Supplementary Material:
ERL-Net: Entangled Representation Learning for Single Image De-Raining

Guoqing Wang[1,2], Changming Sun[2,1] and Arcot Sowmya[1]
[1]University of New South Wales, Australia, [2]CSIRO, Data61, Australia
{guoqing.wang,changming.sun}@csiro.au, a.sowmya@unsw.edu.au

The supplementary material consists of:

1. Analysis on the working mechanism of ERL-Net.
2. Illustrating the detailed procedures to search the hard and easy examples.
3. Providing much more de-raining results and corresponding analysis on both synthetic and real rainy images.
4. Providing comparative analysis on the computational efficiency of ERL-Net.
5. Investigating the generalization ability of ERL-Net.

1. More clear understanding of ERL-Net

To obtain insight on what features are learned by both the main branch and the residual branch and also how the proposed simple entanglement of them achieves better de-raining results, we visualize the feature maps in different branches as shown in Fig. S1:

As can be seen from Fig. S1 and as expected, many background related patterns are learned by the residual branch, which plays a complementary role in providing the background information missed by the main branch (this can be especially observed by the feature maps in the red rectangle regions of Fig. S1). By combining the main branch and the residual branch via the entanglement manner, more complete encoding of the important background patterns is obtained, thus resulting in much better de-raining results with most of the background information being recovered well.

2. Illustration on how to determine hard and easy samples

In Section 4.3 of the paper, we have demonstrated the effectiveness of ERL-Net by estimating the statistical distribution of the de-raining results on both hard and easy samples. Here we simply describe how the hard and easy samples are selected from the whole testing dataset.

As shown in Fig. S2, two stages are involved for determining the subset of hard samples and subset of easy samples:

Stage-I: In the first stage, the baseline model (MEMD in the paper and the network structure is shown in Fig. S2) is trained and then tested to obtain the average PSNR value;

Stage-II: In this stage, the specific PSNR value of each sample in the testing set is calculated with the trained MEMD model. Then the sample in the testing set is determined as either hard sample or easy sample by comparing the specific PSNR with a threshold, which is defined as the average PSNR value obtained in the first stage.

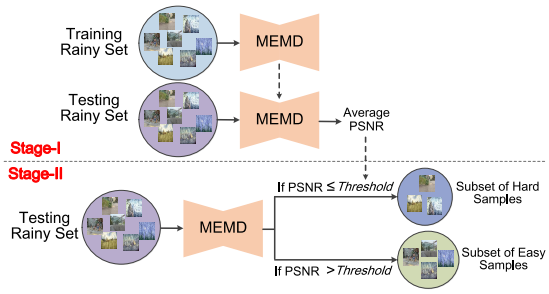

Figure S2: Demonstration of the hard and easy samples selection mechanism.

After figuring out the hard and easy samples subset, the improvement achieved by ERL-Net over the simple baseline is investigated in each subset respectively, thus providing a deeper analysis on the effectiveness of ERL-Net.

3. More de-raining results on synthetic and real rainy images

In the paper, some visual results are shown in Fig. 4, Fig. 5, and Fig. 6 for understanding the effect of ERL-Net on different datasets. In this section, more visual results are provided for better understanding on the effectiveness of the proposed model.

As evidenced by the additional de-raining results on the synthetic rainy datasets shown in Fig. S3, Fig. S4, and Fig. S5, the proposed model produces much better results with thorough raindrops/rain streaks removal and nearly perfect detail recovery [1].

For the real image rain streak removal task, as shown in Fig. S6, very promising results on real rainy images are obtained by ERL-Net, which can on the one hand remove the rain streaks thoroughly, and on the other hand recover the detailed structures well with high contrast (*e.g.*, the results in the 1^{st}/2^{nd}/5^{th}/7^{th}/8^{th} rows). Compared with ERL-Net, the other methods usually generate much more blurry results with important details missing. Even for JORDER, which

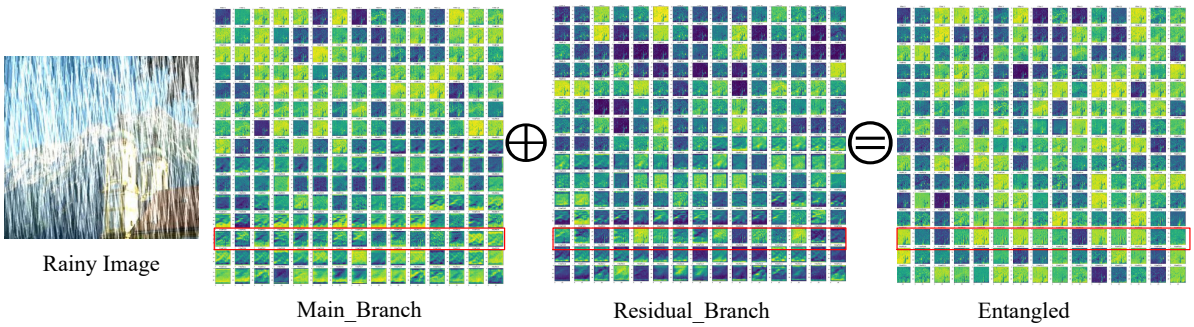

Figure S1: Visualization of activations (before ReLU) for the intermediate layer of ERL-Net (see Figure 2 in the paper) in the main branch, residual branch, and their combinations by entanglement (click `here` to see more visualizations on feature maps distributions).

