Contrastive Learning for Controllable Blind Video Restoration

Givi Meishvili Abdelaziz Djelouah Sally Hattori Christopher Schroers

BMVC 2022

Disney Research Studios

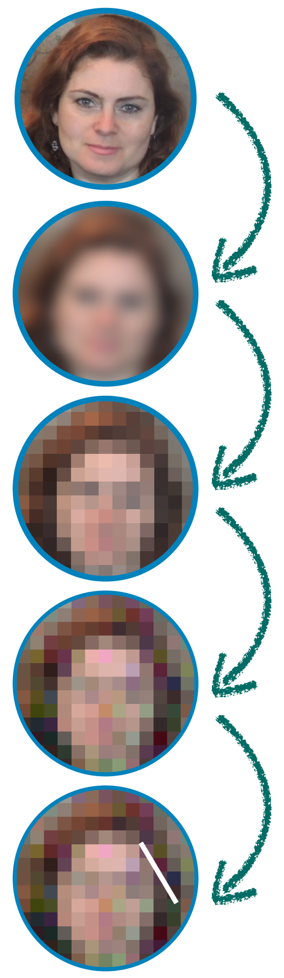
Motivation & Contribution

» Multi-degradation image restoration models have shown impressive results

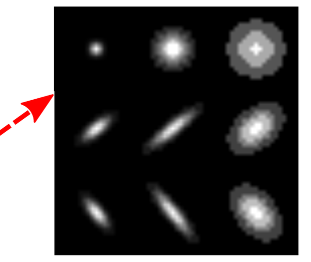
» For real-world applications:
- initial restoration in a blind setting
- possibility to interpret and manipulate results

» In summary
- A multi-degradation blind video restoration model
- Contrastive training to learn an interpretable and controllable degradation representation
- State-of-the-art results in blind video restoration

