Deep Video Demoiréing via Compact Invertible Dyadic Decomposition (Supplemental Material)

1. More Details of Alignment Block (AB)

The AB is implemented by the pyramid cascading deformable (PCD) module [4], whose structure is illustrated in Fig. 1. It first extracts features at different scales from both the reference frame and the neighboring frames using standard convolutional layers, and then aligns the features in each scale using deformable convolutions with their offsets predicted in the coarser scales. See also Fig. 2 for an example of the aligned features produced by the AB;.

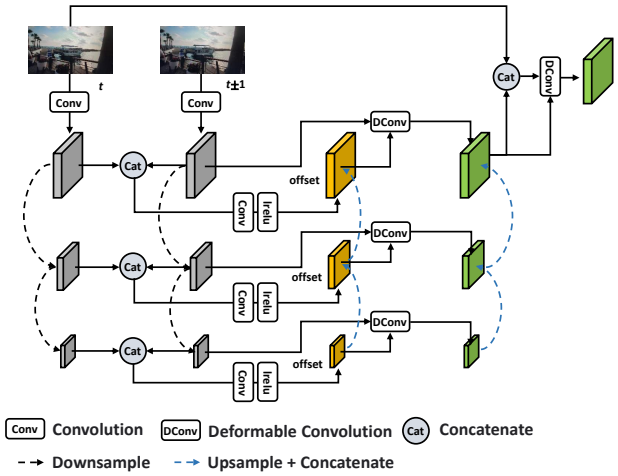


Figure 1: Structure of PCD module used for our AB.

2. Visual Comparison for Ablation Study

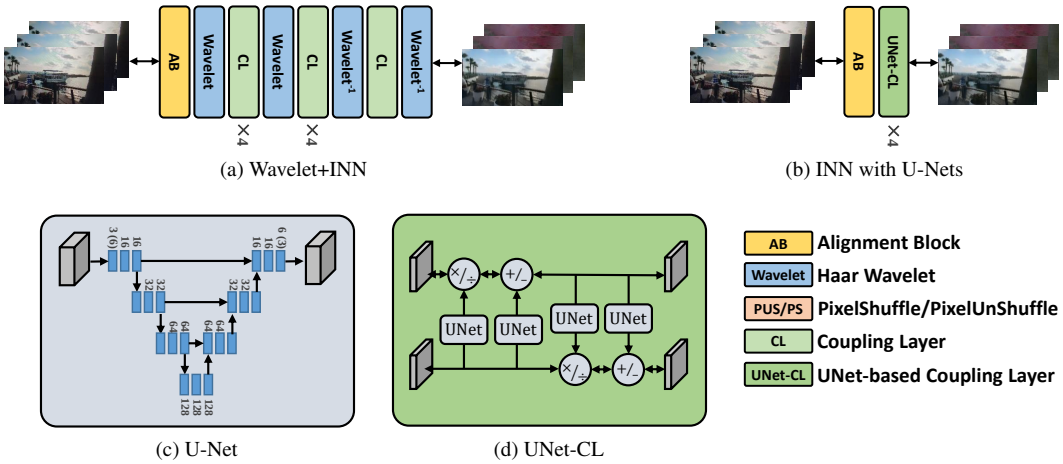
See Fig. 3 for the recovered frames produced by different baselines in our ablation study, with comparison to our original CIDNet. We can observe that the results produced by the original CIDNet are more perceptually satisfactory.

3. Comparison in FPS

See Table 1 for the comparison on FPS between our CIDNet and some video-oriented methods. Our CIDNet is faster than VDN. Note that the ESDNet and DMCNN are even faster due to the original design for single images (extended to videos by retraining on multi-frame input), but they lack specific designs on video processing and thus perform worse than our CIDNet. Therefore, we do not include their

Figure 2: Visualization of the aligned features produced by AB.

Figure 3: Visual comparison for ablation studies.



(a) Wavelet+INN (b) INN with U-Nets

(c) U-Net (d) UNet-CL

Figure 4: Detailed structures of "Wavelet+INN" and "INN with U-Nets".

results for comparison.

Table 1: FPS(↑) on frame size 1280 × 720 using a single RTX 3090, without counting the I/O time.

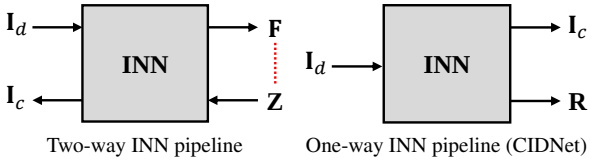
MBCNN	InvDN	VDN	CIDNet (Ours)
2.94	7.75	10.20	11.63

4. Details of Two Multi-Scale Baselines

The details of "Wavelet+INN" and "INN with U-Nets" used in the ablation study are shown in Fig. 4. In "Wavelet+INN" (Fig. 4(a)), the wavelet transform is inserted into a series of coupling layers (CLs) for multi-scale analysis, and the free-form functions in CLs are defined as Dense-Blocks [2]. In "U-Nets in INN" (Fig. 4(b)), each free-form function of CLS is defined as a U-Net (Fig. 4(c)) with the same number of scales as that of the CIDNet. For a fair comparison, the AB used in the CIDNet is also added to the font of each of these two baselines.

