DeblurGAN v2

Image & Video Enhancement

Members: Eduin Hernandez, Edwin Rios, Felipe Manosalvas

0860862 0860802 0860801

Introduction

Nowadays, people are unable to separate themselves from their smartphone. We carry them to work, classes, parties, and on trips. It our main tool to capture the moment via images and videos, but it doesn't always turn out well, especially when our hands are the main source of support for the camera. The end result are blurry images that fail to capture the moment for us, but what if we could recover them?

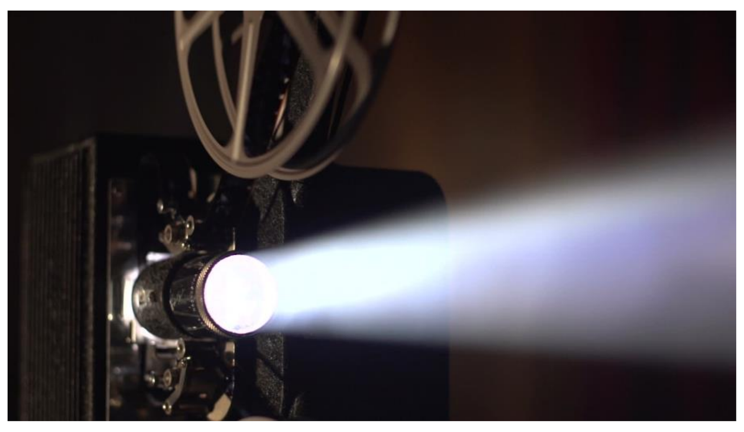
DeblurGAN is an end-to-end learned method developed by Kupyn et al. [6] for motion deblurring which provided state-of-the art performance for structure similarity and visual appearance. DeblurGAN v2 is the further improves the deblurring efficiency, quality, and flexibility.

Benefits

- Image Deblurring
- Video Deblurring
- Interframe Enhancement
- Super Image Resolution
- Image Segmentation
- Object Detection

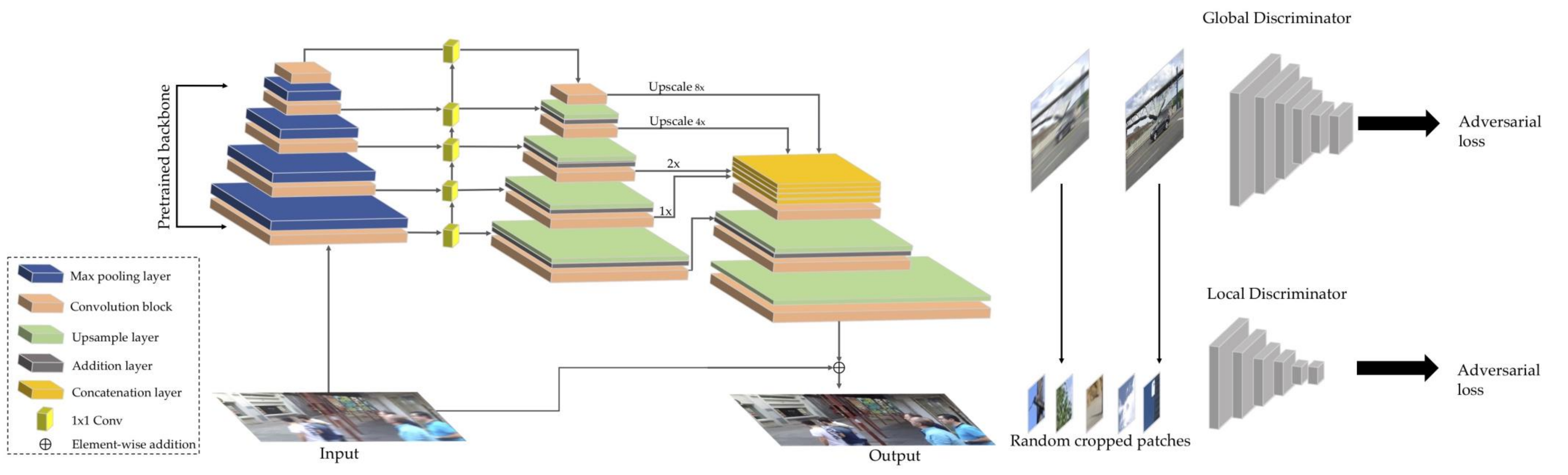
Practical Uses

- Smartphone App
- Camera Filter
- Video Editing



Architecture

Feature Pyramid Deblurring



Structure

- Bottom-up Pathway
- Top-down Pathway
- Lateral Connections

Backbones

- Inception ResNet v2
- MobileNet v2
- MobileNet DSC

Discriminators

- Local
- Global

Evaluation Metric

- Pixel Space Loss (L_p)
- Perceptual Distance/ Content Loss (L_X)
- Adverbial Loss (L_{adv})

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2$$

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}$$

$$L_G = 0.5 * L_p + 0.006 * L_X + 0.01 * L_{adv}$$

$$L_D^{RaLSGAN} = \mathbb{E}_{x\sim p_{data}(x)}\left[(D(x) - \mathbb{E}_{z\sim p_z(z)}D(G(z)) - 1)^2\right] + \mathbb{E}_{z\sim p_z(z)}\left[(D(G(z)) - \mathbb{E}_{x\sim p_{data}(x)}D(x) + 1)^2\right]$$

$$PSNR = 10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\cdot\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) = 20\cdot\log_{10}(MAX_I) - 10\cdot\log_{10}(MSE)$$

Experiments

Training Dataset

- GoPro
- DVD
- NFS

Performance

Table 5: Ablation Study on the GoPro dataset, based on DeblurGAN-v2 (Inception-ResNet-v2).

