Paper Review

Old Photo Restoration via Deep Latent Space Translation

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University

Wan, Ziyu, et al. "Old photo restoration via deep latent space translation." IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2022.

Old Photo Restoration via Deep Latent Space Translation

Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Fang Wen, Jing Liao

Abstract—We propose to restore old photos that suffer from severe degradation through a deep learning approach. Unlike conventional restoration tasks that can be solved through supervised learning, the degradation in real photos is complex and the domain gap between synthetic images and real old photos makes the network fail to generalize. Therefore, we propose a novel triplet domain translation network by leveraging real photos along with massive synthetic image pairs. Specifically, we train two variational autoencoders (VAEs) to respectively transform old photos and clean photos into two latent spaces. And the translation between these two latent spaces is learned with synthetic paired data. This translation generalizes well to real photos because the domain gap is closed in the compact latent space. Besides, to address multiple degradations mixed in one old photo, we design a global branch with a partial nonlocal block targeting the structured defects, such as scratches and dust spots, and a local branch targeting the unstructured defects, such as noises and blurriness. We also extend the global branch with a more memory-efficient scheme, named multi-scale patch-based attention to processing high-resolution photos. Two branches are fused in the latent space, leading to improved capability to restore old photos from multiple defects. Furthermore, we apply another face refinement network to recover fine details of faces in the old photos, thus ultimately generating photos with enhanced perceptual quality. With comprehensive experiments, the proposed pipeline demonstrates superior performance over state-of-the-art methods as well as existing commercial tools in terms of visual quality for old photos restoration. Both code and models could be found here.

Index Terms—Image Restoration, Image Generation, Latent Space Translation, Mixed degradation

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

Main purpose

- **Unstructured defects:** noise, blurriness, color fading, and low resolution
- **Structured defects:** holes, scratches, and spots

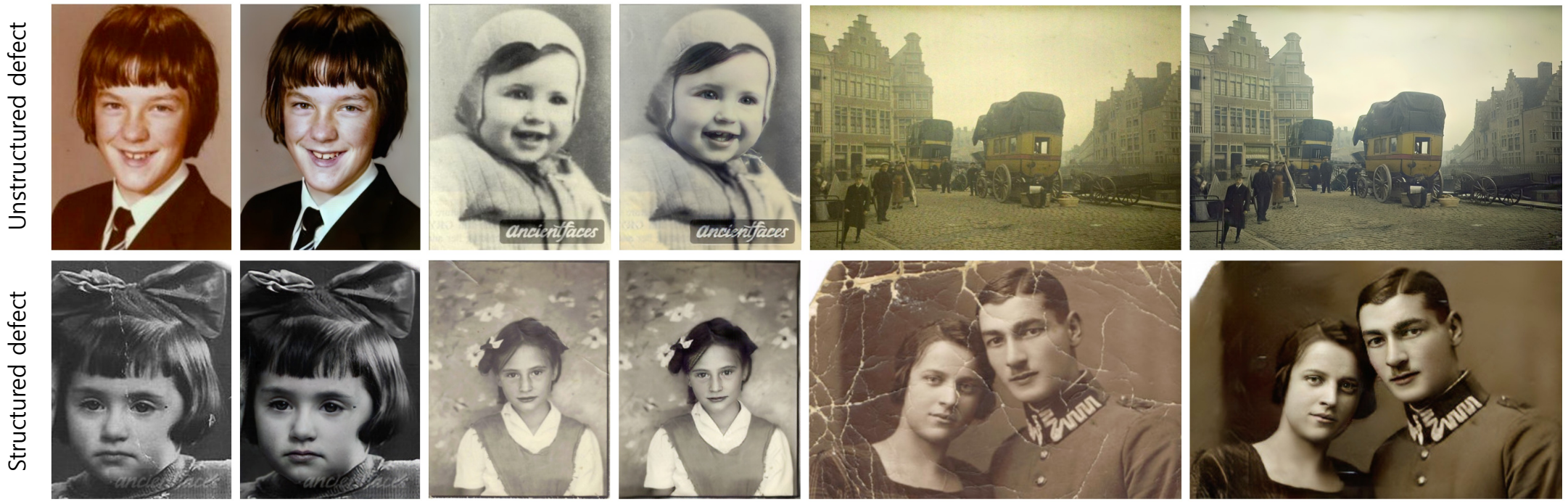


Fig. 1: **Old photo restoration results produced by our method.** Our method can handle the complex degradation mixed with both unstructured and structured defects in real old photos. In particular, we recover high-frequency details for face regions, further improving the perceptual quality for portraits. For each image pair, left is the input while the retouched output is shown on the right.

Bring old photos back to life!

