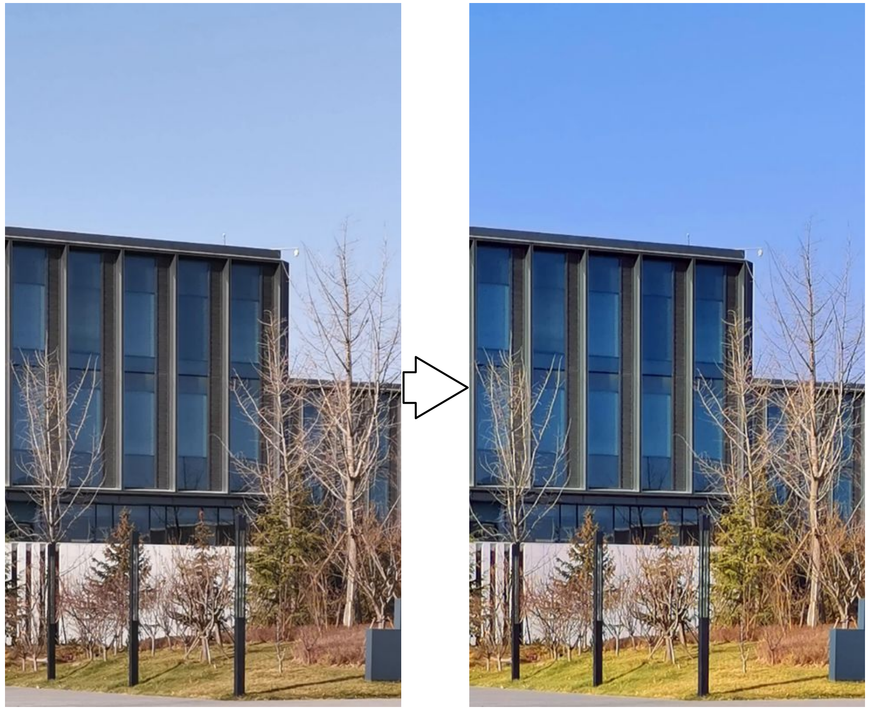
Compound Quality Mapping on High-resolution Images and Videos

Danda Pani Paudel & Zhiwu Huang

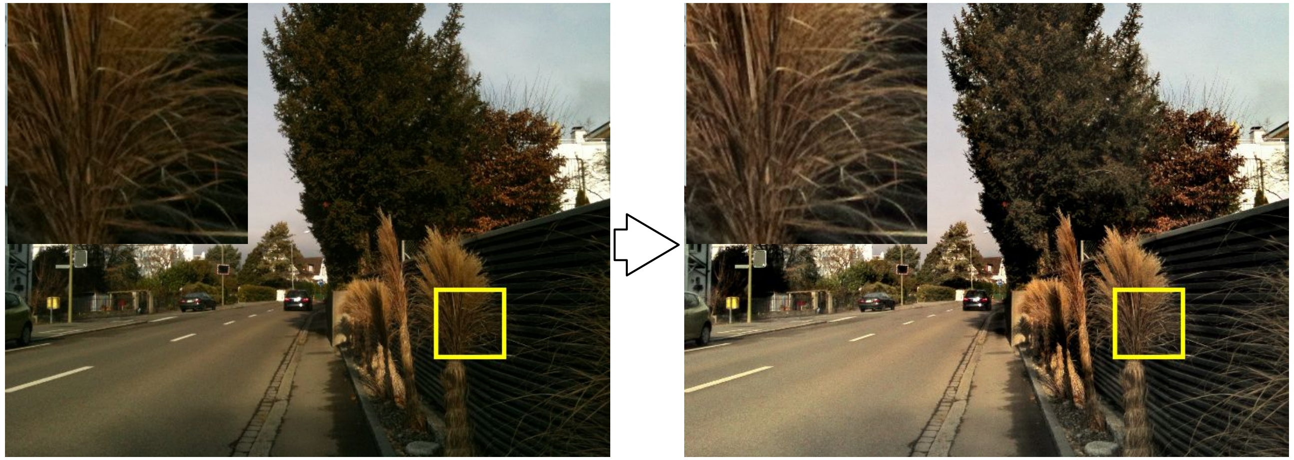
Computer Vision Lab @ ETH Zurich

Compound Quality Mapping



Color Adjustment

Illumination Enhancement



Texture Sharpening

etc....

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Data Collection for Compound Image and Video Quality Mapping

Data Collection I: Paired Image Retouching (expensive expert effort)

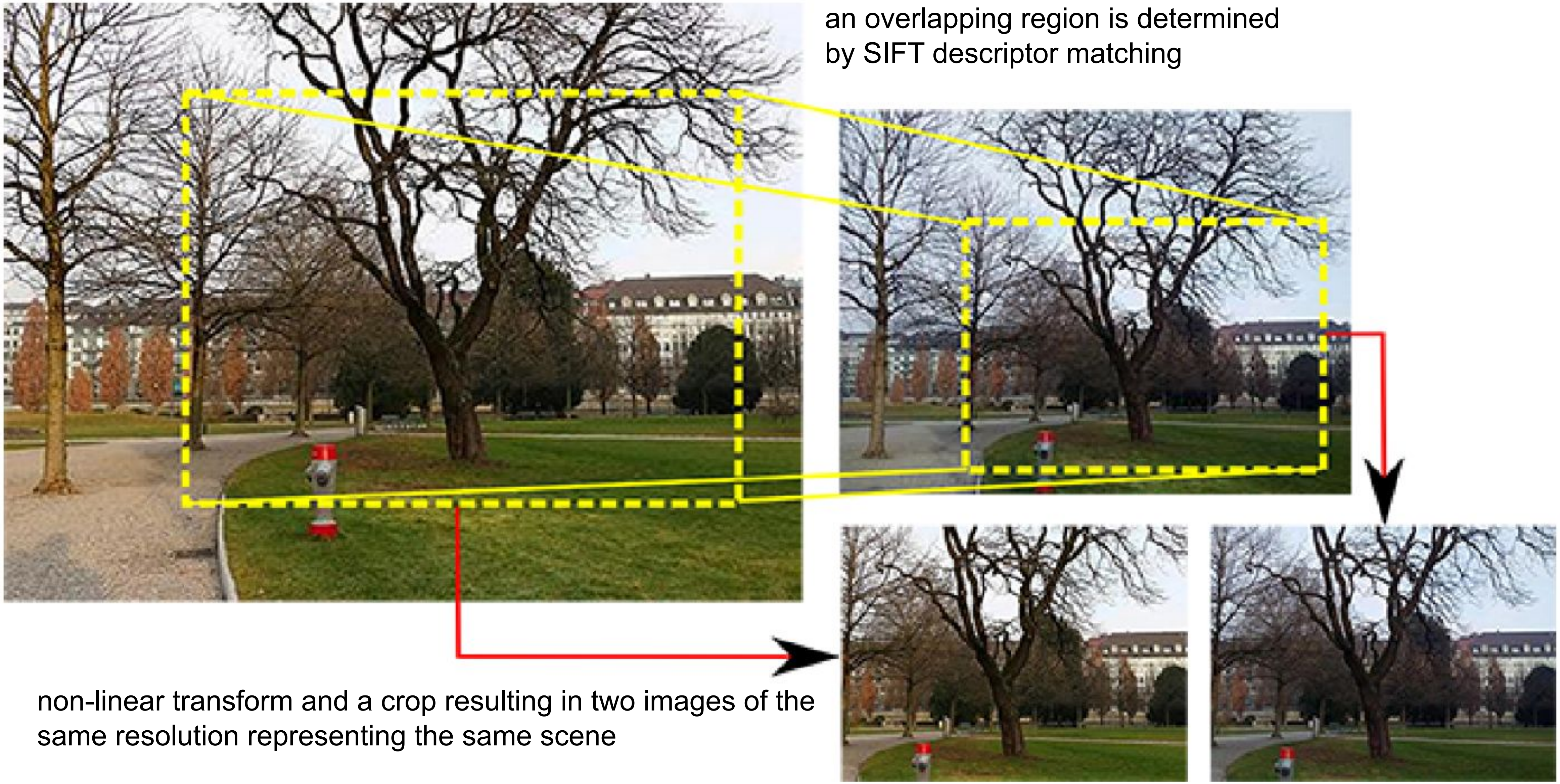
Original | Retouched by expert-A | Retouched by expert-B

Data Collection II: Weakly-paired Collection (expensive alignment)

Camera	Sensor	Image size	Photo quality
iPhone 3GS	3 MP	2048 × 1536	Poor
BlackBerry Passport	13 MP	4160 × 3120	Mediocre
Sony Xperia Z	13 MP	2592 × 1944	Average
Canon 70D DSLR	20 MP	3648 × 2432	Excellent

DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks, Ignatov et al., ICCV 2017.

an overlapping region is determined
by SIFT descriptor matching



non-linear transform and a crop resulting in two images of the
same resolution representing the same scene