performs the best among the comparative methods, the results are still very blurry (*e.g.*, the results in the 2nd/7th/9th rows) and some important structural details of the background cannot be recovered/preserved well (*e.g.*, the result in the 1st row is very dark with low contrast caused by the over-deraining effect, and the structure of the wall in the 7th row is not recovered well).

Failure cases: Due to the lack of an explicit control on the learning of the representation in the residual branch, some negative results may be obtained because: If the original representation from the main branch is good enough, the introduction of the residual representation may play negative effect by destroying the original one, thus resulting in some under-controlled results (*e.g.*, the over-deraining phenomenon as in the 8th row of Fig. S6). In the future, it is possible to add some prior [2] to regularize the learning of residual representation, thus eliminating the unexpected residuals if the original representation from the main branch are good enough to guarantee a satisfactory de-raining result.

4. Running time analysis

By performing the de-raining task on a computer equipped with a Tesla P100 GPU, the running of time of different models are reported in Table S1:

DDN	JORDER	DID	NLEDN	PReNet	AGAN	ERL-Net
0.26	6.89	1.89	7.65	0.11	0.85	0.46

Table S1: Running time (s) of different models on a 320×320 sized rainy image. Red color indicates the SOTA rain streak removal methods, Cyan color indicates the SOTA raindrop removal method.

As can be seen from the comparison, our model is able to process rainy image with a comparable running time to other methods while achieving new SOTA de-raining results for both rain streak and raindrop removal tasks.

5. Generalization ability analysis

To test the cross-dataset performance of the proposed model, we record the results on each dataset with ERL-Net trained on different datasets, and the results are listed in Table S2:

As shown in Table S2, even when trained on one dataset (*e.g.*, Rain100H) and tested on another dataset (*e.g.*, DDN), the ERL-Net still achieves better result than the current SOTA, which is obtained by the model trained and tested on the same dataset. Such a comparison fully demonstrates the powerful generality of ERL-Net.

	Different datasets for training ERL-Net			SOTA results
	DDN	DID	Rain100H	
DDN	33.92/0.9502	32.69/0.9460	32.78/0.9463	32.60/0.9458
DID	34.28/0.9365	34.62/0.9403	34.39/0.9372	33.48/0.9229
Rain100H	34.12/0.9379	34.03/0.9371	34.57/0.9387	30.38/0.8939

Table S2: Average PSNR/SSIM values obtained by ERL-Net trained on different datasets. (The dataset with red color indicates the training set while the dataset with blue color indicates the testing set).

References

[1] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *CVPR*, 2019.

[2] Tschannen Michael, Bachem Olivier, and Lucic Mario. Recent advances in autoencoder-based representation learning. In *NeurIPS*, 2018.