Wan, Ziyu, et al. "Bringing old photos back to life." proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020.
Project page: http://raywzy.com/Old_Photo

Prior Work

CVPRVIRTUAL

Bringing Old Photos Back to Life

Ziyu Wan[1*], Bo Zhang[2], Dongdong Chen[3], Pan Zhang[4], Dong Chen[2], Jing Liao[1†], Fang Wen[2]
[1]City University of Hong Kong [2]Microsoft Research Asia [3]Microsoft Cloud + AI
[4]University of Science and Technology of China

Figure 1: **Old image restoration results produced by our method.** Our method can handle the complex degradation mixed by both unstructured and structured defects in real old photos.

About First Author

Ziyu Wan

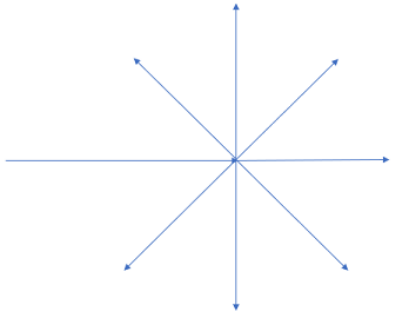
FOLLOW

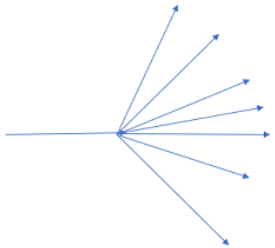
City University of Hong Kong
Verified email at my.cityu.edu.hk - Homepage
Computer Vision Computational Photography Computer Graphics

TITLE	CITED BY	YEAR
Bringing Old Photos Back to Life Z Wan, B Zhang, D Chen, P Zhang, D Chen, J Liao, F Wen IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2747-2757	76	2020
Transductive Zero-Shot Learning with Visual Structure Constraint Z Wan, D Chen, Y Li, X Yan, J Zhang, Y Yu, J Liao Advances in Neural Information Processing Systems (NeurIPS)	56	2019
PD-GAN: Probabilistic Diverse GAN for Image Inpainting H Liu, Z Wan, W Huang, Y Song, X Han, J Liao IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	55	2021
High-Fidelity Pluralistic Image Completion with Transformers Z Wan, J Zhang, D Chen, J Liao IEEE/CVF International Conference on Computer Vision (ICCV)	44	2021
Meta-PU: An arbitrary-scale upsampling network for point cloud S Ye, D Chen, S Han, Z Wan, J Liao IEEE Transactions on Visualization and Computer Graphics (TVCG)	19	2021
Old Photo Restoration via Deep Latent Space Translation Z Wan, B Zhang, D Chen, P Zhang, D Chen, J Liao, F Wen IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	16	2022
DeFLOCNet: Deep Image Editing via Flexible Low-level Controls H Liu, Z Wan, W Huang, Y Song, X Han, J Liao, B Jiang, W Liu IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	8	2021

TITLE	CITED BY	YEAR
Transductive Zero-Shot Learning via Visual Center Adaptation Z Wan, Y Li, M Yang, J Zhang Proceedings of the AAAI Conference on Artificial Intelligence 33 (01), 10059 …	3	2019
Visual Structure Constraint for Transductive Zero-Shot Learning in the Wild Z Wan, D Chen, J Liao International Journal of Computer Vision (IJCV) 129 (6), 1893-1909	2	2021
Bringing Old Films Back to Life Z Wan, B Zhang, D Chen, J Liao IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	1	2022
FDNeRF: Few-shot Dynamic Neural Radiance Fields for Face Reconstruction and Expression Editing J Zhang, X Li, Z Wan, C Wang, J Liao		2022
Adaptive Joint Optimization for 3D Reconstruction with Differentiable Rendering J Zhang, Z Wan, J Liao IEEE Transactions on Visualization and Computer Graphics (TVCG)		2022
Supplementary Material-Bringing Old Films Back to Life Z Wan, B Zhang, D Chen, J Liao		

Summary


Isotropic scattering


Anisotropic scattering

Purpose
- Unpaired old (degraded) photo restoration

Contributions
- Domain alignment: well generalization on real photos by training synthetic data
- Partial nonlocal block: reconstruction of blind spots / corruption area
- Multi-scale patch-based fusion: reduction of memory cost and better performance
- Coarse-to-fine generator: reconstruction of the high-resolution face

Limitations
- Anisotropic pollution is not totally restored.
- Shading along the folds of the photo hindrance the restoration process.