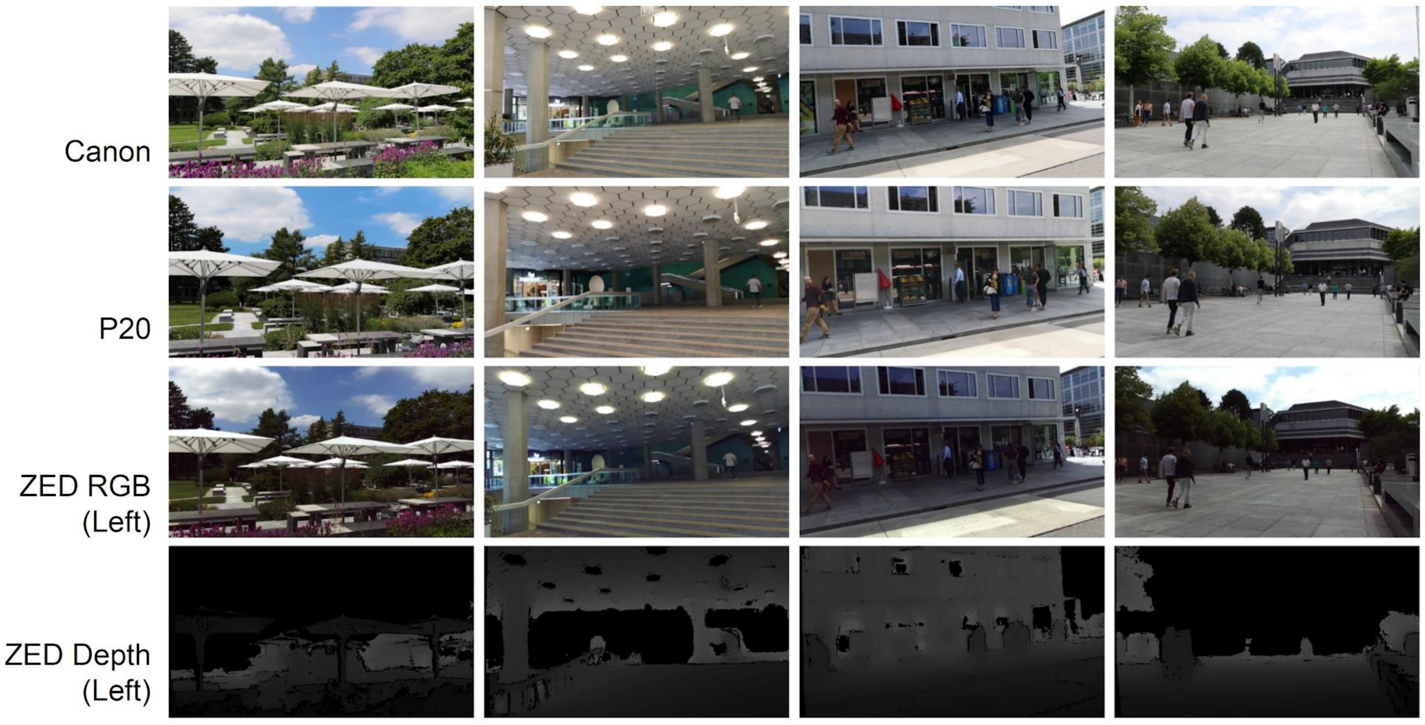
Data Collection II: Weakly-paired Video Collection? (more expensive)

Vid3oC Dataset

- Three cameras
- 82 recordings
- 328 videos
- Includes stereo depth

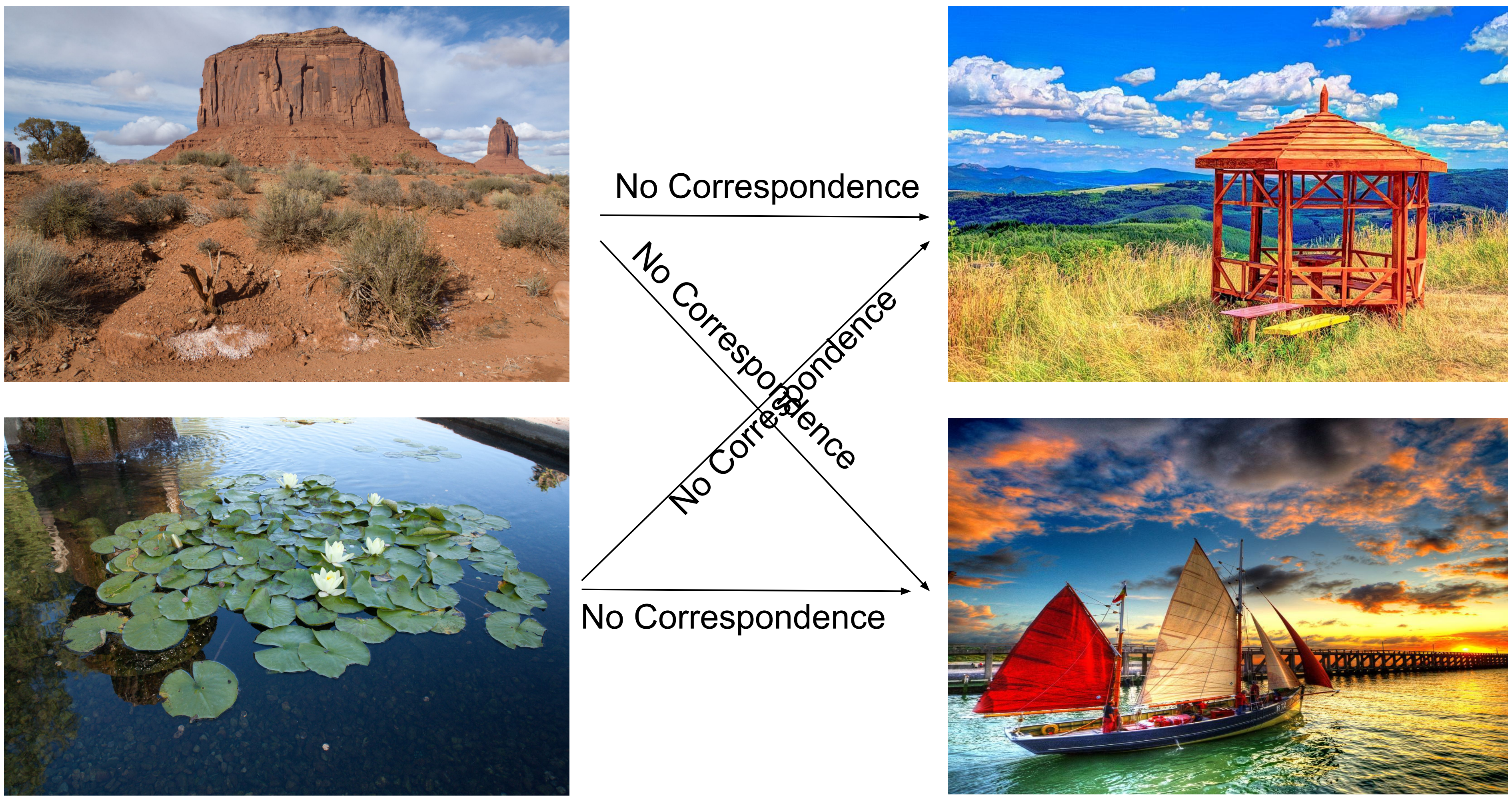
- **Canon 5D Mark IV** high quality DSLR
- **Huawei P20** high-end smartphone
- **ZED** stereo camera

- **AIM 2019** Video SR



The Vid3oC and IntVID Datasets for Video Super Resolution and Quality Mapping. Sohyeong Kim, Guanju Li, Dario Fuoli, Martin Danelljan, Zhiwu Huang, Shuhang Gu and Radu Timofte. ICCV 2019 Workshops.

Data Collection II: Unpaired Collection (Cheaper)



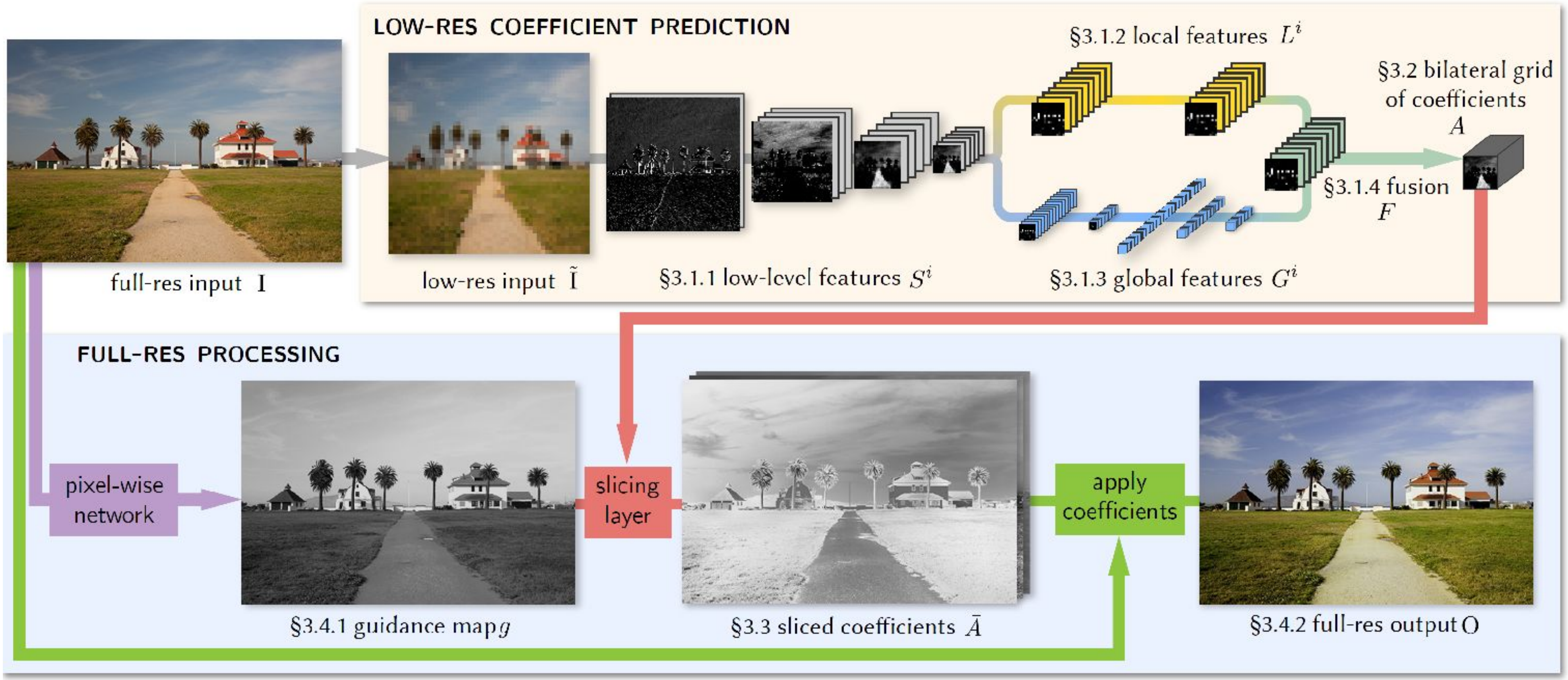
Poor-quality Image Dataset

High-quality Image Dataset

Supervised Deep Learning Methods for Compound Image Quality Mapping

Deep Bilateral Learning for Real-Time Image Enhancement (HDRNet)

Idea: consumes a low-resolution version of the input image, followed by an edge-preserving upsampling to the full-resolution image in a bilateral filtering fashion



Optimized by L2-loss

$$\mathcal{L} = \frac{1}{|\mathcal{D}|} \sum_i \|\mathrm{I}_i - \mathrm{O}_i\|^2$$

Deep Bilateral Learning for Real-Time Image Enhancement, GHARBI et al., TOG 2017.