	PSNR	SSIM
DeblurGAN (starting point)	28.70	0.927
+ FPN	29.26	0.931
+ FPN + Global D	29.29	0.932
+ FPN + Global D + RaGAN-LS	29.37	0.933
DeblurGAN-v2 (FPN + Global D + RaGAN-LS + MSE Loss)	**29.55**	**0.934**
Removing perceptual loss (replace 0.5 with 0 in L_G)	28.81	0.924

Table 1: Performance and efficiency comparison on the GoPro test dataset, All models were tested on the *linear* image subset.

	Sun *et al.* [43]	Xu *et al.* [51]	DeepDeblur [33]	SRN [45]	DeblurGAN [21]	Inception-ResNet-v2	MobileNet	MobileNet-DSC
PSNR	24.64	25.10	29.23	**30.10**	28.70	29.55	28.17	28.03
SSIM	0.842	0.890	0.916	0.932	0.927	**0.934**	0.925	0.922
Time	20 min	13.41s	4.33s	1.6s	0.85s	0.35s	0.06s	**0.04s**
FLOPS	N/A	N/A	1760.04G	1434.82G	678.29G	411.34G	43.75G	**14.83G**

Visual Results

Figure 1: Sharp Image (Left), Sharp Image after DeblurGANv2 – Inception ResNetv2 (Middle), Sharp Image after DeblurGANv2 – MobileNet v2 (Right)

Figure 2: Blurred Image (Left), Deblurred Image after DeblurGANv2 – Inception ResNetv2 (Middle), Deblur Image after DeblurGANv2 – MobileNet v2 (Right)

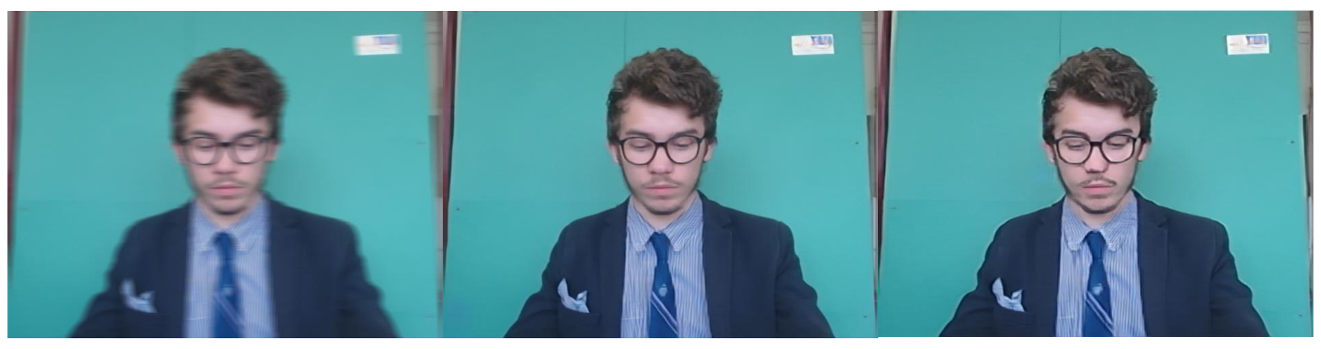


Figure 3: Blurred Image (Left), Deblurred Image after DeblurGANv2 – Inception ResNetv2 (Middle), Deblur Image after DeblurGANv2 – MobileNet v2 (Right)

Figure 4: Sharp Image (Left), Deblurred Image after DeblurGANv2 – Inception ResNetv2 (Middle), Deblur Image after DeblurGANv2 – MobileNet v2 (Right)

		Figure 1	Figure 2	Figure 3	Figure 4
PSNR	Inception	42.549	26.02	27.48	36.39
	MobileNet	49.24	29.67	26.32	34.75
SSIM	Inception	0.994	0.866	0.893	0.972
	MobileNet	0.998	0.93	0.868	0.959

Conclusion

1. DeblurGAN-v2 provides good results and fast performance, but it is not fast enough for live-video deblurring.
2. DeblurGAN-v2 is flexible enough to switch between different backbones, in exchange of tradeoffs between performance and efficiency.
3. Can be used to increase the framerate of animated videos to obtain more fluid fps versions of them.

Future Work

There's room for improvement for the technology. By implementing a parallel implementation of the framework, near-live video deblurring and FPS increase could be obtained. Fine-tuning of the network weights conditioned to a smaller training dataset similar to whatever video style we desire to perform enhancements on, higher performance might be achieved. It would require prior knowledge of when will enhancements be needed, but it is worth investigating.

A final line of improvements come from a GAN specifically dedicated to FPS increase, by using high-FPS videos down sampled to lower-FPS and using these pairs as inputs/outputs for training the network, it could theoretically learn to produce high-FPS videos without outside transformations, and as shown by the impressive results obtained by deblurGAN-v2, obtain state-of-the-art video frame interpolation. We baptized this conceptual network as fpsGAN and plan to further continue doing research on the topic.

References

[1] GoPro Camera Specifications Comparison.

[2] iPhone 11 - Technical Specifications.

[3] PyTorch.

[4] Specifications j Samsung Galaxy S10e, S10 & S10+.

[5] TAMU-VITA/DeblurGANv2, December 2019. original-date: 2019-08-10T09:02:40Z.

[6] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. arXiv:1908.03826 [cs], August 2019. arXiv: 1908.03826 version: 1.

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com