Problem Definition

Microsoft Research, Bringing Old Photos Back to Life, https://www.youtube.com/watch?v=Q5bhszQq9eA

Problems

- **Unstructured defects:** noise, blurriness, color fading, and low resolution
- **Structured defects:** holes, scratches, and spots

Unstructured distortion

Noise, blurring, color fading, sepia issue, …

Structured distortion

Scratches, dust, pollution, ….

Single / Mixed degradation image restoration

- Single degradation = unstructured or structured defects
- Mixed degradation = unstructured and structured defects
- Classical methods have over smoothness issue.
- Deep learning methods have a limitation of generalization.
  - The deep neural network learns synthetic data due to unpaired dataset.
  - So, it has a limitation to restore the real photo.

Face restoration

- Wild-portraits have various posed faces.
- Face identity should be preserved while the face restoration task.

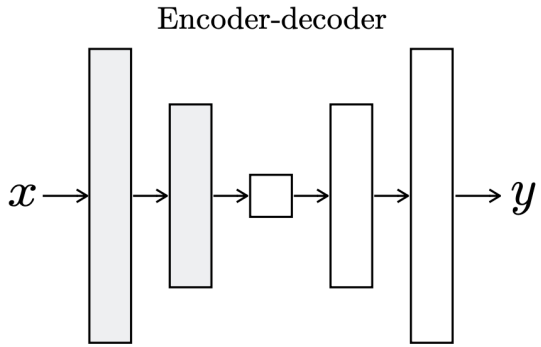
Old photo restoration

- Prior studies focus on inpainting task to remove structured defects.
- Also, those are not consider to restore unstructured defects.

Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2017.
Yu, Jiahui, et al. "Generative image inpainting with contextual attention." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2018.

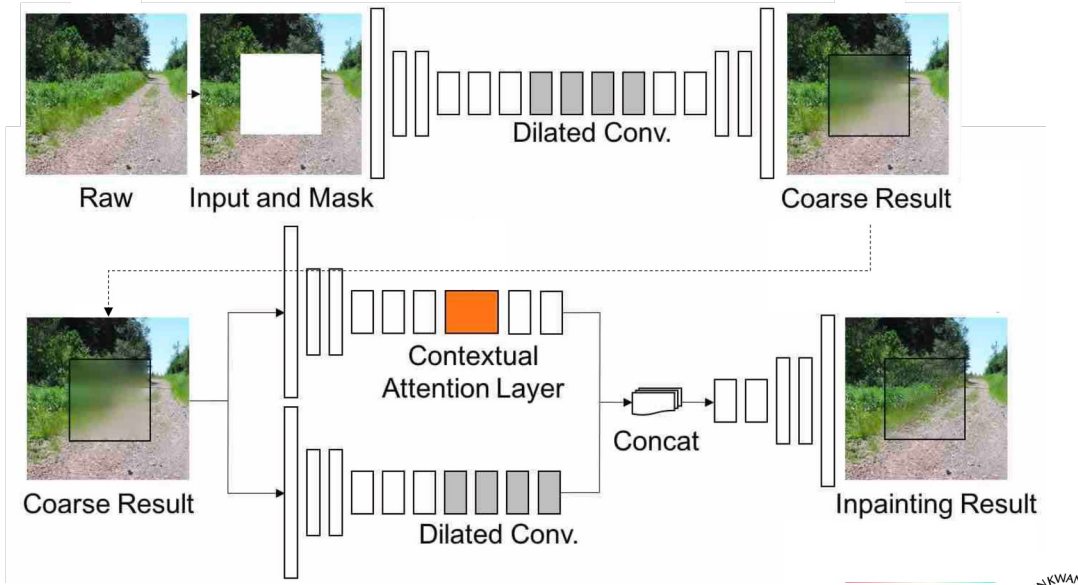
Single degradation image restoration

Pix2Pix



Pix2Pix cannot fill hole.

