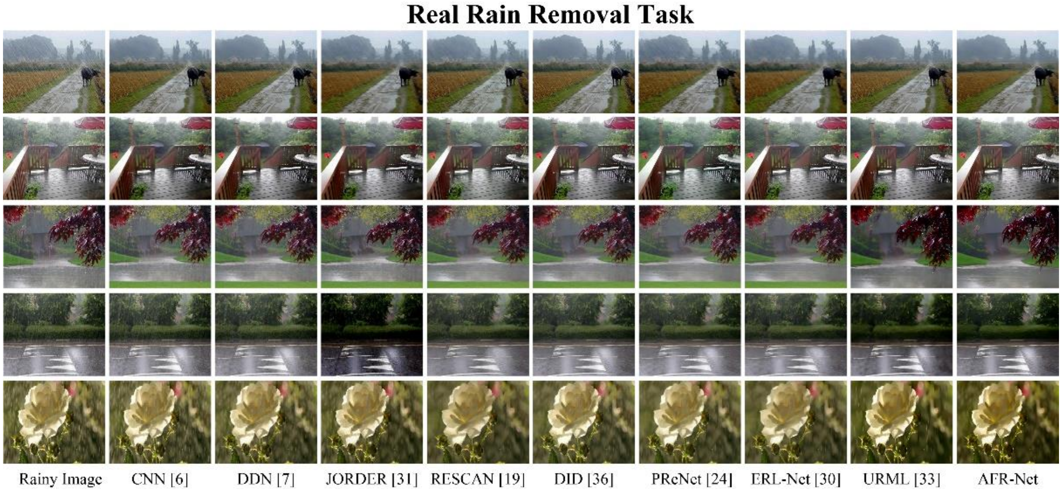
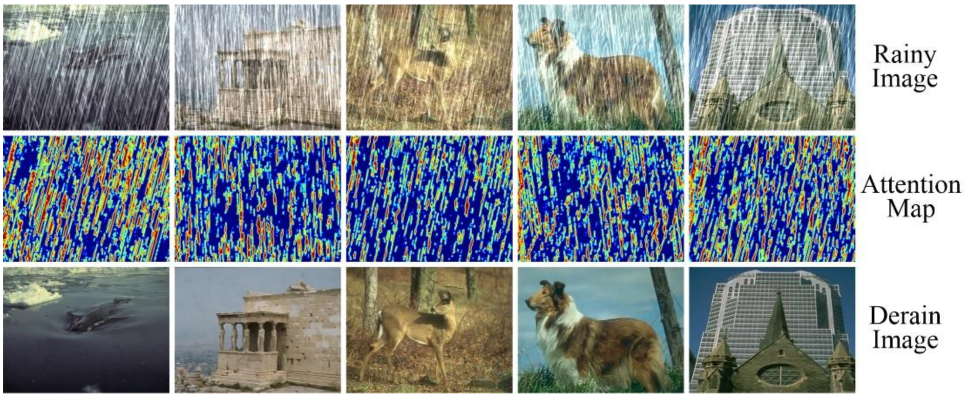


Fig. R1 More visual comparisons on different rainy datasets.

Response to specific comments (#1):

Q1: Rain100H is a benchmark dataset for heavy rainstreak removal task. We have tried AFR-Net on Rain100H, the attention maps are shown in Fig. R2. Besides, both quantitative and qualitative comparisons with other models (as mentioned in Table 3 of the paper) are carried out, and AFR-Net outperform the competitive models by a large margin of 1.28/0.016 on PSNR/SSIM respectively. Both quantitative and qualitative comparisons are provided in Fig. R2 for your reference.



Quantitative Results

		DSC	GMM	CNN	DDN	JORDER	RESCAN	DID-MDN	PReNet	ERL-Net	URML	AFR-Net
Rain100H	PSNR	14.63	15.05	21.53	21.92	26.54	28.88	28.37	29.46	29.58	29.36	30.86
	SSIM	0.423	0.431	0.684	0.764	0.835	0.866	0.842	0.899	0.912	0.908	0.928

Qualitative Results

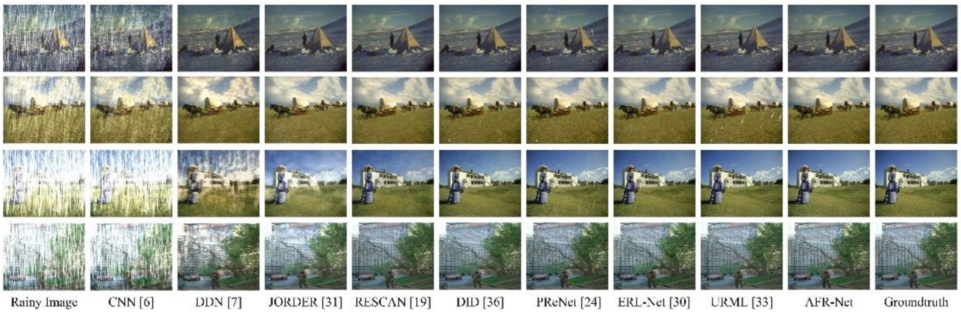


Fig. R2 Experimental results of AFR-Net on Rain100H dataset.

Q2: Attention maps have been widely explored by existing de-raining frameworks (e.g., AGAN and URML), however our design differs from theirs on the following aspects: (1) The motivation, networks structure, and the function of attention map are different from the other designs, and the results that our design outperforms AGAN/URML quantitatively and qualitatively verify that our design provides a much better solution on using attention maps; (2) Our framework is very general, and by replacing the attention map as keypoints map, our models can be directly used for keypoint guided pose transfer and landmark guided face manipulation. In contrast, the other designs can only be applied for de-raining task. See Sec1.1 for clearer explanation on the contributions of our design.