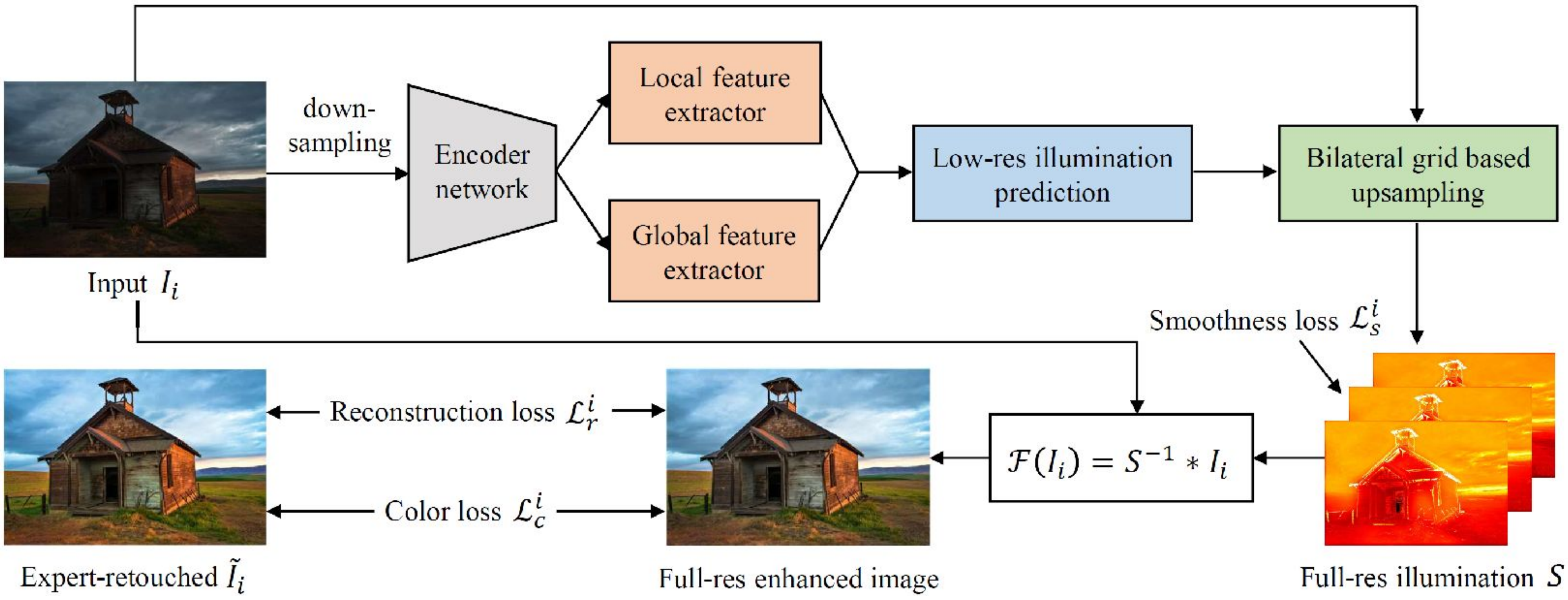
Deep Bilateral Learning for Real-Time Image Enhancement (HDRNet)

12 megapixel 16-bit linear input (tone-mapped for visualization)	tone-mapped with HDR+ **400 – 600 ms**	processed with our algorithm **61 ms**, PSNR = **28.4 dB**

Deep Bilateral Learning for Real-Time Image Enhancement, GHARBI et al., TOG 2017.

Underexposed Photo Enhancement (UPE)

Idea: learn an image-to-illumination (instead of image-to-image) mapping



Underexposed Photo Enhancement using Deep Illumination Estimation, Wang et al., CVPR 2019.

Underexposed Photo Enhancement (UPE)

Loss function:

$$\mathcal{L} = \sum_{i=1}^{N} \omega_r \mathcal{L}_r^i + \omega_s \mathcal{L}_s^i + \omega_c \mathcal{L}_c^i$$

Reconstruction

$$\mathcal{L}_r^i = \|I_i - S * \tilde{I}_i\|^2,$$
$$s.t. \quad (I_i)_c \leq (S)_c \leq 1\ , \quad \forall \text{ pixel channel } c$$

Smoothness

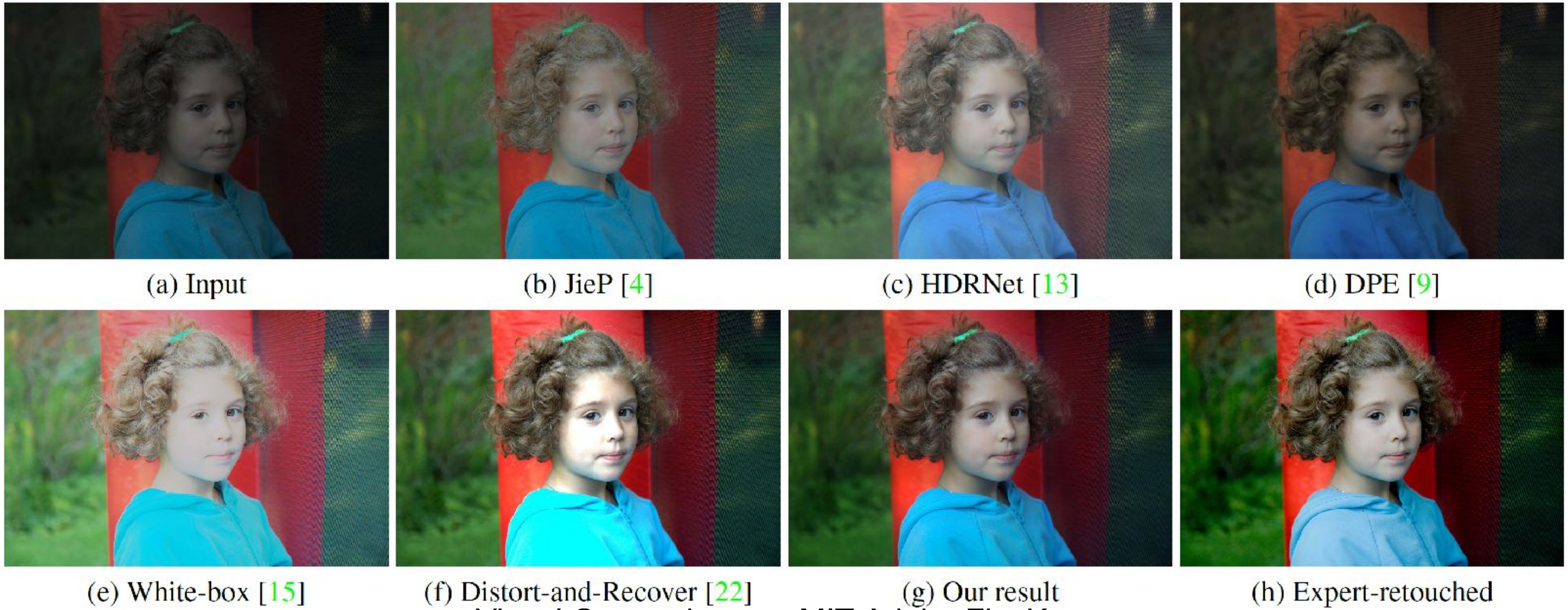
$$\mathcal{L}_s^i = \sum_p \sum_c \omega_{x,c}^p (\partial_x S_p)_c^2 + \omega_{y,c}^p (\partial_y S_p)_c^2$$

Color

$$\mathcal{L}_c^i = \sum_p \angle((\mathcal{F}(I_i))_p, (\tilde{I}_i)_p)$$

Underexposed Photo Enhancement using Deep Illumination Estimation, Wang et al., CVPR 2019.

Underexposed Photo Enhancement (UPE)



Visual Comparison on MIT-Adobe FiveK

Underexposed Photo Enhancement using Deep Illumination Estimation, Wang et al., CVPR 2019.

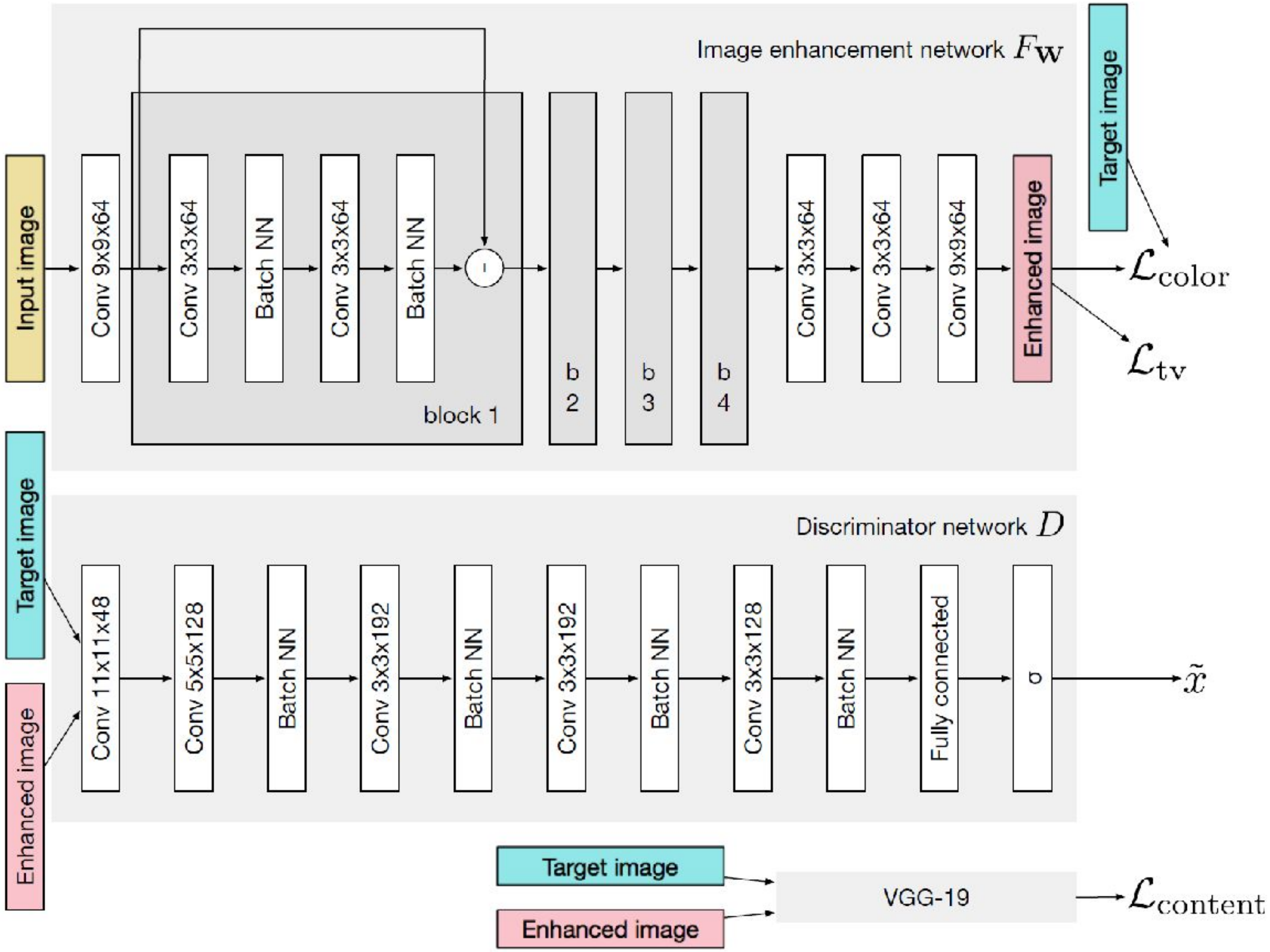
Underexposed Photo Enhancement (UPE)