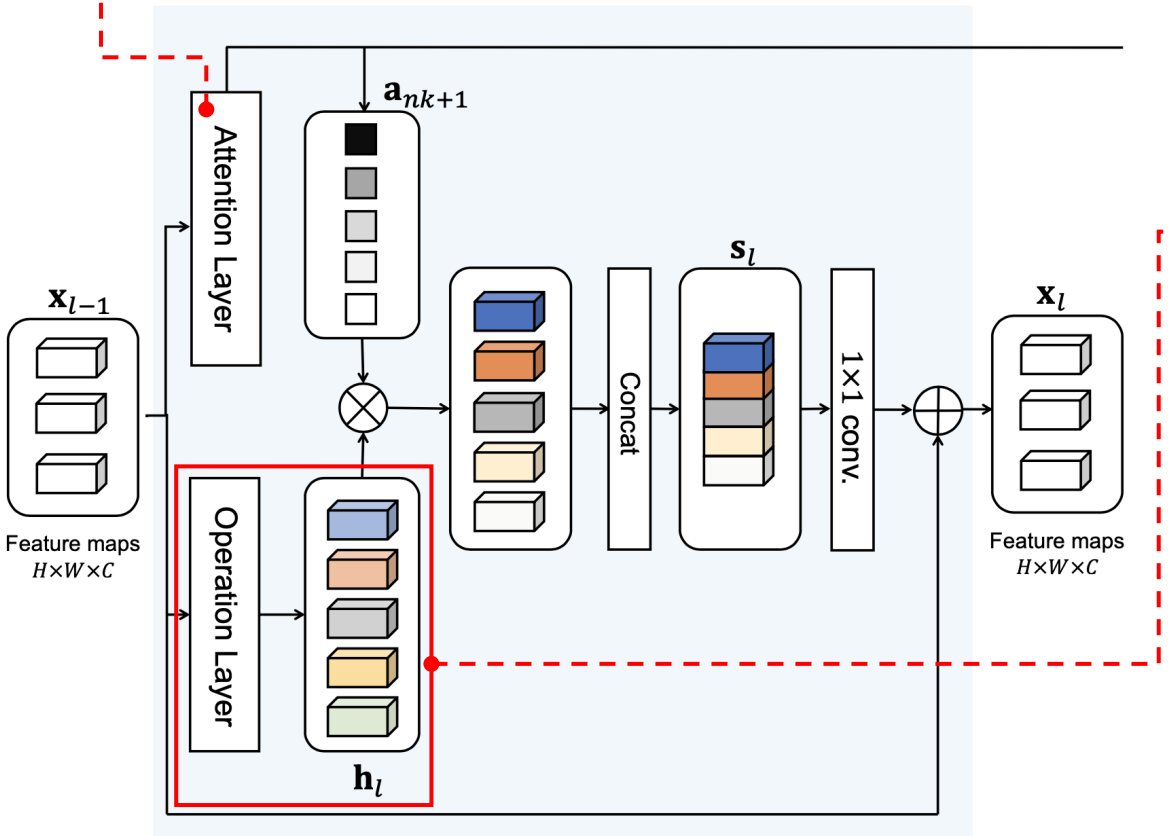
Contextual Attention

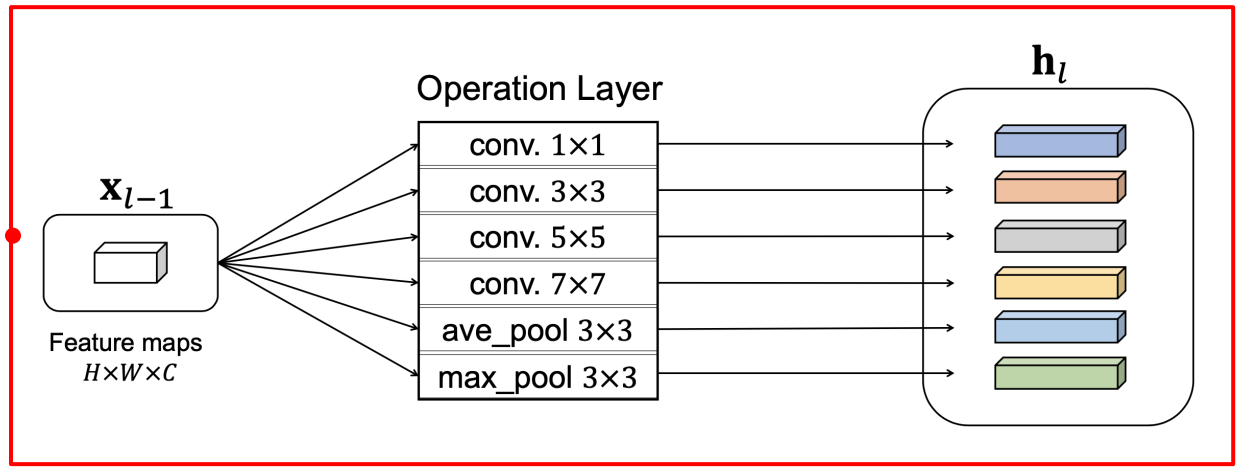


Mixed degradation image restoration [1/2]

Operation-wise attention

- 1 x 1 convolution & global average pool





- k x k convolution (k=1, 3, 5, 7)
- k x k dilated convolution
- 3 x 3 average pool
- 3 x 3 max pool

Mixed degradation image restoration [2/2]