Degradation Model

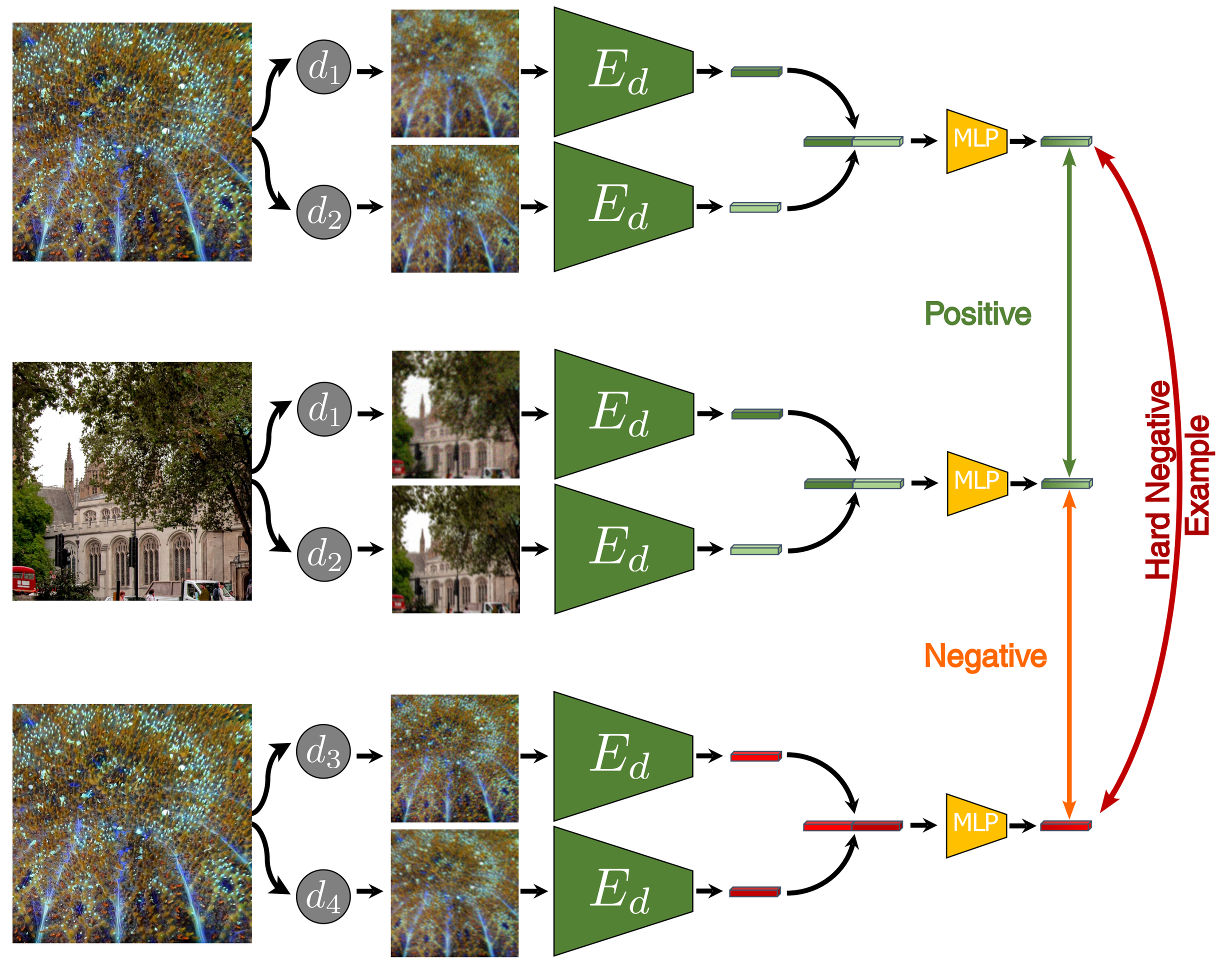


Convolution w. Gaussian Kernel	$x * k$
Downsampling	$(x * k) \downarrow_s$
Noise	$(x * k) \downarrow_s + n$
Film Scratches	$S \circ \left((x * k) \downarrow_s + n \right)$

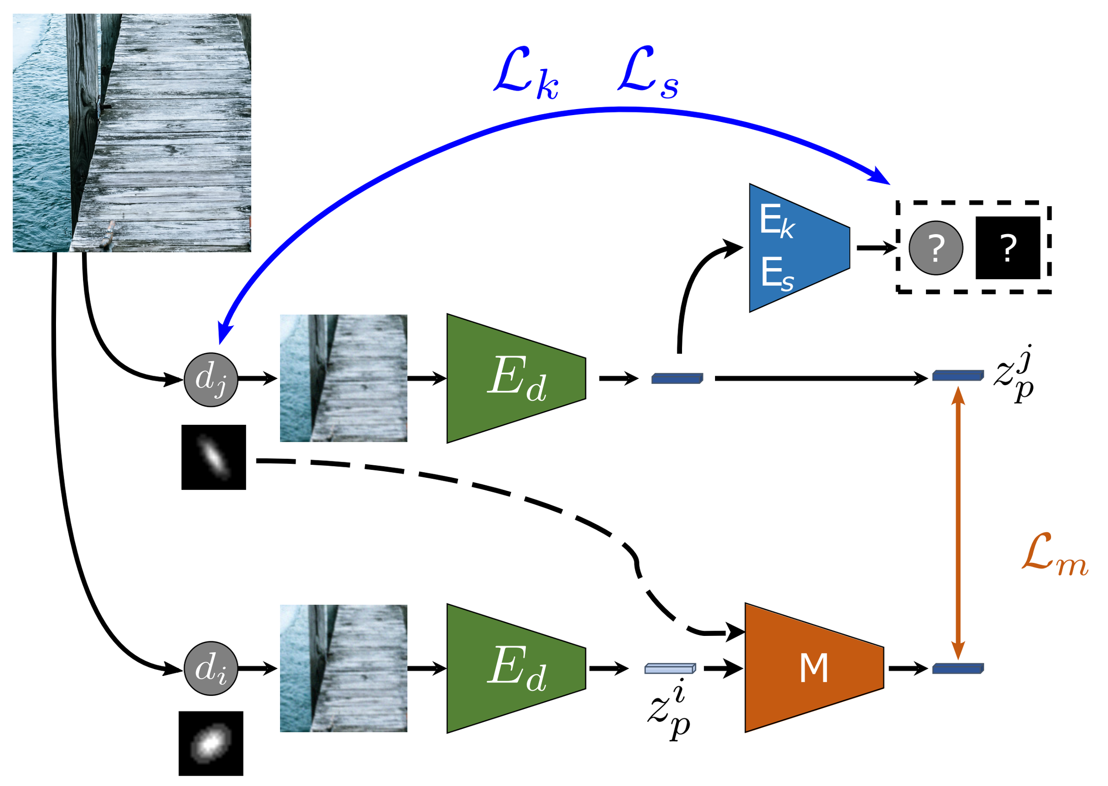


Method

Pairwise contrastive loss ($\mathcal{L}_c$) with hard negatives



Degradation regressor and mutator



Degradation representation learning

Using the Moco [1] representation learning framework and regression losses for blur and noise