Response to specific comments (#2):

Q1: The expected proportion should show like: F_R plays a dominant role for de-raining task thus the proportion of F_R should be much larger than that of F_{NR} at each stage. Only in this way, sufficient manipulation over rainy regions can be guaranteed and unnecessary manipulation over non-rainy regions can be avoided. To demonstrate this, we train an AFR-Net with three refinement modules and calculate the proportion of both F_R and F_{NR} at each refinement stage. Direct comparison of proportion distributions of AFR-Net with the one in Fig.4 of the paper are provided as in Table R1.

Table R1: Comparison of filter proportion at different stages of different models.

	Raindrop Removal Task						Rainstreak Removal Task					
	Stage-I		Stage-II		Stage-III		Stage-I		Stage-II		Stage-III	
	F_R	F_{NR}	F_R	F_{NR}	F_R	F_{NR}	F_R	F_{NR}	F_R	F_{NR}	F_R	F_{NR}
PReNet	58%	42%	39%	61%	30%	70%	86%	14%	58%	42%	46%	54%
AFR-Net	87%	13%	80%	20%	79%	21%	90%	10%	84%	16%	82%	18%

As shown in Table R1 and as expected, F_R shows a much larger proportion than F_{NR} regardless of the stages. When performing de-raining by models with such filter proportion distribution, unexpected phenomenon of over/under deraining can be avoided as shown in Fig. 3 of the paper. Furthermore, by removing the attention branch, the proposed model shows a similar filter proportion distribution as PReNet, fully demonstrating the effectiveness of our novel attentive feature refinement design on overcoming problems of the progressive de-raining model design. Please see Sec1.1 for a clearer analysis on the contributions of our design.

Q2: Our design has nothing to do with the assumption of "a rainy image or a coarse-refined result should contain balanced region distribution". For both heavy and light rainy condition, the proportion of F_R and F_{NR} can be kept reasonable as the one in Table R1, thus producing SOTA results on different rainy conditions as demonstrated by visual results in the paper and supplementary material.

By carrying out deeper analysis on the behavior of different stages when dealing with heavy and light rainy condition, it was found that the proportion of filter with very ignorable response was large at later stage when dealing with light rainy condition, indicating that these filters can be removed by model compression strategy. In the future, we will base on AFR-Net to design a rainy-condition adaptive de-raining strategy. By this way, adaptive number of stages will be activated when dealing with images with different conditions, thus reducing the time consumption during inference.

Q3: The datasets used for results in Fig.5 is the same with the ones used in ablation study, and we will add the description in a future version.

Q4: The biggest difference between our design and existing network architecture is that: most of the other architecture are black-box, and they are designed empirically and runs in a blind way. However, the construction of our model is guided by our findings on analyzing the filter behavior, and we for the first time provide an explainable de-raining network design. Besides, see Sec1.1 and our response to R1 for explanation of other difference.

Q5: For the ablation study, FRB and AGB in each AFR module have been designed with lightweight convolution, and cannot be modified into smaller structure. To answer the concern that performance improvement may come from more parameters than the baseline, we try to modify the baseline to a model with similar parameters as the AFR-Net. Specifically, for models in Table 1 of the paper, we modify the baseline as the combination of two encoders (one as replacement of AGB1), one FRB, and one decoder. The modified model (denoted as BaseR) shows nearly the same complexity as AFR-Net with one refinement module. However, the PSNR/SSIM on RS-Data is 26.35/0.907, on RD-Data is 29.11/0.908, and the results are much worse than AFR-B as described in Table 1 of the paper. Furthermore, the same strategy is used to modify the baseline model to be the same complexity as AFR-Net with different numbers of AFR modules as in Table 2, and the results are similar that the modified baseline models are outperformed by the AFR-Net variants by a large margin. Such results indicate that the performance improvement is not from the more parameters but from the novel attention guided refinement design. The above results and analysis will be added to the paper in a future version.

Q6: For the knowledge distillation setup, owing to the extra supervision by the feature level loss introduced from the pre-trained teacher encoder, the student encoder can be trained to approximate the behavior of the highly non-linear semantic space learned by a large amount of parameters in the teacher model, and then the representation from the student encoder can be used by the decoder shared from the teacher model to produce satisfactory results. However, for the teacher model which can only be trained by the reconstruction loss, such a highly non-linear representation space can only be modelled with encoder consisting of a large number of parameters. See the supplementary material for a cleared description on the knowledge distillation setup.

Q7: Better quantitative results are obtained by deeper layers (AFR-Net with more refinement modules) is understandable because more parameters are needed to model the complex rain distribution for images with heavier rainy conditions. See previous response for more detailed explanation on the difference of dealing with rainy images with different rainy conditions.

Q8: “guidance information” in L98 refers to density label or rain detection map, “unsatisfactory results” in L319 and “issues in existing refinement network” in L312 refers to de-raining outputs with artefacts of over/under de-raining, “prior assumption” in L320 refers to the low-rank or sparsity assumption for solving a optimization problem. All the unclear and confusing statements will be revised in a future version.

Q9: Different rainy regions of raindrop images may show different property, thus can be

processed by models with a series of stages, with each stage being tailored for a certain region containing raindrops of similar distribution. See previous response on explaining how AFR-Net deals with images with different rainy conditions. Furthermore, unexpected light reflection/refraction when transmitting through raindrops may also cause some artefacts. Images with such artefacts and the raindrops may need a cascaded structure for artefact removal.
Q10: The size of visual results is kept small due to space limitation. See supplementary material and here for results with larger size.

Response to specific comments (#3):

Thank you for your positive comments. Please see Sec1 and our responses to other reviewers for a clearer explanation on our design.