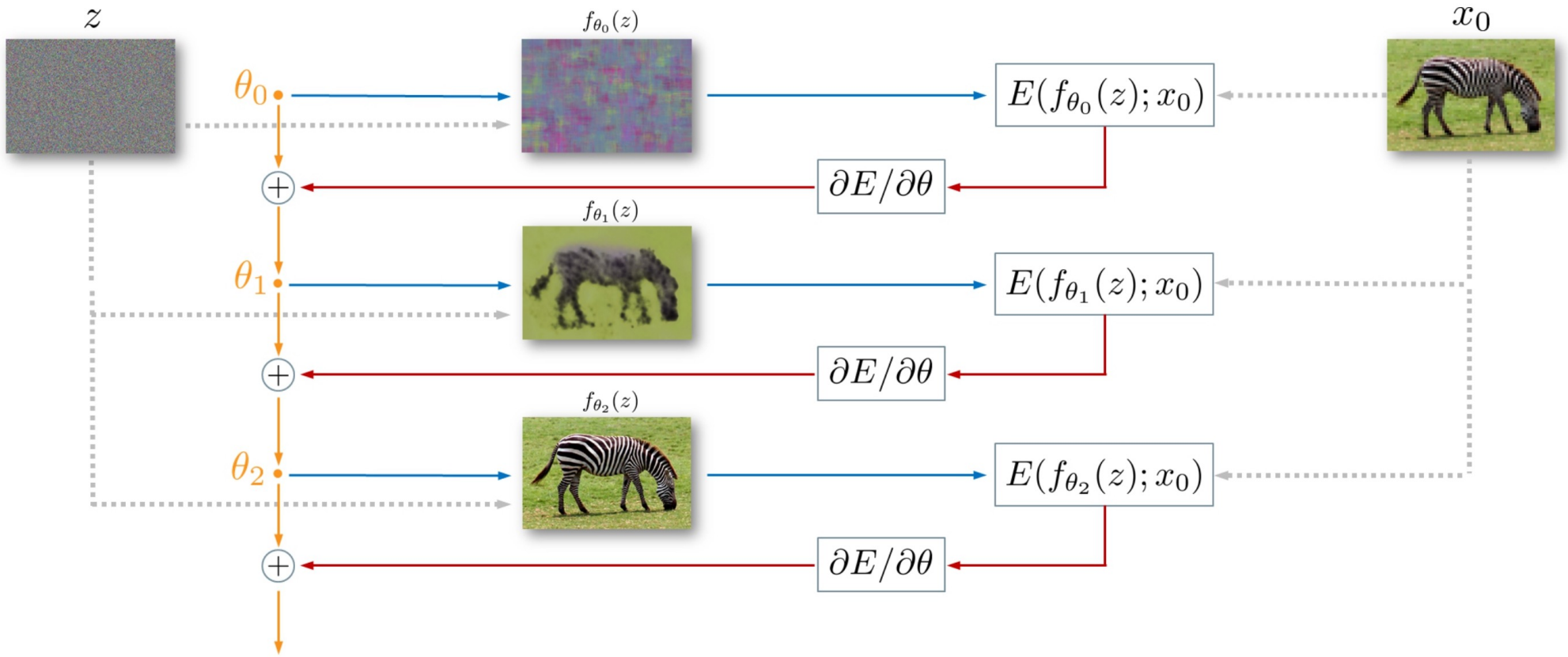


Fig. 2: **Image restoration using the deep image prior.** Starting from a random weights θ_0, we iteratively update them in order to minimize the data term eq. (2). At every iteration the weights θ are mapped to an image $x = f_\theta(z)$, where z is a fixed tensor and the mapping f is a neural network with parameters θ. The image x is used to compute the task-dependent loss $E(x, x_0)$. The gradient of the loss w.r.t. the weights θ is then computed and used to update the parameters.

Old Photo Restoration Model

Wan, Ziyu, et al. "Bringing old photos back to life." proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020.