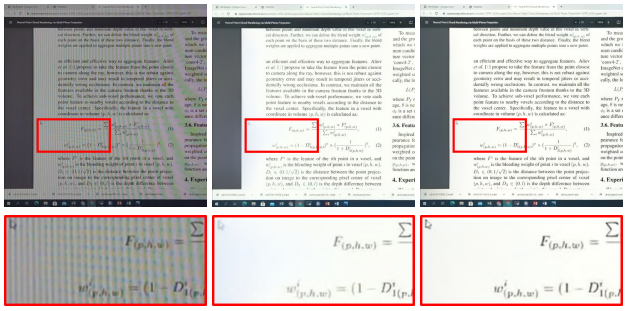
5. Details of The One-Way Baseline

See Fig. 5 for a comparison of the two INN pipelines, two-way INN versus one-way INN used in our ablation study. The two-way INN (with a similar spirit to existing works) first extracts features $\mathbf{F}$ from a degraded image $\mathbf{I}_d$ in the forward pass, then zero out a part of the features in $\mathbf{F}$ to eliminate the undesired pattern, and finally feed the features into the INN in a backward pass to obtain the recovery image $\mathbf{I}_c$. In comparison, the proposed one-way scheme directly decomposes the input degraded image into the latent clean one and the moiréing component. We also show some visual results in Fig. 6 for comparison, where the "one-way CIDNet" can preserve more details for the texts.



Two-way INN pipeline One-way INN pipeline (CIDNet)

Figure 5: Illustration of different INN pipelines.



Input Two-way scheme One-way scheme

Figure 6: Visual comparison of images recovered by the two-way scheme and the one-way scheme.

6. Influence Analysis on Number of Scales/CLs

The number of scales and the number of CLs in each scale are important hyper-parameters for the CIDNet. We further investigate their influence on the performance and complexity. See Table 2 for the results. The performance of CIDNet is improved as the number of scales increases, but the number of FLOPs also increases accordingly. The results of different numbers of CLs are listed in Table 3, where k_i denotes the number of CLs in the ith scale. Not surprisingly, larger k_i achieves better results but leads to higher complexity.

#Scale	LPIPS↓	PSNR (dB)↑	SSIM↑	#Params	#FLOPs
2	0.190	21.90	0.730	4.70	44.54
3	0.184	22.27	0.735	4.57	28.10
4	0.193	21.93	0.729	4.75	19.00

Table 2: Results of CIDNEt with different number of scales.

k_1	k_2	k_3	LPIPS↓	PSNR (dB)↑	SSIM↑	#Params	#FLOPs
1	1	1	0.208	21.18	0.723	2.32	17.98
1	1	2	0.198	21.64	0.726	2.72	18.39
1	2	3	0.187	22.20	0.730	3.52	20.44
2	3	4	0.184	22.27	0.735	4.57	28.10
3	4	5	0.184	22.32	0.736	5.93	37.69

Table 3: Results of CIDNet with different number of CLs in each scale, where k_i for ith scale.

7. More Results in Text Recognition

See Fig. 7 for one more visual example in the downstream text recognition task, where our proposed CIDNet outperforms VDNNet.

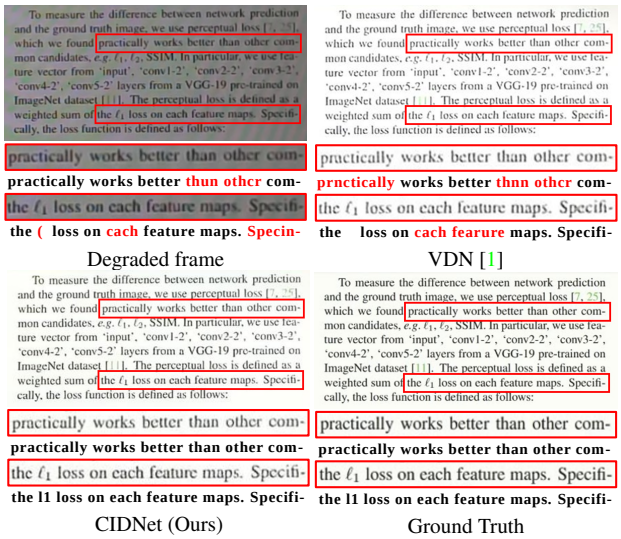


Figure 7: An example of text recognition. 1st row: the degraded/demoiréd frame, 2nd row: zoom-ins, 3rd row: recognized texts, whose erroneous words are marked in RED .