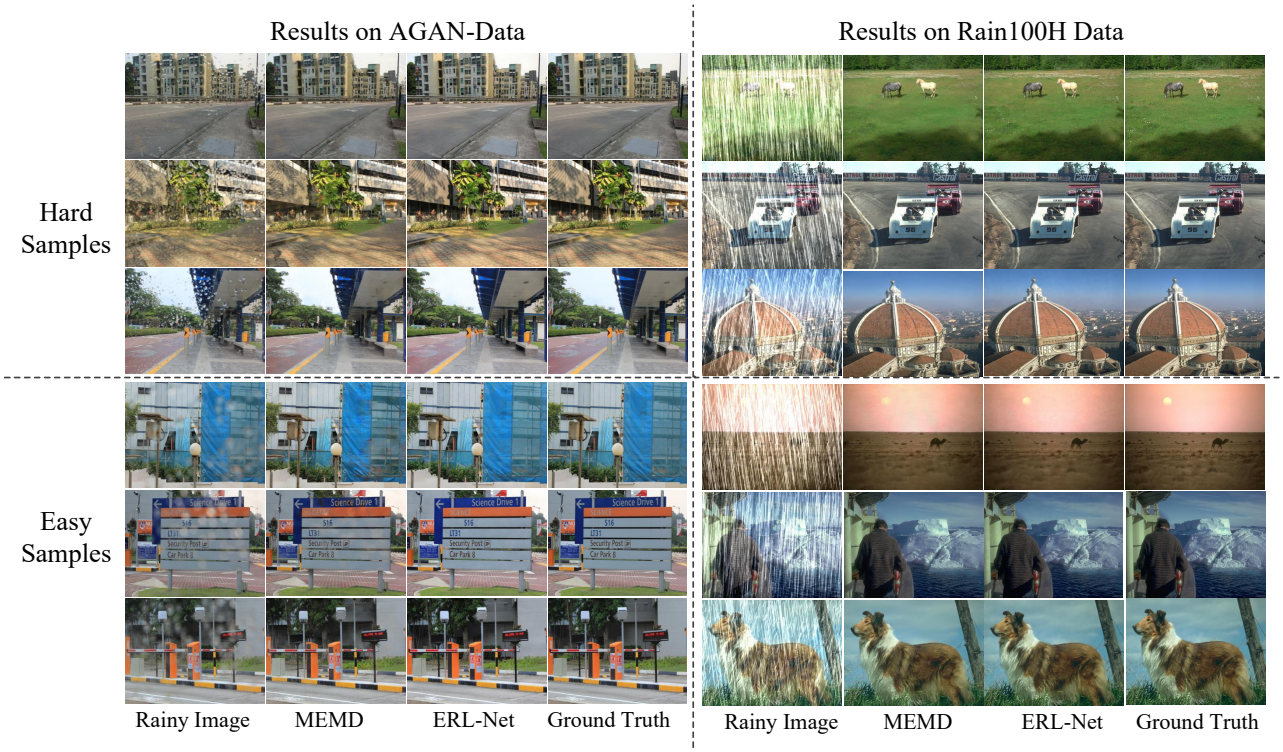

Figure S3: Additional results showing improvement by ERL-Net over baseline MEMD on both hard and easy samples.

Figure S4: Additional visual comparison of raindrop removal results with the AGAN-Data.