Overview

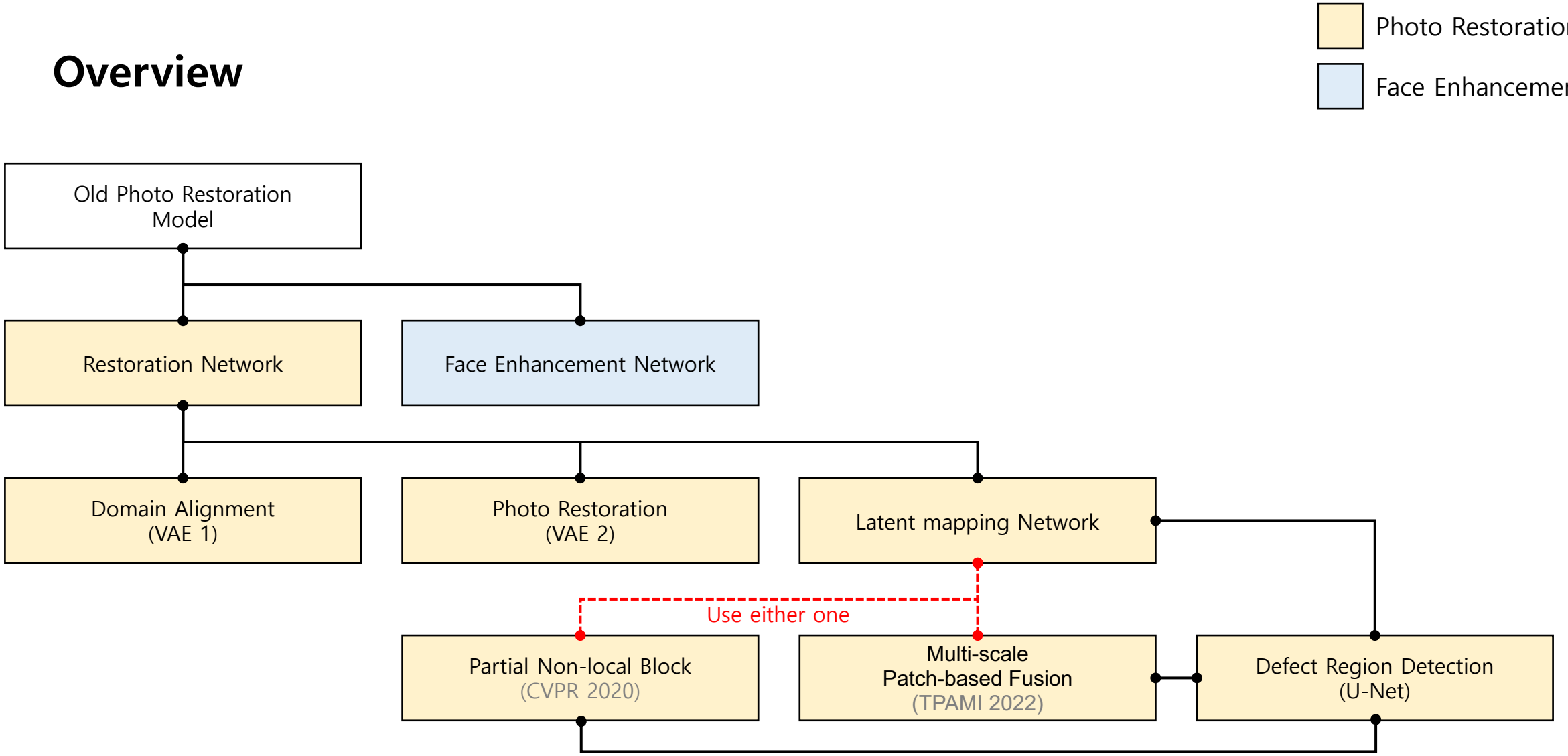


Photo Restoration [1/11]

Latent Space Translation (Intersection Maximization)