8. More Visual Results

See Fig. 8 and Fig. 9 for more visual comparisons, where our proposed CIDNet generates better results than other compared methods. Video demonstrations can be found at our GitHub site.

References

[1] Peng Dai, Xin Yu, Lan Ma, Baoheng Zhang, Jia Li, Wenbo Li, Jiajun Shen, and Xiaojuan Qi. Video demoireing with relation-based temporal consistency. In *Proccedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17622–17631, 2022. 3, 4

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proccedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. 2

[3] Yang Liu, Zhenyue Qin, Saeed Anwar, Pan Ji, Dongwoo Kim, Sabrina Caldwell, and Tom Gedeon. Invertible denoising network: A light solution for real noise removal. In *Proccedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13365–13374, 2021. 4

[4] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proccedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 1

[5] Xin Yu, Peng Dai, Wenbo Li, Lan Ma, Jiajun Shen, Jia Li, and Xiaojuan Qi. Towards Efficient and Scale-Robust Ultra-High-Definition Image Demoiréing. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Proceedings of the European Conference on Computer Vision*, pages 646–662, Cham, 2022. Springer Nature Switzerland. 4

[6] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. Image demoireing with learnable bandpass filters. In *Proccedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3636–3645, 2020. 4

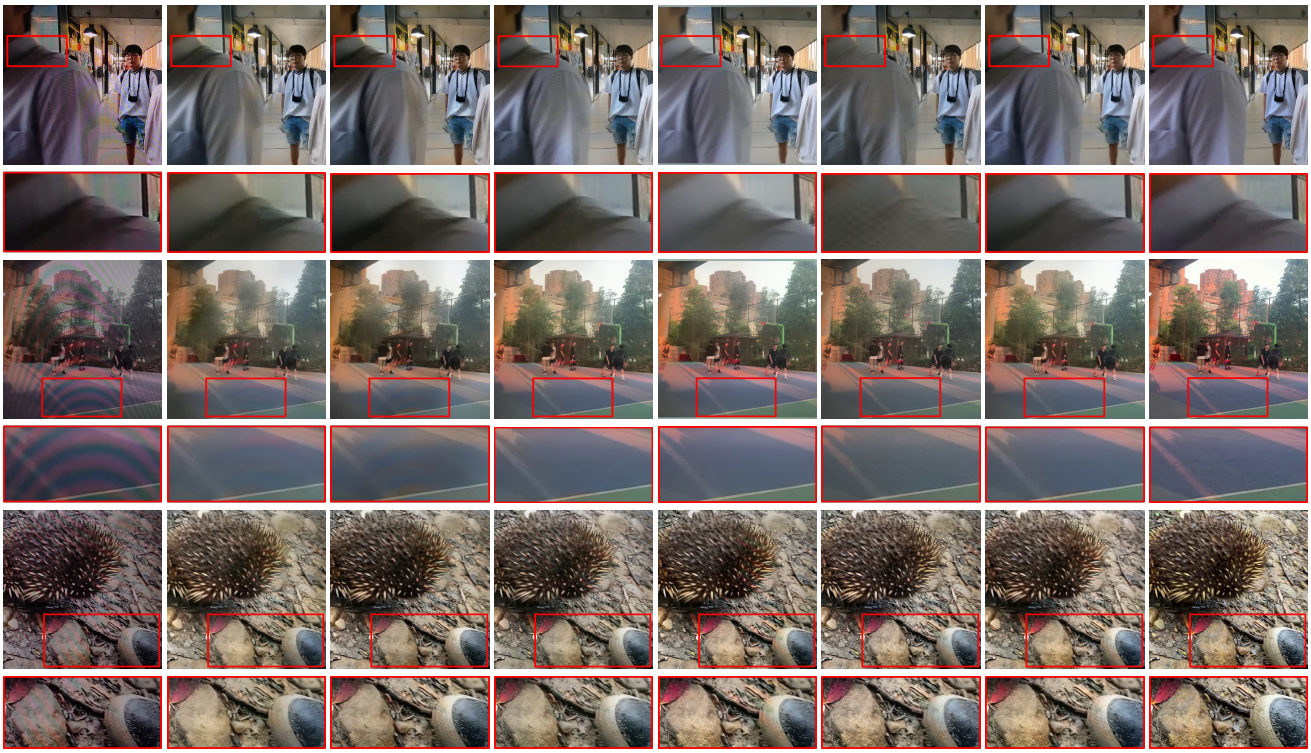


Figure 8: Visual comparison of some selected demoiréd frames in the TCL20 Pro setting.

Figure 9: Visual comparison of some selected demoiréd frames in the IPhoneXR setting.