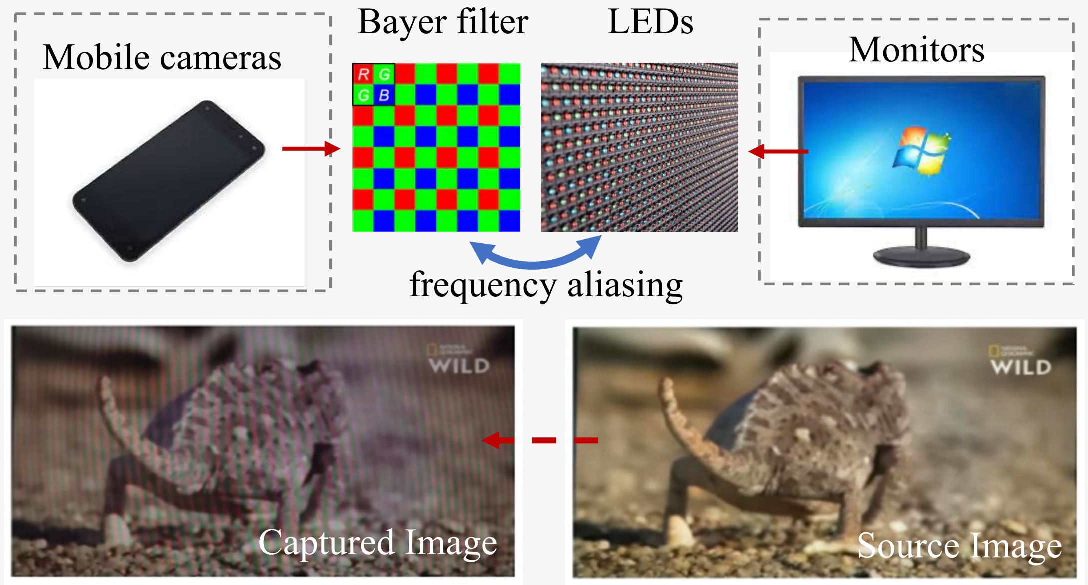
Towards Practical Moiré Artifacts Removal for High-Resolution Images and Videos

INTRODUCTION

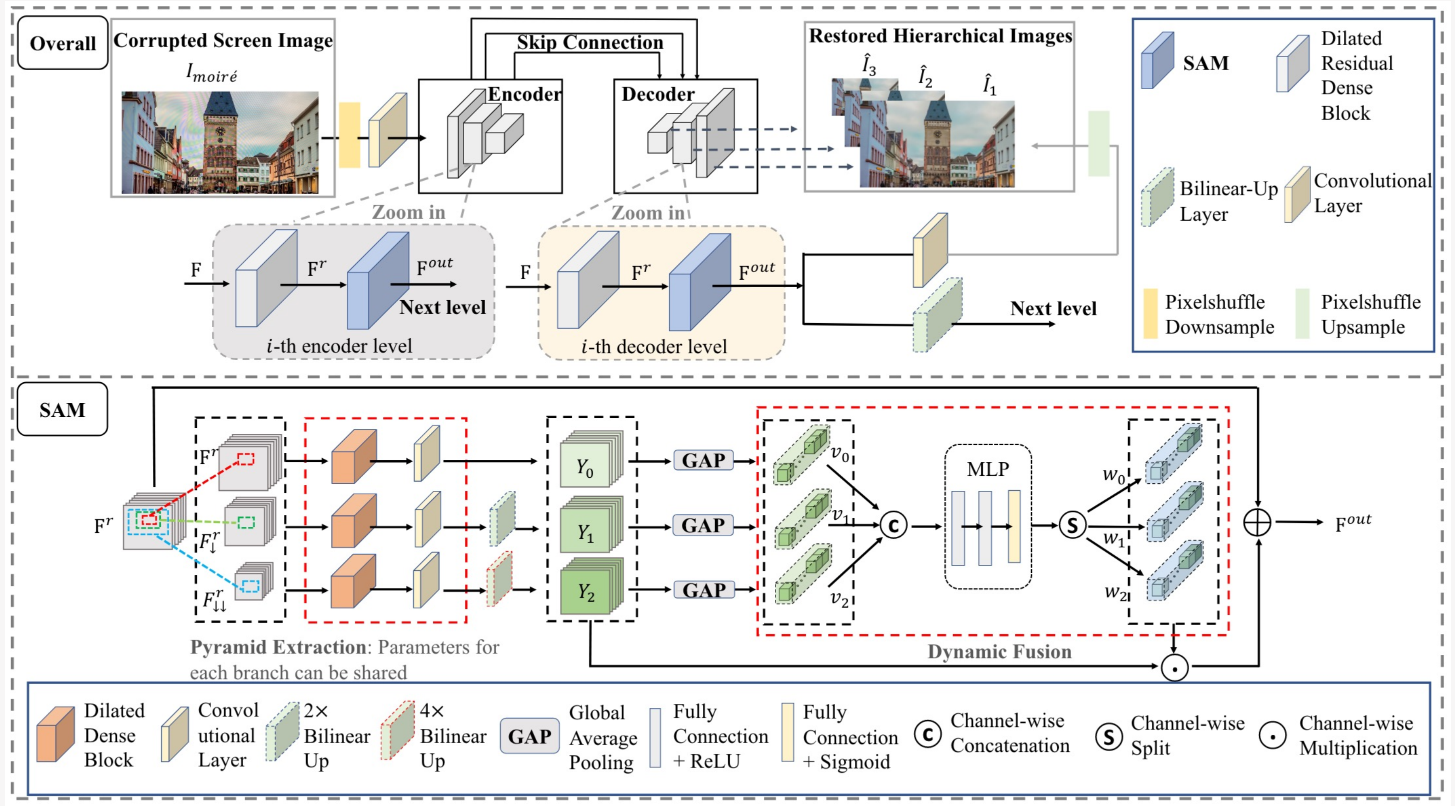
- When photographing the contents displayed on the digital screen, an inevitable frequency aliasing between the camera's color filter array (CFA) and the screen's LCD subpixel widely exists. The captured images are often mixed with colorful stripes, named moiré patterns, which severely degrade the images' perceptual qualities.