Method	PSNR	SSIM
HDRNet [2]	21.96	0.866
DPE [1]	22.15	0.850
White-Box [3]	18.57	0.701
Distort-and-Recover [4]	20.97	0.841
Ours w/o $\mathcal{L}_r$, w/o $\mathcal{L}_s$, w/o $\mathcal{L}_c$	21.97	0.867
Ours with $\mathcal{L}_r$, w/o $\mathcal{L}_s$, w/o $\mathcal{L}_c$	22.31	0.871
Ours with $\mathcal{L}_r$, with $\mathcal{L}_s$, w/o $\mathcal{L}_c$	22.89	0.884
Ours	**23.04**	**0.893**

Quantitative Comparison on MIT-Adobe FiveK

Underexposed Photo Enhancement using Deep Illumination Estimation, Wang et al., CVPR 2019.

DSLR Photo Enhancement (DSLR-PE)



Idea: learn the translation function using a residual convolutional neural network with a composite perceptual error function that combines content, color and adversarial texture losses

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{content}} + 0.4 \cdot \mathcal{L}_{\text{texture}} + 0.1 \cdot \mathcal{L}_{\text{color}} + 400 \cdot \mathcal{L}_{\text{tv}}$$

$$\mathcal{L}_{\text{color}}(X, Y) = \|X_b - Y_b\|_2^2$$

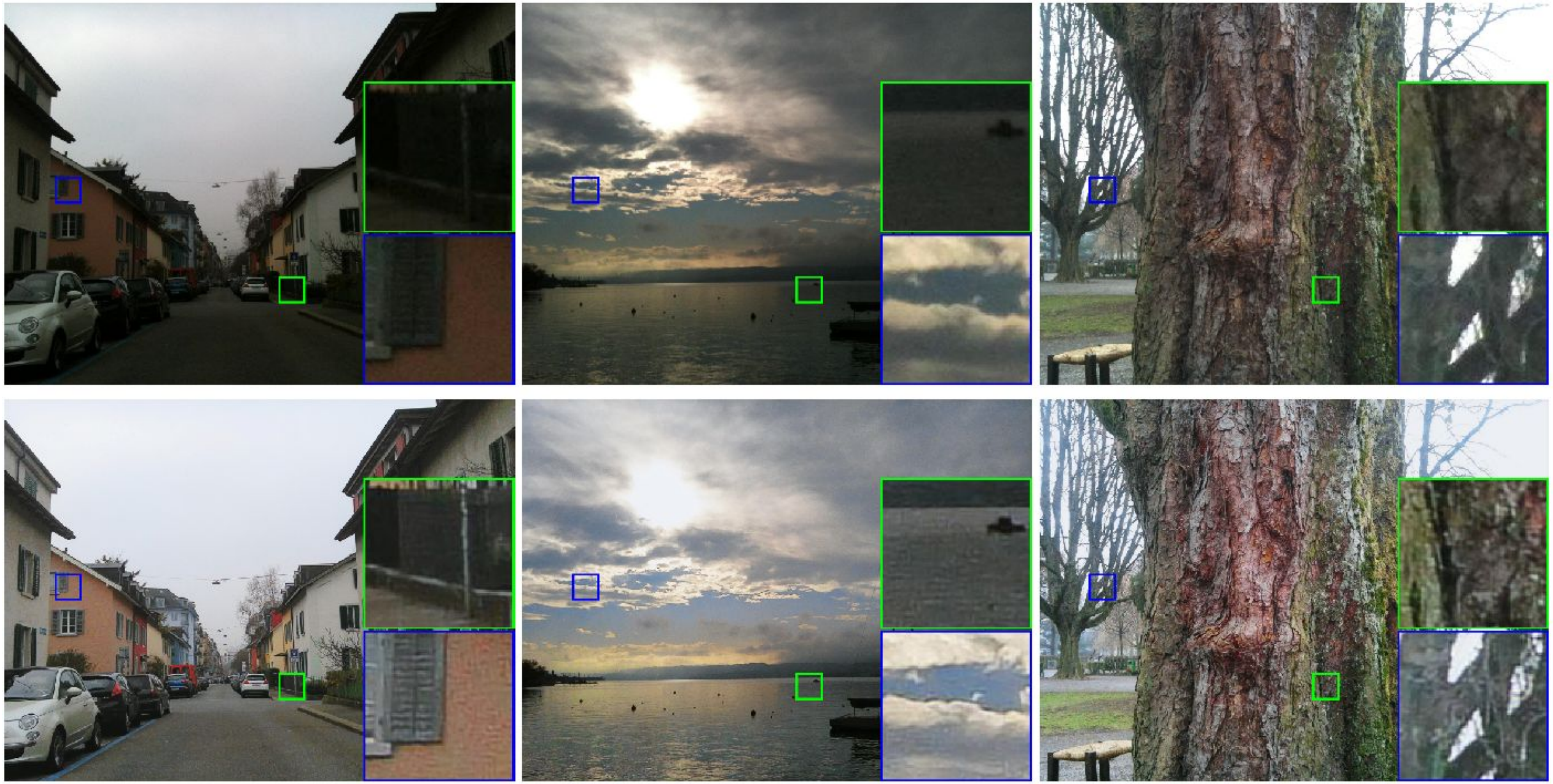
$$\mathcal{L}_{\text{content}} = \frac{1}{C_j H_j W_j} \|\psi_j(F_{\mathbf{W}}(I_s)) - \psi_j(I_t)\|$$

$$\mathcal{L}_{\text{tv}} = \frac{1}{CHW} \|\nabla_x F_{\mathbf{W}}(I_s) + \nabla_y F_{\mathbf{W}}(I_s)\|$$

$$\mathcal{L}_{\text{texture}} = -\sum_i \log D(F_{\mathbf{W}}(I_s), I_t)$$

DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks, Ignatov et al., ICCV 2017.

DSLR Photo Enhancement (DSLR-PE)



Typical artifacts generated by our method (bottom) compared with original iPhone images (top)

DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks, Ignatov et al., ICCV 2017.

Weakly-Supervised Deep Learning Methods for Compound Image Quality Mapping

Weakly Supervised Photo Enhancer (WESPE)

-Content Consistency Loss

$$\mathcal{L}_{\text{content}} = \frac{1}{C_j H_j W_j} \|\psi_j(x) - \psi_j(\tilde{x})\|,$$

-Adversarial Color Loss

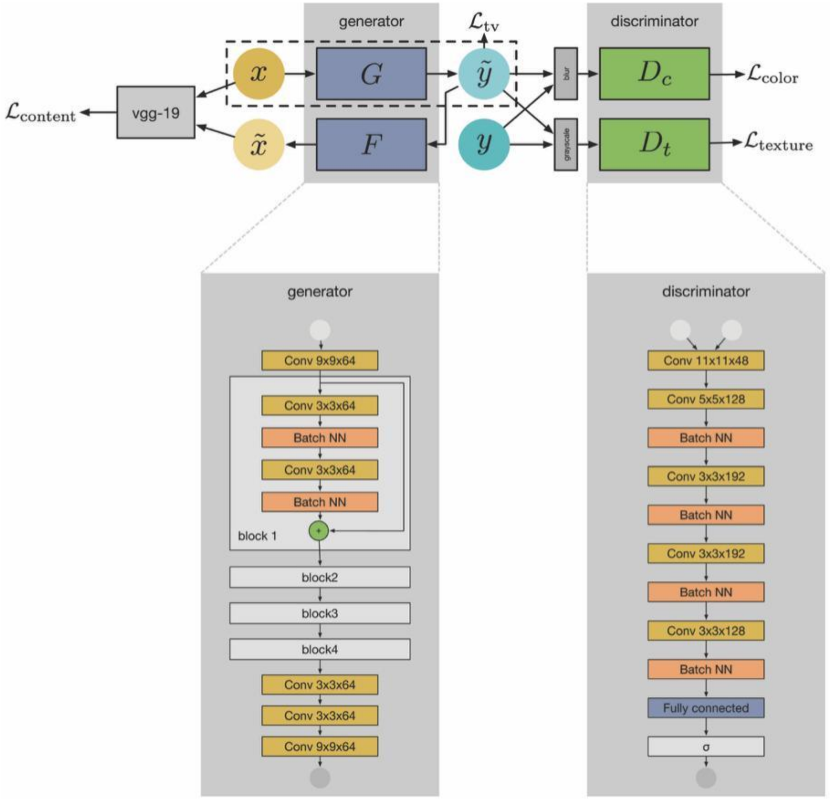
$$\mathcal{L}_{\text{color}} = -\sum_i \log D_c(G(x)_b).$$