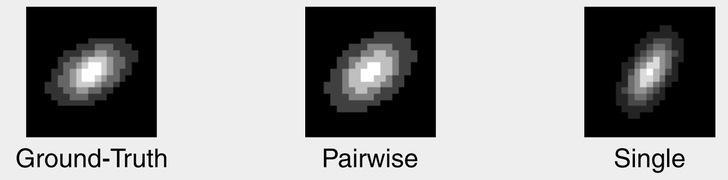
$$\mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_k \mathcal{L}_k + \lambda_s \mathcal{L}_s$$

Degradation Mutator

$$\mathcal{L}_m = \sum_p^{\mathcal{V}} \sum_{i,j}^{\mathcal{D}} \left| M(z_p^i, k_j, \sigma_j) - z_p^j \right|$$

» Benefits of pairwise training strategy

Example

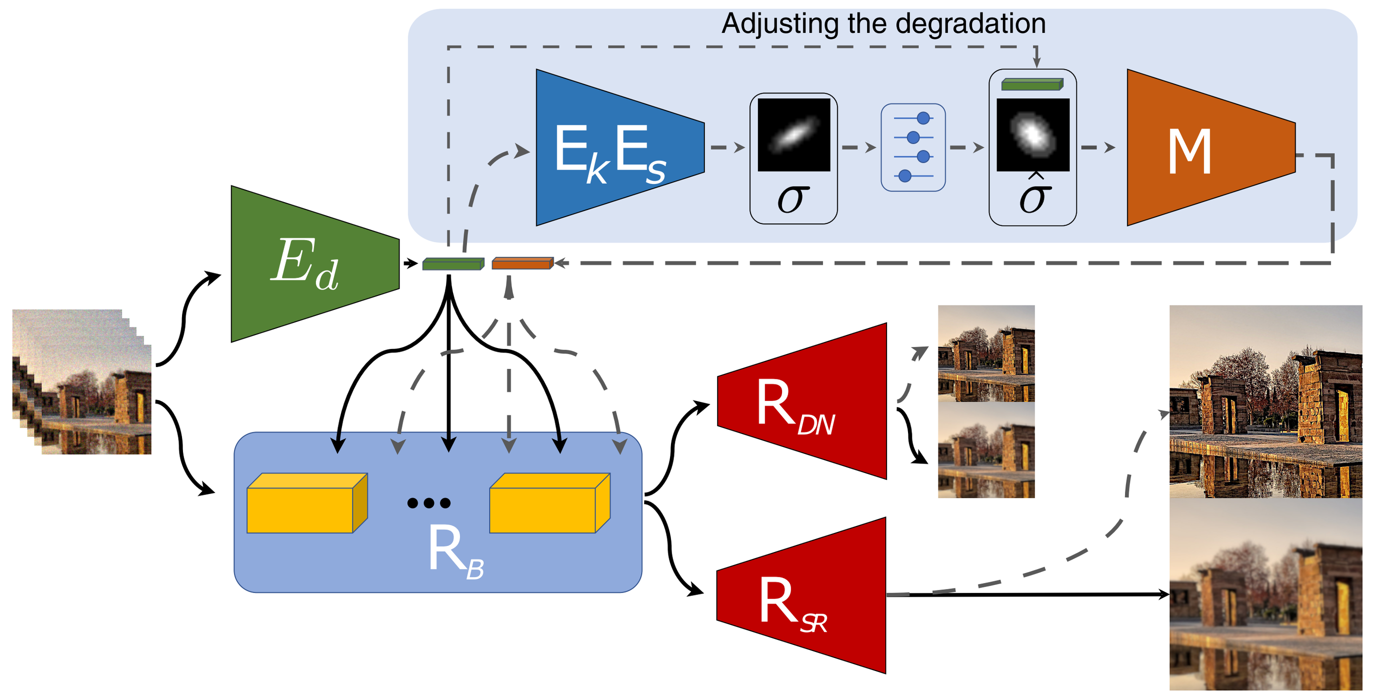


Feature Contrasting	MAE↓	Kernel Similarity↑
Single	0.0008	0.9438
KernelGAN[2]	0.0006	0.9446
Pairwise	0.0005	0.9821

Training for Conditional Video Restoration

Continue training the degradation representation along the restoration tasks

$$\mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_k \mathcal{L}_k + \lambda_s \mathcal{L}_s + \underbrace{\lambda_{SR} \mathcal{L}_{SR}}_{\text{Super-resolution loss}} + \underbrace{\lambda_{DN} \mathcal{L}_{DN}}_{\text{Denoising loss}}$$