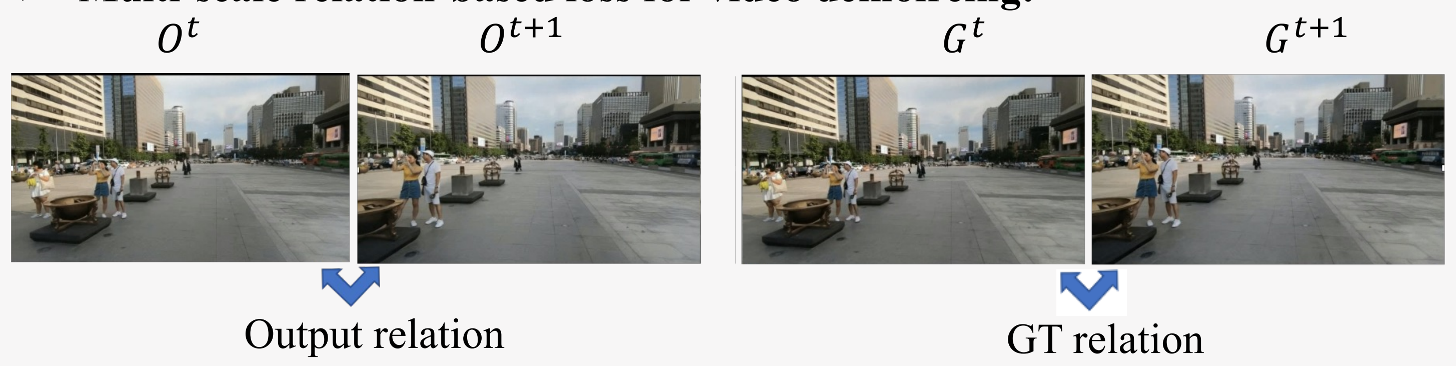
- With the fast development of mobile devices, modern widely used mobile-phones often allow capturing 4K resolution images and 720P videos. Therefore, we build the first 4K resolution moiré image dataset and the first 720P resolution moiré video dataset for exploring and tackling practical real-world high-resolution image and video demoireing problems.
- For moiré image processing, we present an efficient baseline model ESDNet for tackling 4K moiré images, wherein we build a semantic-aligned scale-aware module to address the scale variation of moiré patterns.
- For moiré video processing, we propose a relation-based temporal consistency loss to encourage the model to learn temporal consistency priors directly from ground-truth reference videos, which facilitates producing temporally consistent predictions and effectively maintains frame-level qualities.