-Adversarial Texture Loss

$$\mathcal{L}_{\text{texture}} = -\sum_i \log D_t(G(x)_g).$$

-Total Variation Loss

$$\mathcal{L}_{\text{tv}} = \frac{1}{CHW} \|\nabla_x G(x) + \nabla_y G(x)\|,$$



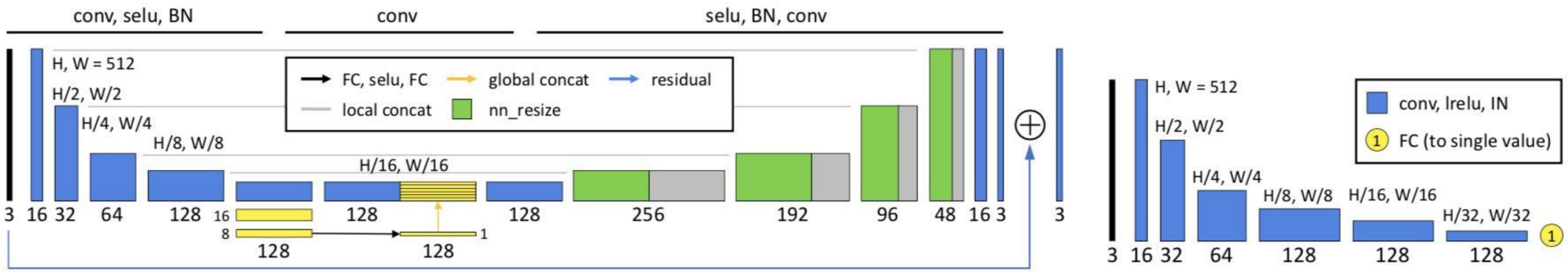
WESPE: weakly supervised photo enhancer for digital cameras, Ignatov et al., CVPRW 2018.

Weakly Supervised Photo Enhancer（WESPE）

DPED images	APE		Weakly Supervised WESPE [DIV2K]		Weakly Supervised WESPE [DPED]		Fully Supervised [13]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
iPhone	17.28	0.86	17.76	0.88	18.11	0.90	**21.35**	**0.92**
BlackBerry	18.91	0.89	16.71	0.91	16.78	0.91	**20.66**	**0.93**
Sony	19.45	0.92	20.05	0.89	20.29	0.93	**22.01**	**0.94**

WESPE: weakly supervised photo enhancer for digital cameras, Ignatov et al., CVPRW 2018.

Deep Photo Enhancer (DPE)



-Identity Mapping Loss:

$$I = \mathop{\mathbb{E}}_{x,y'}\Big[MSE(x,y')\Big] + \mathop{\mathbb{E}}_{y,x'}\Big[MSE(y,x')\Big].$$

-Cycle-Consistency Loss:

$$C = \mathop{\mathbb{E}}_{x,x''}\Big[MSE(x,x'')\Big] + \mathop{\mathbb{E}}_{y,y''}\Big[MSE(y,y'')\Big].$$

-Adversarial Loss:

$$\begin{aligned} A_D = {} & \mathop{\mathbb{E}}_{x}\Big[D_X(x)\Big] - \mathop{\mathbb{E}}_{x'}\Big[D_X(x')\Big] + \\ & \mathop{\mathbb{E}}_{y}\Big[D_Y(y)\Big] - \mathop{\mathbb{E}}_{y'}\Big[D_Y(y')\Big], \\ A_G = {} & \mathop{\mathbb{E}}_{x'}\Big[D_X(x')\Big] + \mathop{\mathbb{E}}_{y'}\Big[D_Y(y')\Big]. \end{aligned}$$

Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs, *Chen et al., CVPR 2018.*

Deep Photo Enhancer (DPE)

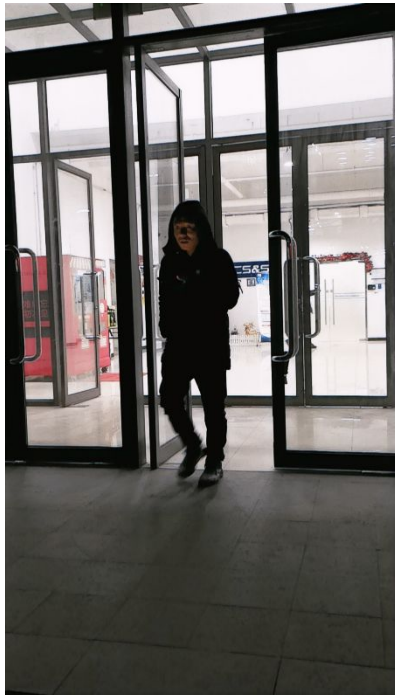
Input Enhanced by DPE

	CycleGAN	DPED	NPEA	CLHE	ours	total
CycleGAN	-	32	27	23	11	93
DPED	368	-	141	119	29	657
NPEA	373	259	-	142	50	824
CLHE	377	281	258	-	77	993
ours	389	371	350	323	-	1433

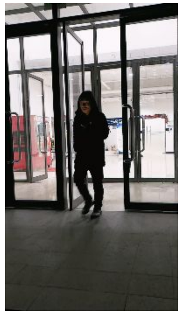
Preference matrix from AMT user study

Deep photo enhancer: Unpaired learning for image enhancement from photographs with GANs, *Chen et al., CVPR 2018.*

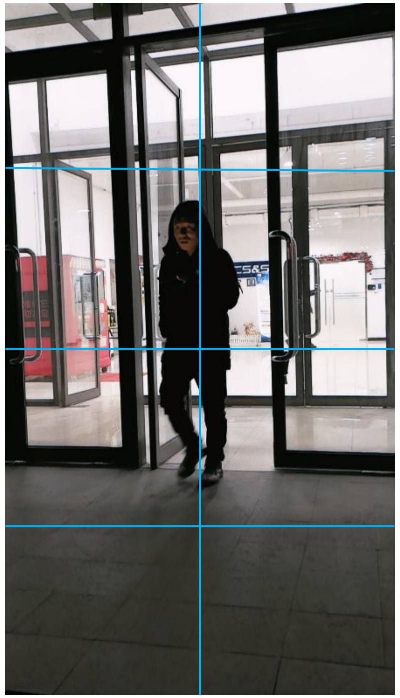
Limitation for Compound Quality Mapping and High-Resolution Image Treatment



Input



Downscaling
(low-res, noisy, blurry)



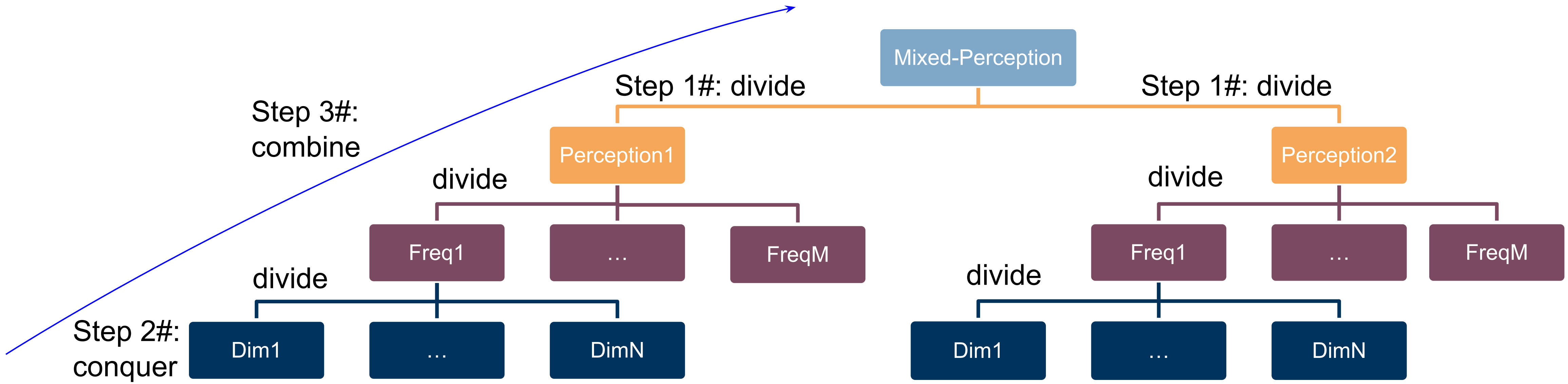
Patch-wise Enhancement
(spatial inconsistency)

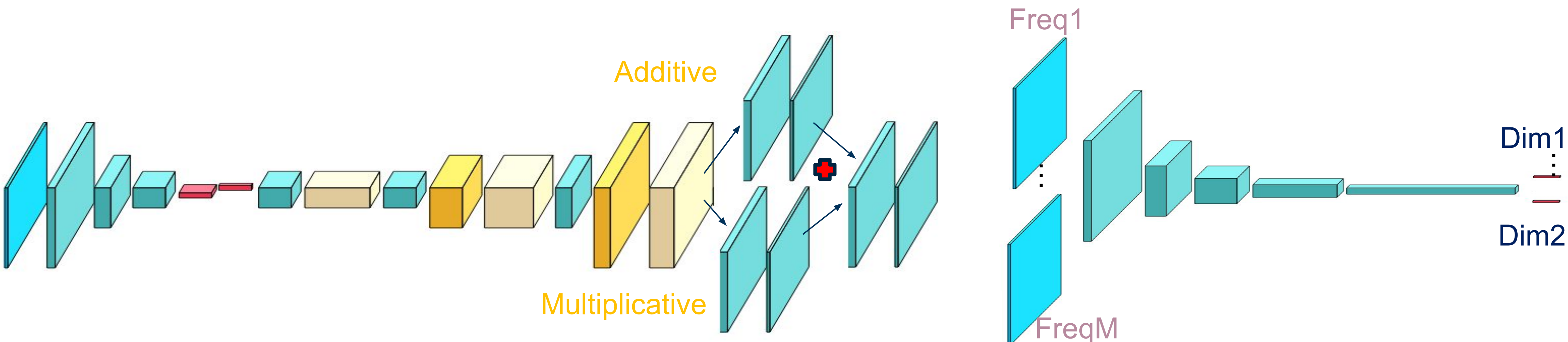
Model	Limitation (Compound Quality)	Limitation (High Resolution)
WESPE	Color and Texture(Not sufficient)	Patch-wise Enhancement
DPE	No consideration on mixed-perceptual improvement	Down-scaling