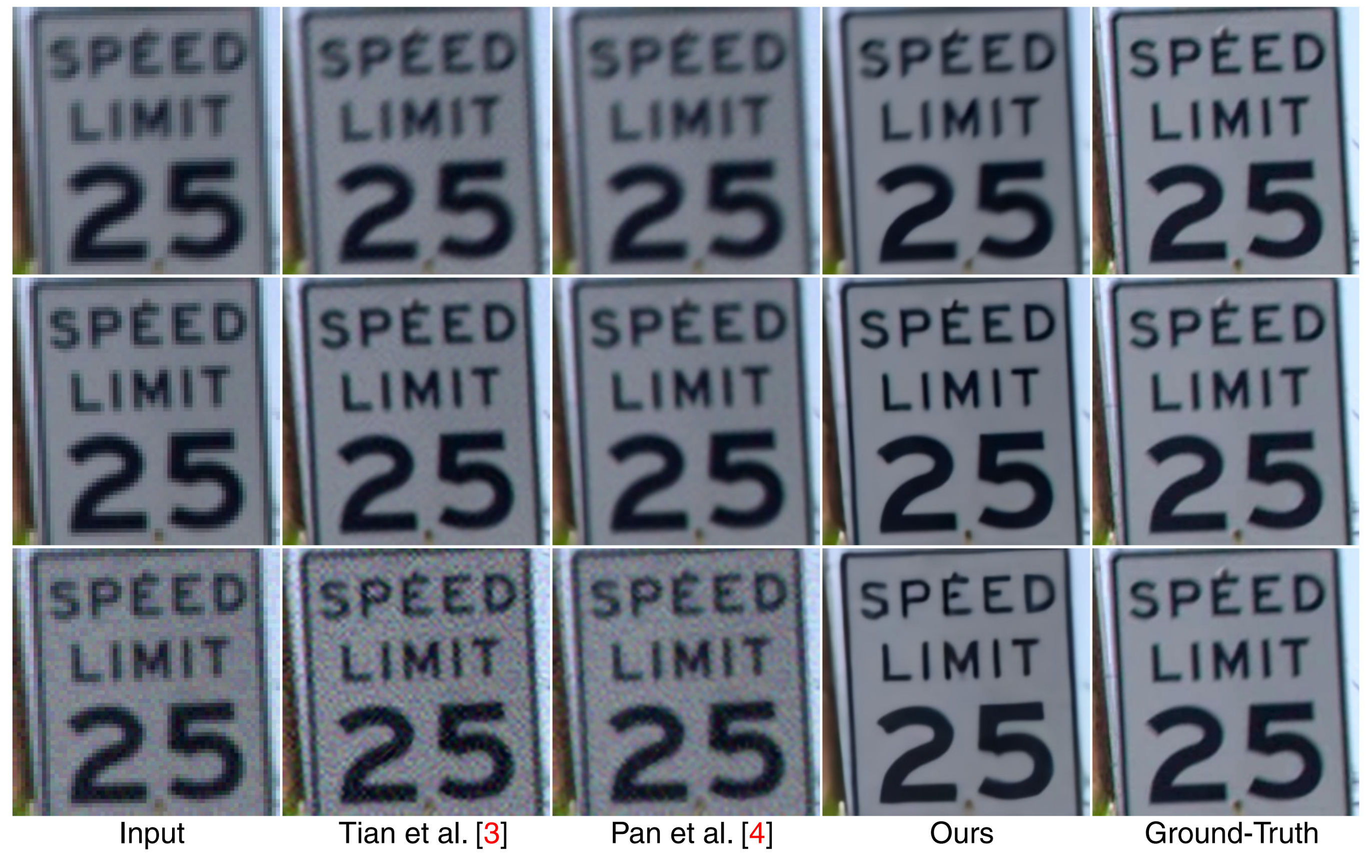
Experimental Results

Denoising & Super-Resolution

	PSNR /SSIM
Ours	**27.07** / **0.76**
Tian et al. [3]	26.09 / 0.68
Pan et al. [4]	25.55 / 0.68
Zhang et al. [5]	25.11 / 0.67

Denoising

	VID4	SET8
Ours	**34.09 /0.99**	**33.65** / 0.98
UDVD [6]	32.52 / 0.97	32.71 / 0.95
DVDnet [7]	33.07 / 0.97	32.7 / 0.95
FastDVD [8]	33.57 / **0.99**	33.14 / **0.99**