METHODS

➢ Efficient 4K image demoireing backbone:



➢ Multi-scale relation-based loss for video demoireing:



Motivations:

- GT videos are already temporally satisfying, and the consistency priors are implicitly encoded into the relationship of consecutive frames.
- Our eyes receive the light intensity from a region (region-level) on the screen rather than a single point (pixel-level).
- Directly learn the temporal consistency by forcing the output relation and the GT relation to be the same.

$$L_{mbr} = \frac{1}{N}\sum_{n=1}^{N} L_n^{k^*}|_{k^*=\arg\min_k\{|(\mathcal{T}_k(O^{t+1})-\mathcal{T}_k(O^t))_n|\}, k\in C},$$

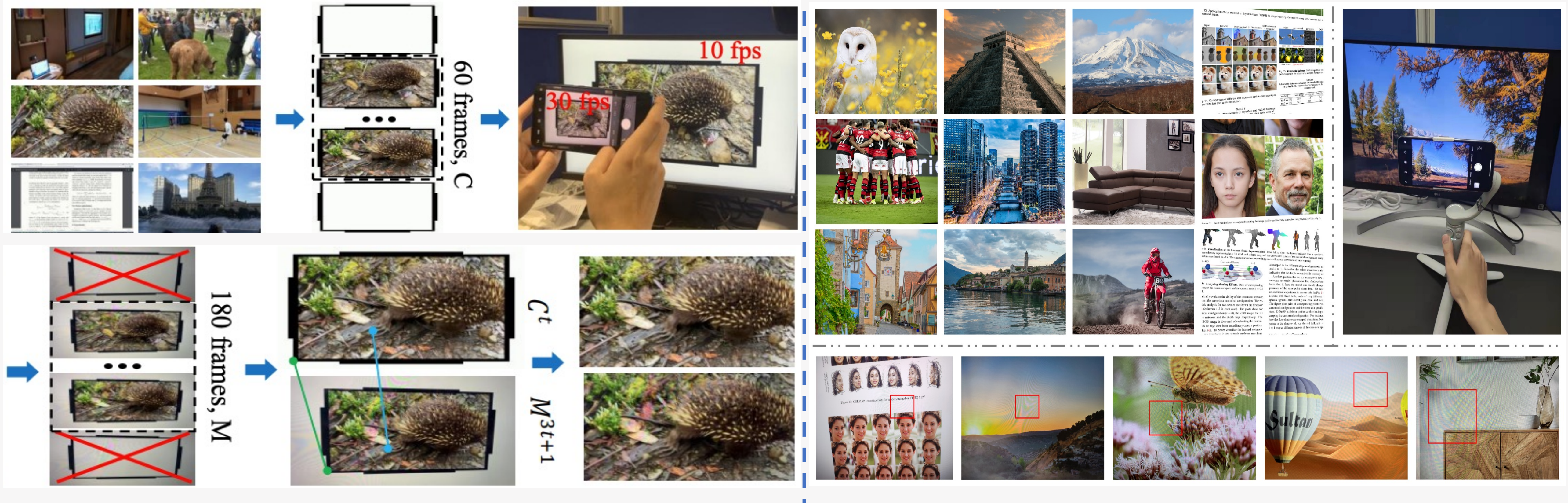
$$L_n^k = |((\mathcal{T}_k(O^{t+1})-\mathcal{T}_k(O^t))_n - (\mathcal{T}_k(G^{t+1})-\mathcal{T}_k(G^t))_n|,$$

Where O^t and O^{t+1} are two output frames, G^t and G^{t+1} are reference frames, $\mathcal{T}_k$ means region-level statistics (i.e., mean), and N is the number of pixels.