Divide-and-Conquer Adversarial Learning for High-resolution Image and Video Enhancement

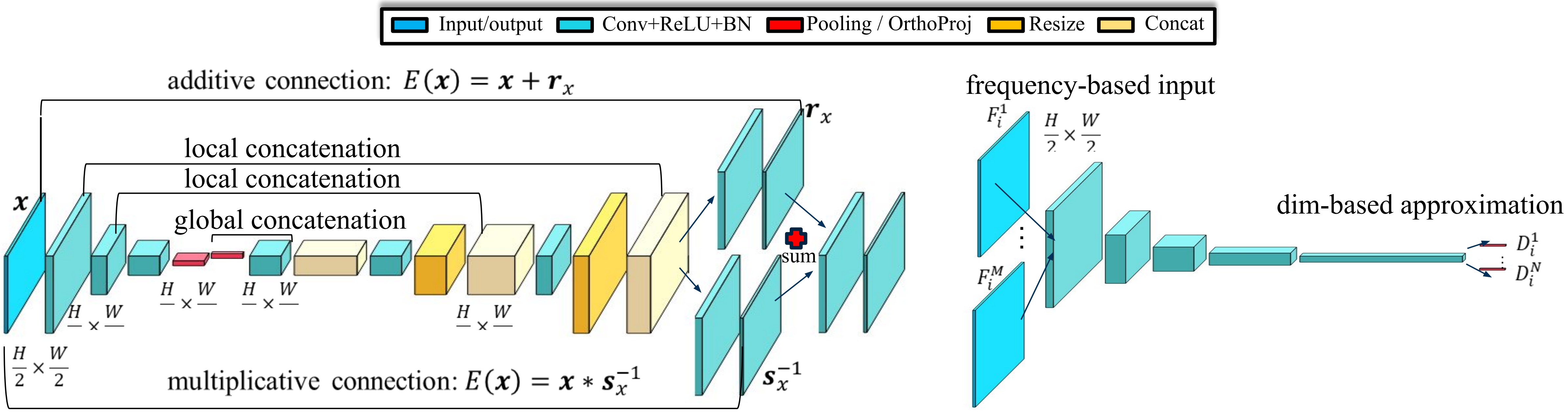
Zhiwu Huang, Danda Pani Paudel, Guanju Li, JiqingWu, Radu Timofte, Luc Van Gool, arXiv preprint arXiv:1910.10455.

Divide-and-Conquer Inspired Method





Network Design



(a) Enhancer for Peception-based Division

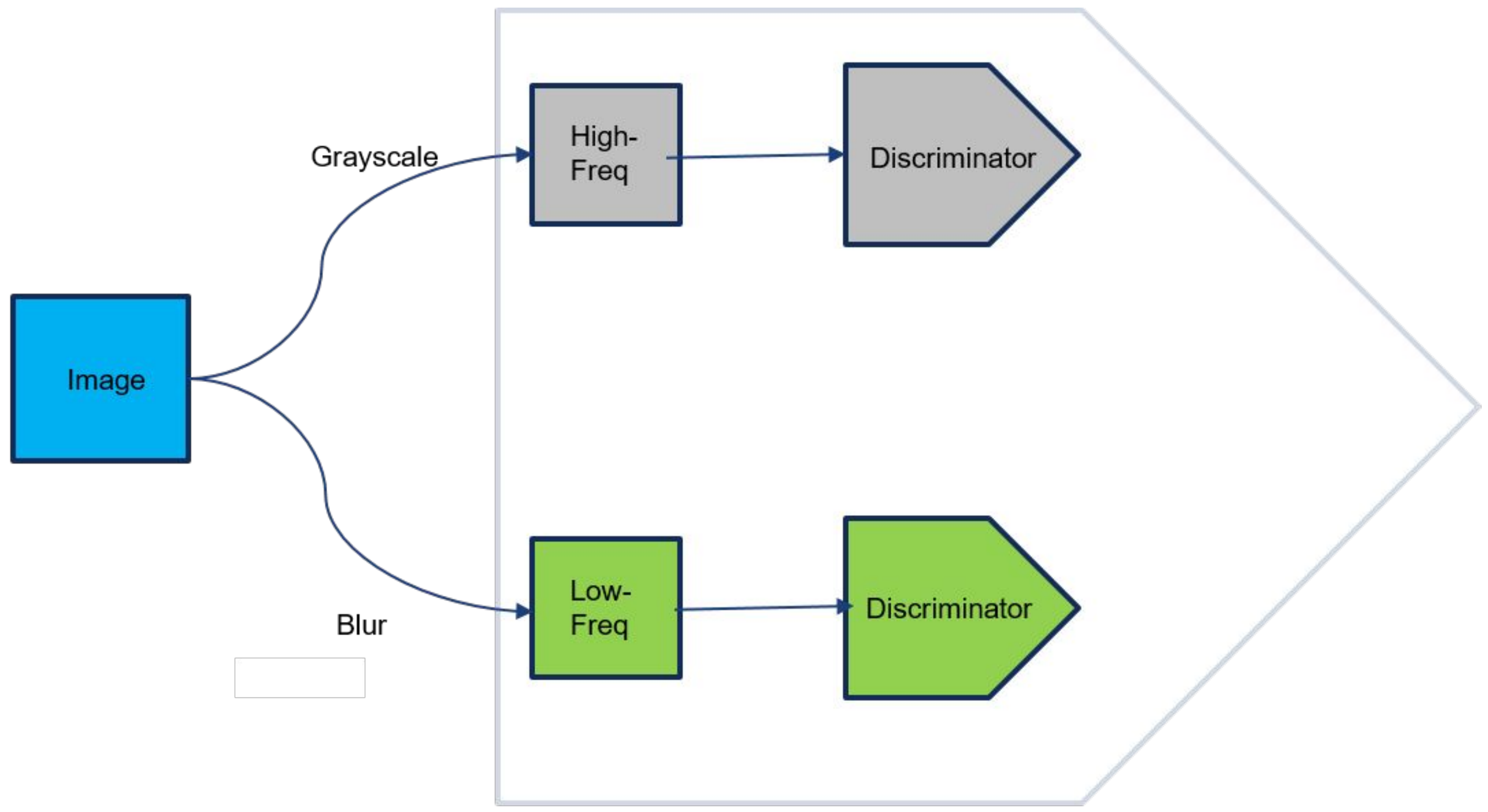
(b) Discriminator for Freq- and Dim-based Division

Perception-based Division

(a) Input (b) Additive component (c) Multiplicative component

(d) Additive map (e) Multiplicative map (f) Fused Map

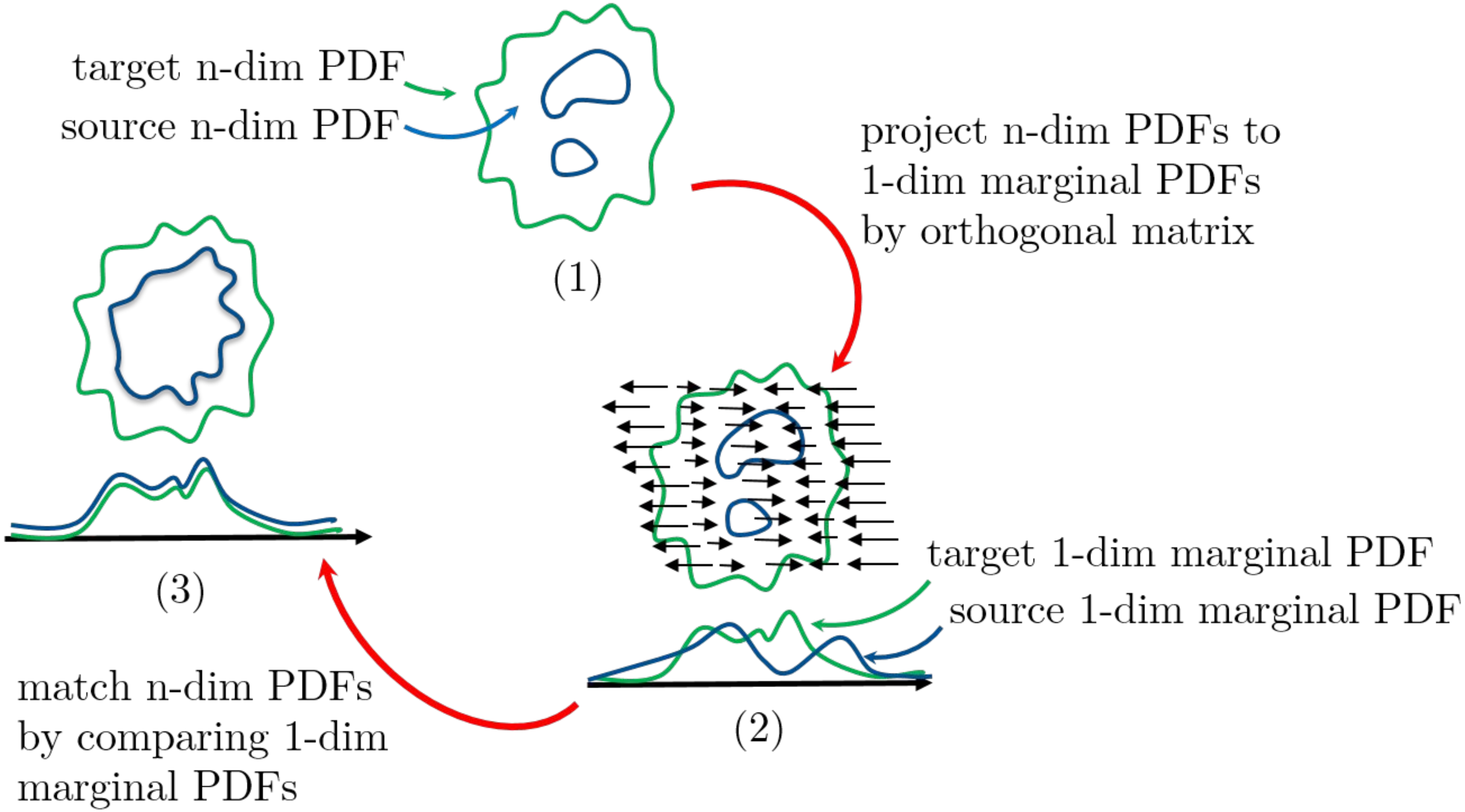
Frequency-based Division



[1] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks.
[2] Anoosheh, Asha, et al. "Night-to-day image translation for retrieval-based localization."