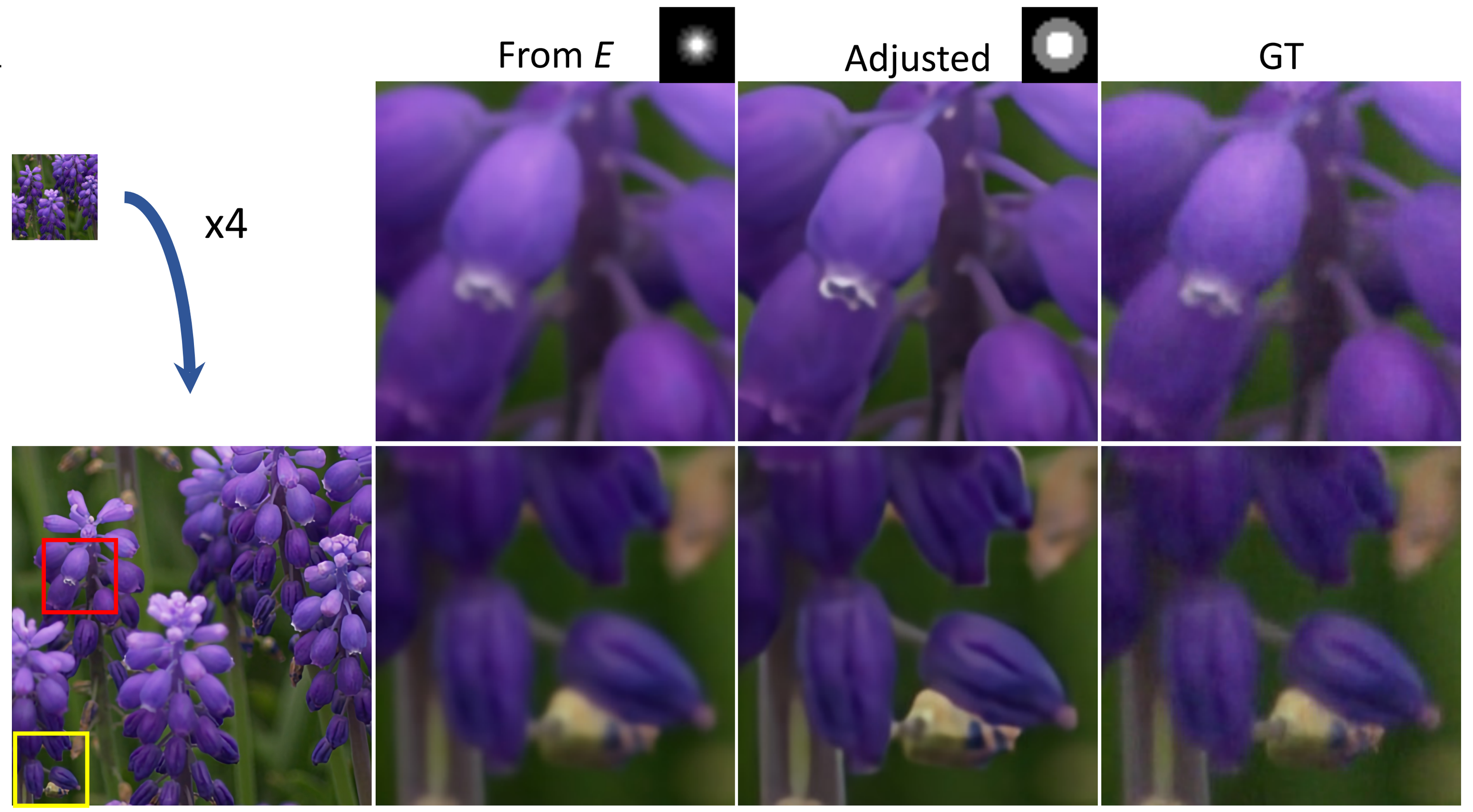
Film Scratch Removal

	VID4	SET8
Ours	**36.09 / 0.99**	**31.93 / 0.98**
Wan et al. [9]	24.54 / 0.83	26.98 / 0.86

Controllable video restoration

» Estimated Kernel is close to ground truth

» Adjusted Kernel allows for sharper results



References

[1] Momentum contrast for unsupervised visual representation learning, CVPR 2020
[2] Blind super-resolution kernel estimation using an internal-gan, NeurIPS. 2019
[3] Tdan: Temporally-deformable alignment network for video SR, CVPR 2020
[4] Deep blind video super-resolution, ICCV 2021
[5] Designing a practical degradation model for deep blind image SR, ICCV 2021
[6] Unsupervised deep video denoising, ICCV 2021
[7] Dvdnet: A fast network for deep video denoising, ICIP 2019
[8] Fastdvdnet: Towards real-time deep video denoising ..., CVPR 2020
[9] Bringing old photos back to life, CVPR 2020