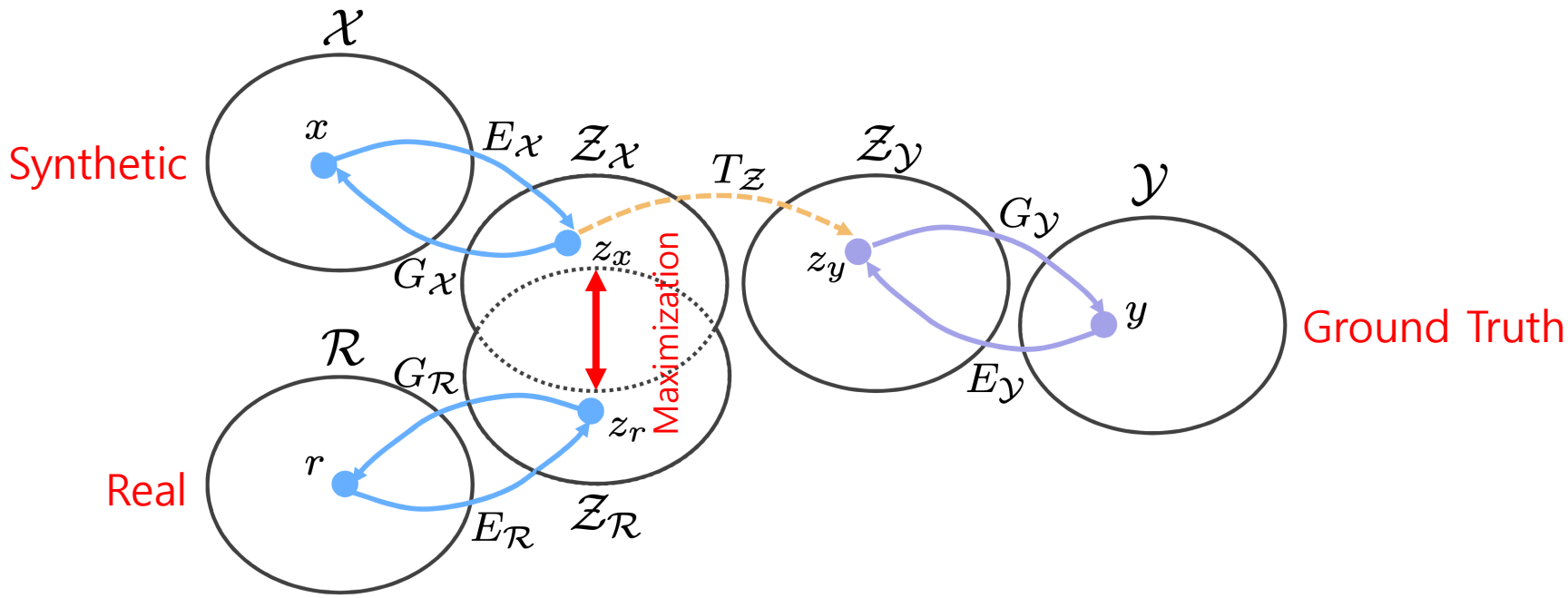


Fig. 2: **Illustration of our translation method with three domains.** The domain gap between $\mathcal{Z}_{\mathcal{X}}$ and $\mathcal{Z}_{\mathcal{R}}$ will be reduced in the shared latent space.

Photo Restoration [2/11]

Out-of-Distribution (Intersection Minimization)