Advantages:

- Easy to implement
- No optical flows (imprecise)
- Consider natural changes

DATASETS

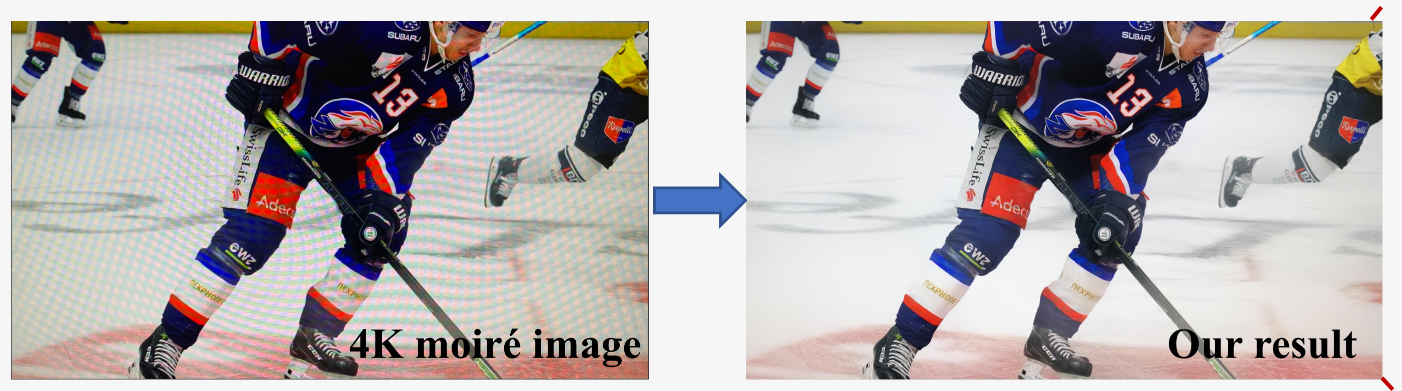


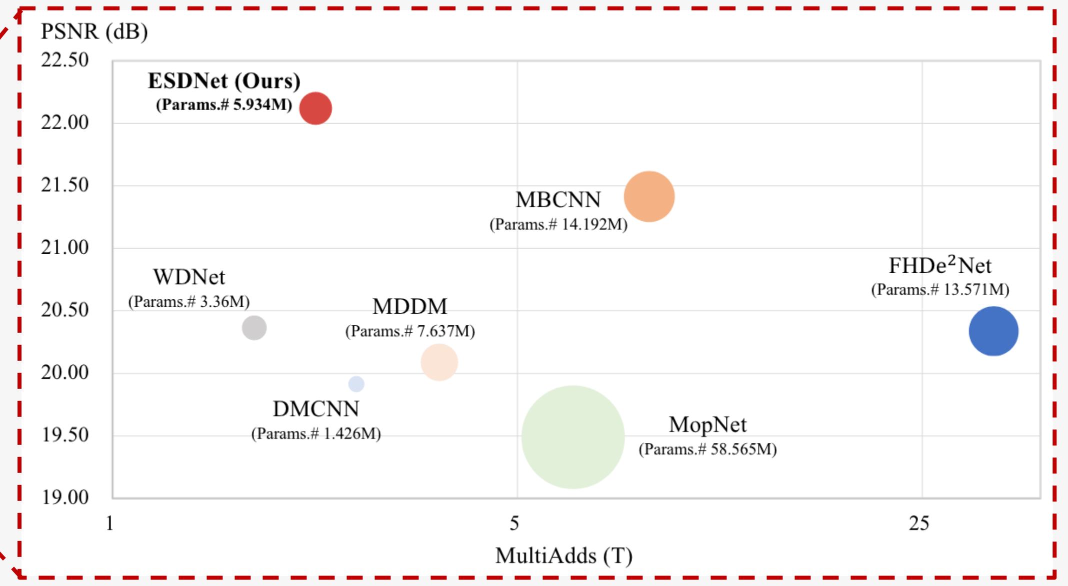
Our 720P video datasets

Our 4K image datasets

RESULTS

➢ Efficient 4K image processing





➢ Effective relation-based loss for video processing

Flow-based Basic relation-based Ours GT

Methods	LPIPS ↓	PSNR ↑	SSIM ↑
MBCNN [55]	0.260	21.534	0.740
DMCNN [40]	0.321	20.321	0.703
U-Net [35]	0.225	20.348	0.720
Ours_S	0.212	21.772	0.729
Ours	0.202	21.725	0.733