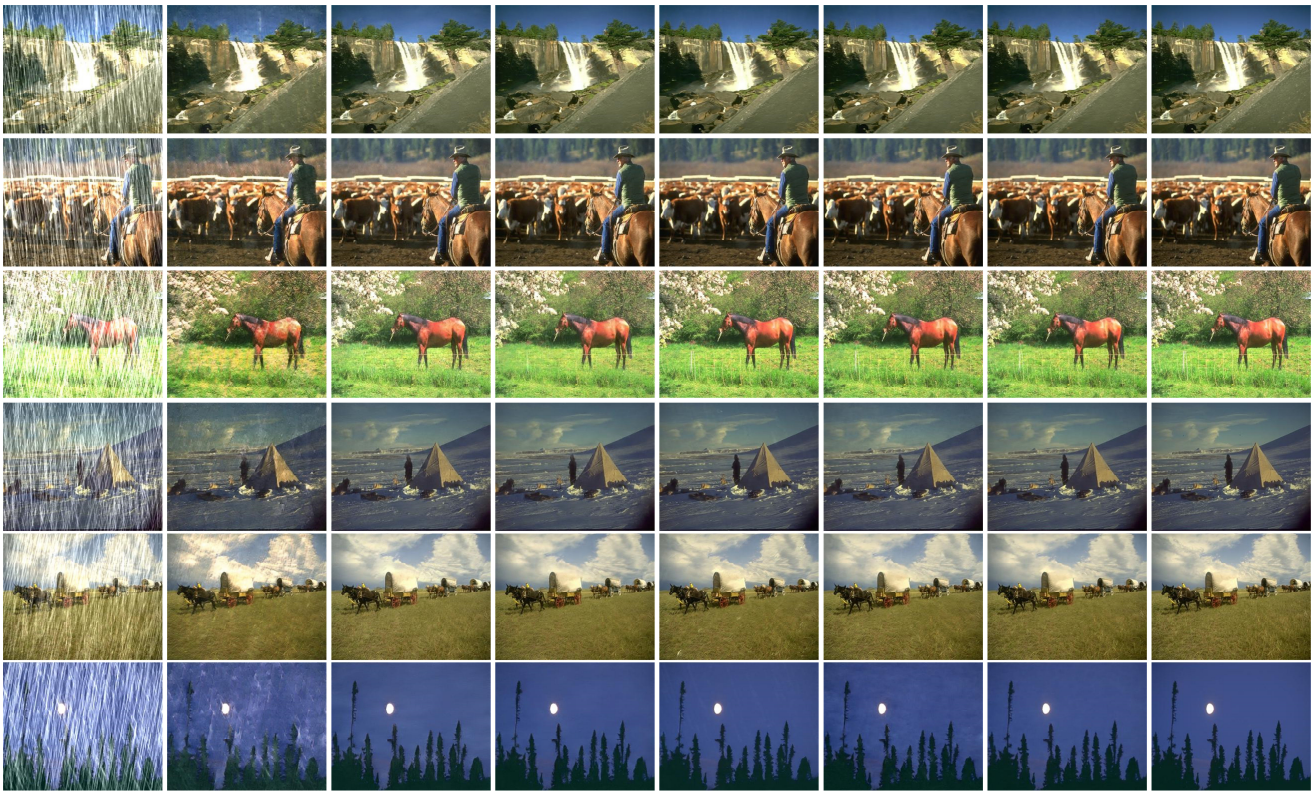

Figure S5: Additional visual comparison of heavy rain streak removal results with synthetic images.

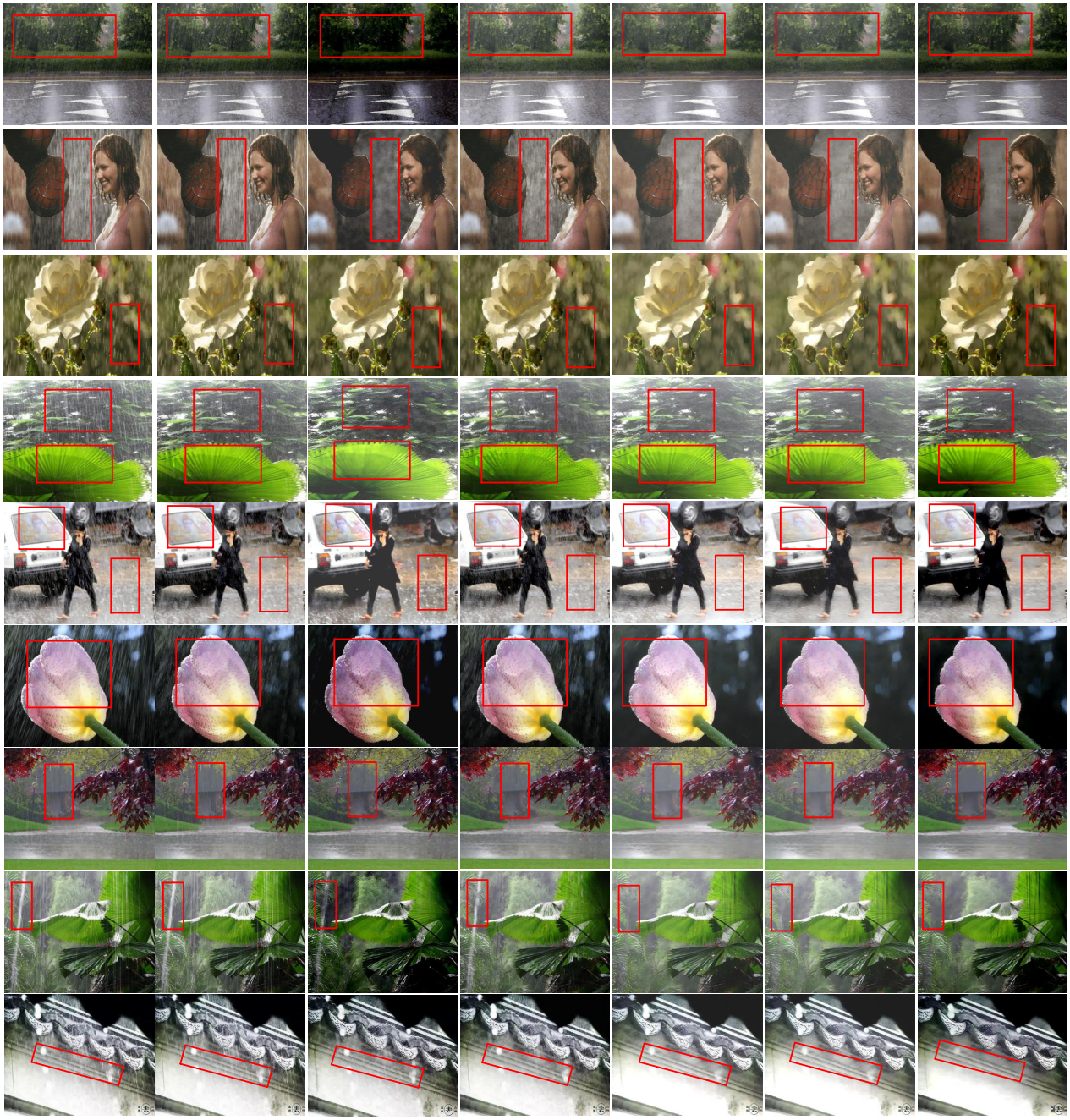

Figure S6: Additional visual comparison of de-raining results with real rainy images.