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

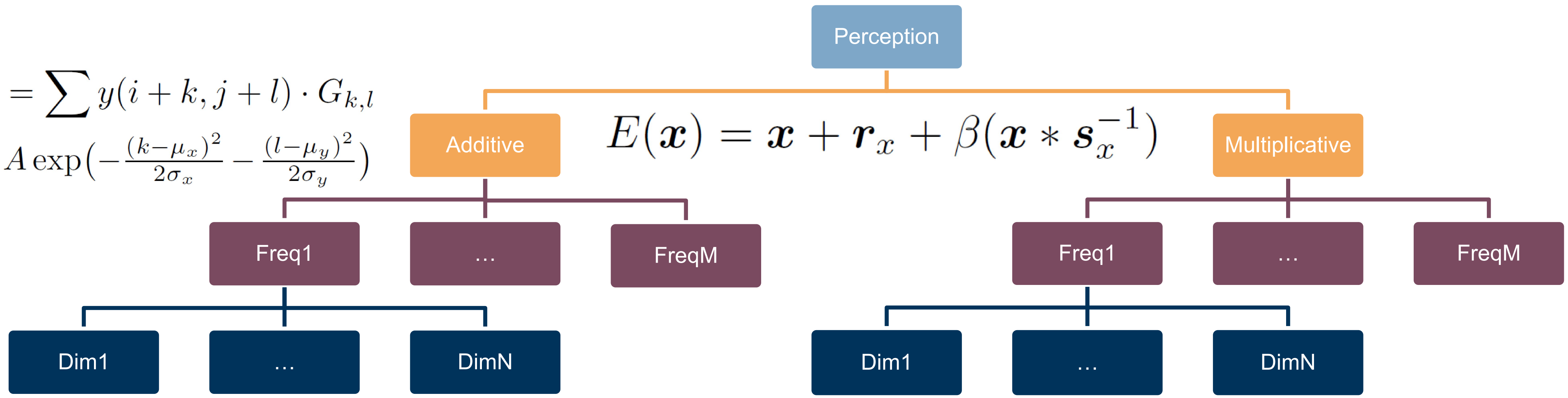
Dimension-based Division and Optimization



Loss Design

$$y_b(i,j)=\sum y(i+k,j+l)\cdot G_{k,l}$$

$$G_{k,l}=A\exp\left(-\frac{(k-\mu_x)^2}{2\sigma_x}-\frac{(l-\mu_y)^2}{2\sigma_y}\right)$$



$$E(\boldsymbol{x})=\boldsymbol{x}+\boldsymbol{r}_x+\beta(\boldsymbol{x}*\boldsymbol{s}_x^{-1})$$

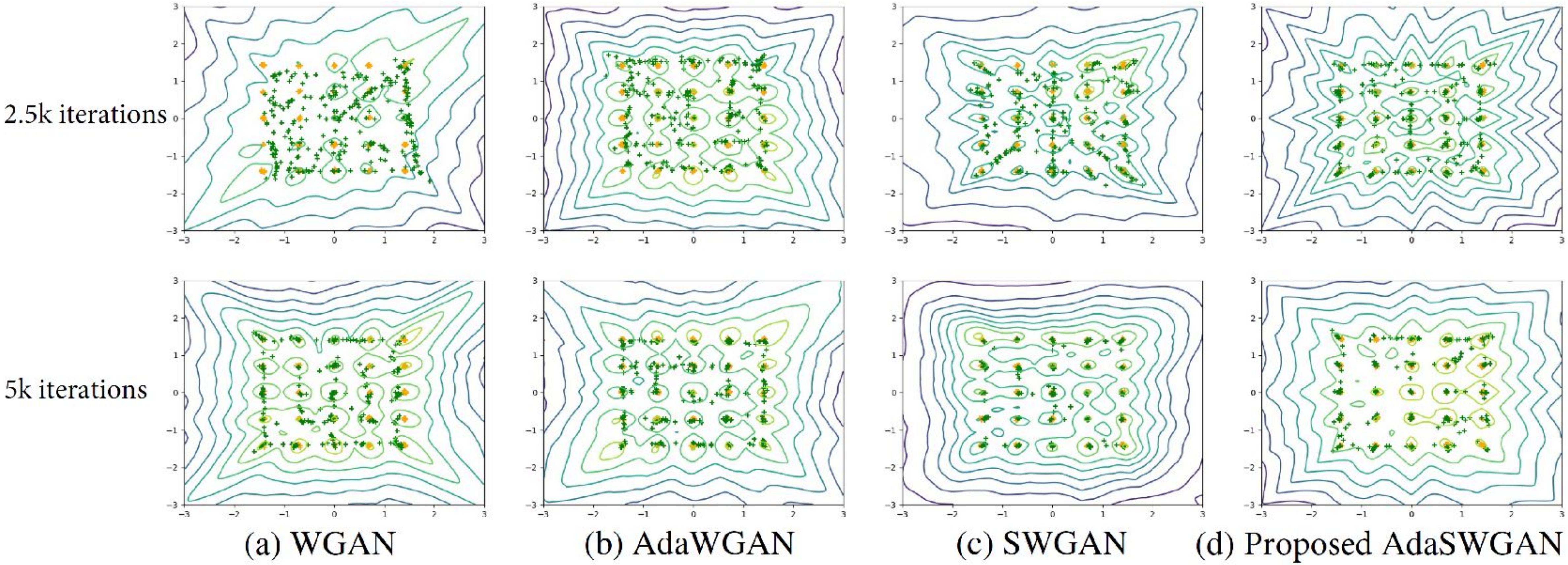
Adaptive SWGAN loss

$$\min_E\max_C\int_{\theta\in\mathbb{S}^{n-1}}\left(\mathbb{E}_{\boldsymbol{y}\sim P_y}[C(\boldsymbol{y})]-\mathbb{E}_{\tilde{\boldsymbol{y}}\sim P_E}[C(E(\boldsymbol{x}))]\right)+\lambda\mathbb{E}_{\hat{\boldsymbol{y}}\sim P_{\hat{\boldsymbol{y}}}}[\max(0,\boxed{\|\nabla_{\hat{\boldsymbol{y}}}C(\hat{\boldsymbol{y}})\|_2}-1)]$$

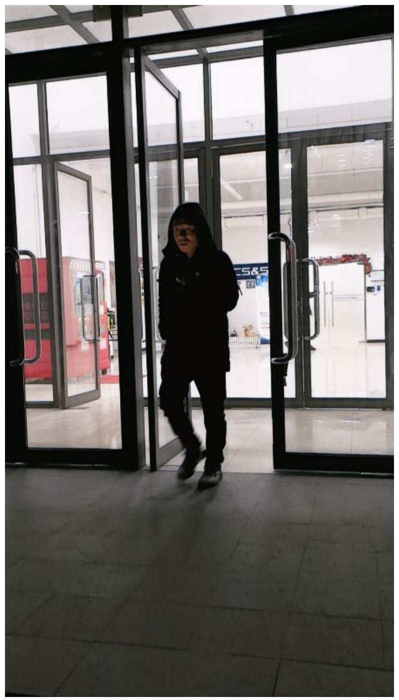
$$\overline{\nabla_{\hat{y}}C_(\hat{\boldsymbol{y}})}=\eta\overline{\nabla_{\hat{y}}C(\hat{\boldsymbol{y}})}+(1-\eta)\boxed{\frac{\nabla_{\hat{y}}C(\hat{\boldsymbol{y}})}{\lambda}}$$

Adaptive Penalty

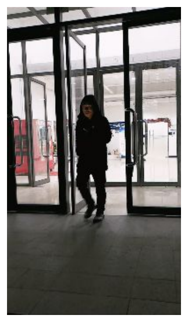
Evaluation on Toy Data



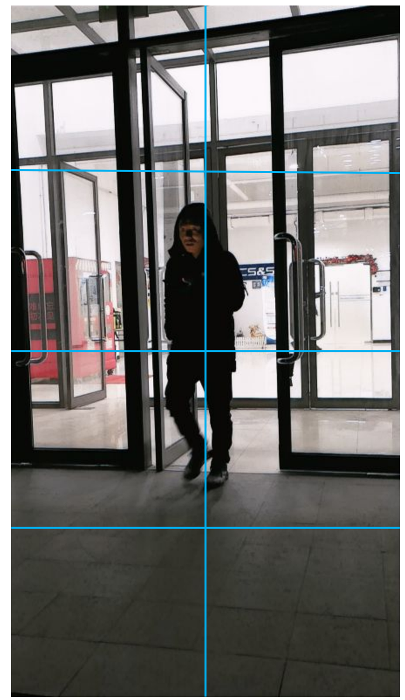
High-resolution Issue for Image Enhancement



Input



Downscaling
(low-res, noisy, blurry)
Deep Photo Ehancer (DPE) [Chen et al in CVPR'18]



Patch-wise Enhancement
(spatial inconsistency)
Weakly Supervised Photo Enhancer (WESPE) [our work in CVPR'18 workshop

coarse

fine

Multi-scale Photo Enhancement (MUSPE)
Our current work

Multi-scale Extension of DACAL for Image Enhancement

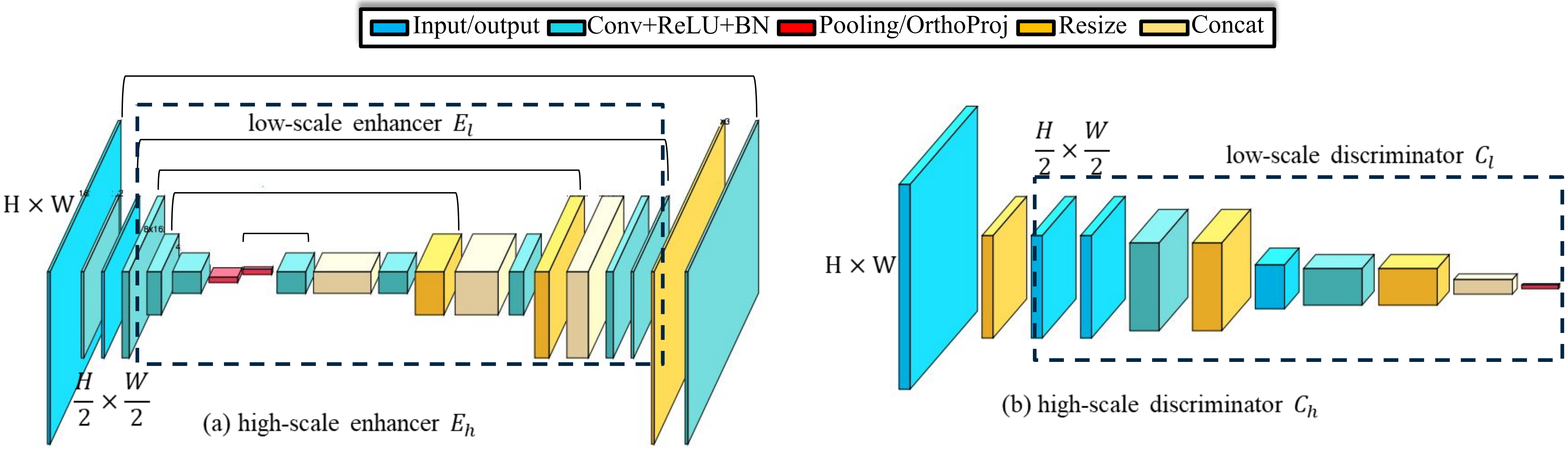


Table 1: PSNR and SSIM results for the MIT-Adobe FiveK [42] test images. Here, WB and DR indicate the White-Box and Distort-and-Recover methods, respectively. MUSPE$_{l_1}$, MUSPE$_{l_2}$, MUSPE$_{l_3}$ and MUSPE$_l$ represent the use of individual additive, individual multiplicative, multiplicative cascaded by additive, and our suggested parallel fusion (two-stream strategy), respectively. MUSPE$_h$ is our higher-scale version. PSNR$_d$/SSIM$_d$ and PSNR$_f$/SSIM$_f$ indicate the results on downscaled images and full-resolution images, respectively.

	WB	DR	DPED	DPE	MUSPE$_{l_1}$	MUSPE$_{l_2}$	MUSPE$_{l_3}$	MUSPE$_l$	MUSPE$_h$
PSNR$_d$	18.86	21.64	21.05	22.10	22.73	22.99	23.01	23.52	**24.15**
PSNR$_f$	19.09	21.52	20.86	21.65	22.43	22.69	23.02	23.56	**24.07**
SSIM$_d$	0.928	0.936	0.922	0.947	0.958	0.942	0.949	0.959	**0.962**
SSIM$_f$	0.920	0.922	0.916	0.894	0.948	0.942	0.940	0.954	**0.956**

Table 2: PSNR and SSIM results for the DPED [14] test 100×100 image patches. Here, l, f, d for MUSPE represent the use of our proposed sliced-perception, sliced-frequency and sliced-dimension learning respectively. MUSPE$_h$ is our higher-scale version.

	WESPE	DPE	MUSPE$_l$	MUSPE$_{l+f}$	MUSPE$_{l+f+d}$	MUSPE$_h$
PSNR$_{100}$	17.45	18.53	19.62	20.01	20.43	**20.90**
SSIM$_{100}$	0.854	0.861	0.868	0.869	0.872	**0.874**

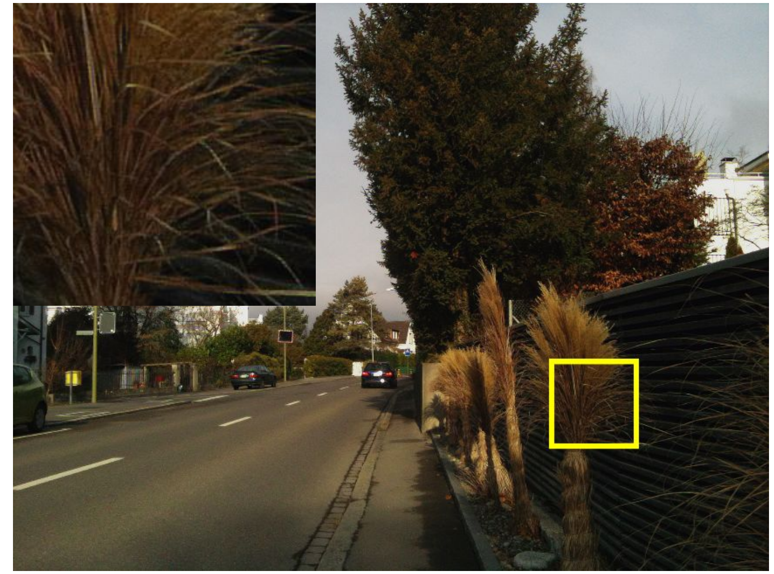
Input

WESPE [Ignatov, CVPRW'18]

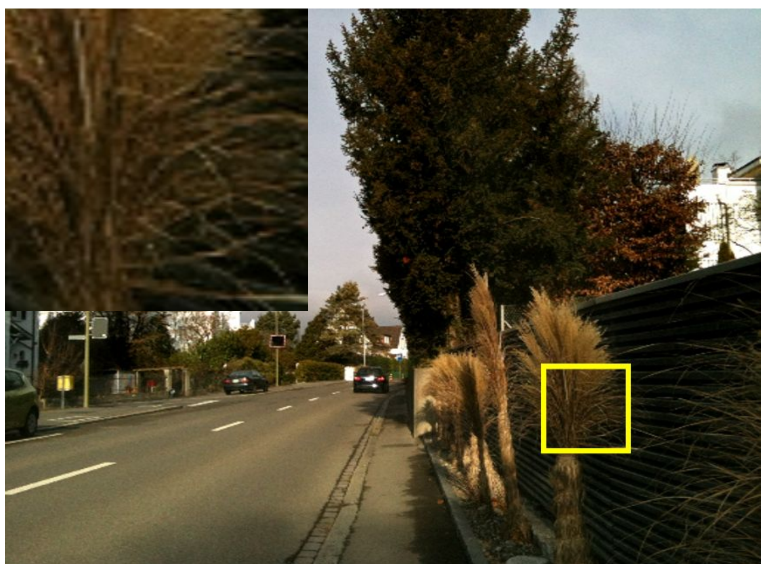
DPE [Chen, CVPR'18]

Proposed MUSPE [ICLR'20 submission]

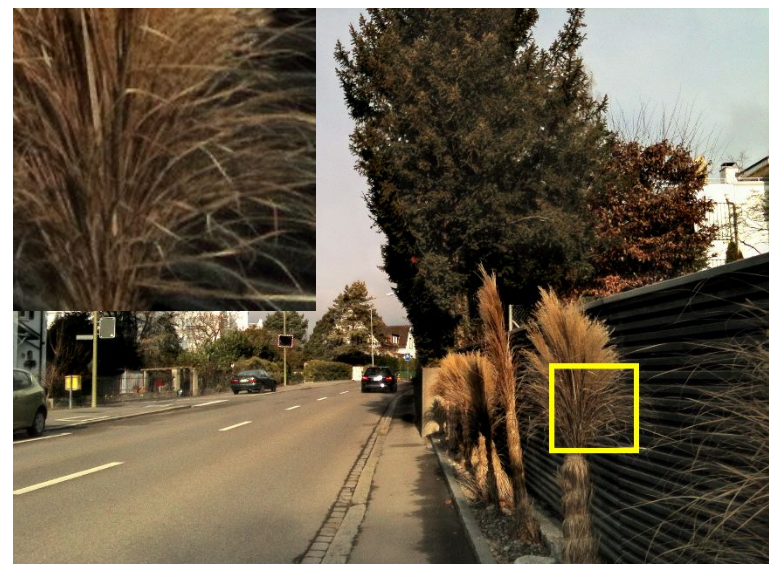
Input



WESPE [Ignatov, CVPRW’18]



DPE [Chen, CVPR’18]

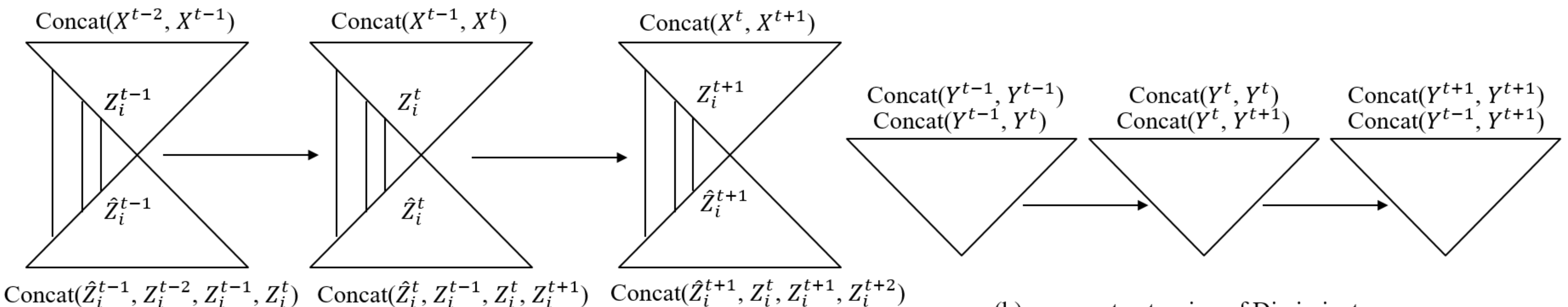


Proposed MUSPE [ICLR’20 submission]

Input | WESPE [Ignatov, CVPRW'18] | DPE [Chen, CVPR'18] | Proposed MUSPE [ICLR'20 submission]

Video Quality Mapping = Image Quality Mapping + Temporal Smoothing

Recurrent Extension of DACAL for Video Enhancement



(b) recurrent extension of Disriminator

(a) recurrent extension of Generator

Perframe-DACAL

Recurrent-DACAL

Perframe-DACAL

Recurrent-DACAL (fine-tuned on Retouched&DSLR images)

Perframe-DACAL

Recurrent-DACAL

Perframe-DACAL

Recurrent-DACAL (fine-tuned on Retouched&DSLR images)

Conclusion

- Supervision
  - Weak-supervision is cheaper
- Compound quality mapping
  - Divide-and-conquer inspired algorithm is promising
- High-resolution image treatment
  - Multiscaled training is helpful
- Video enhancement
  - Recurrent model works well