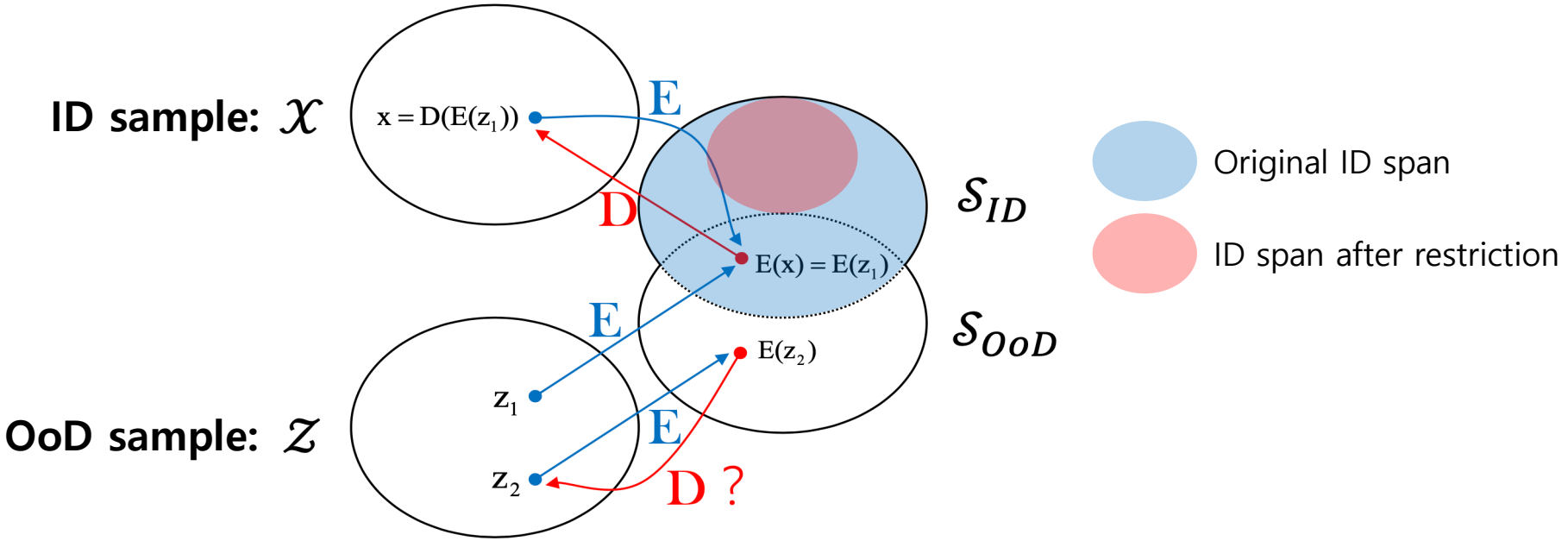
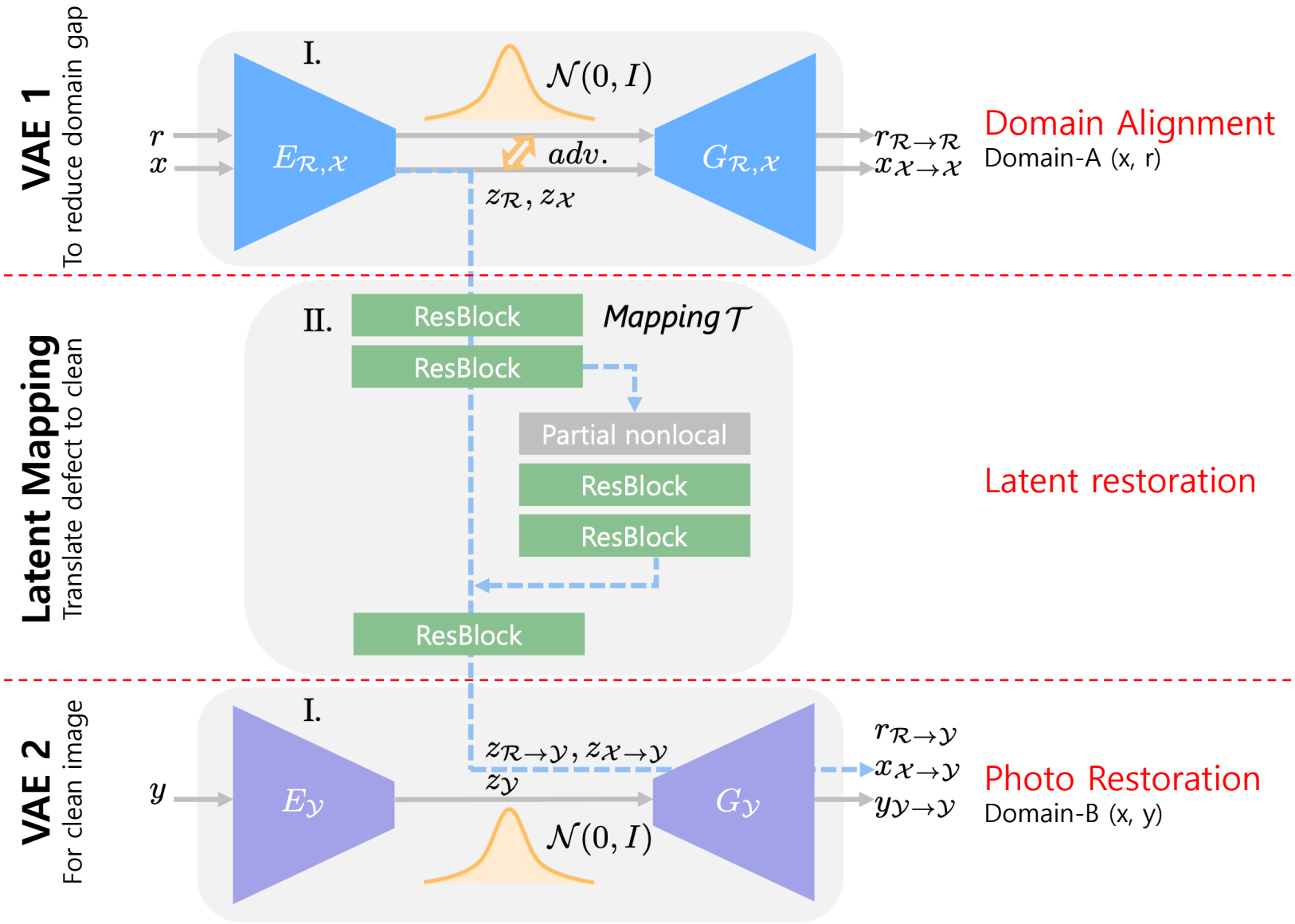


Figure 1. Illustration of the described quadruplet domain translation. For an OoD sample z_1 encoded into $\mathcal{S}_{ID} \cap \mathcal{S}_{OoD}$, its latent representation $E(z_1)$ is equal potentially to that of an ID sample x. Therefore, $E(z_1)$ can be decoded to a different sample x within $\mathcal{X}$, resulting in a large reconstruction error. However, for an OoD sample z_2 with latent representation $E(z_2)$ lying outside $\mathcal{S}_{ID}$, it offers no guarantee that it could not be reconstructed well.

Photo Restoration [3/11]

Networks



Microsoft Research, Bringing Old Photos Back to Life, https://www.youtube.com/watch?v=Q5bhszQq9eA

Photo Restoration [4/11]

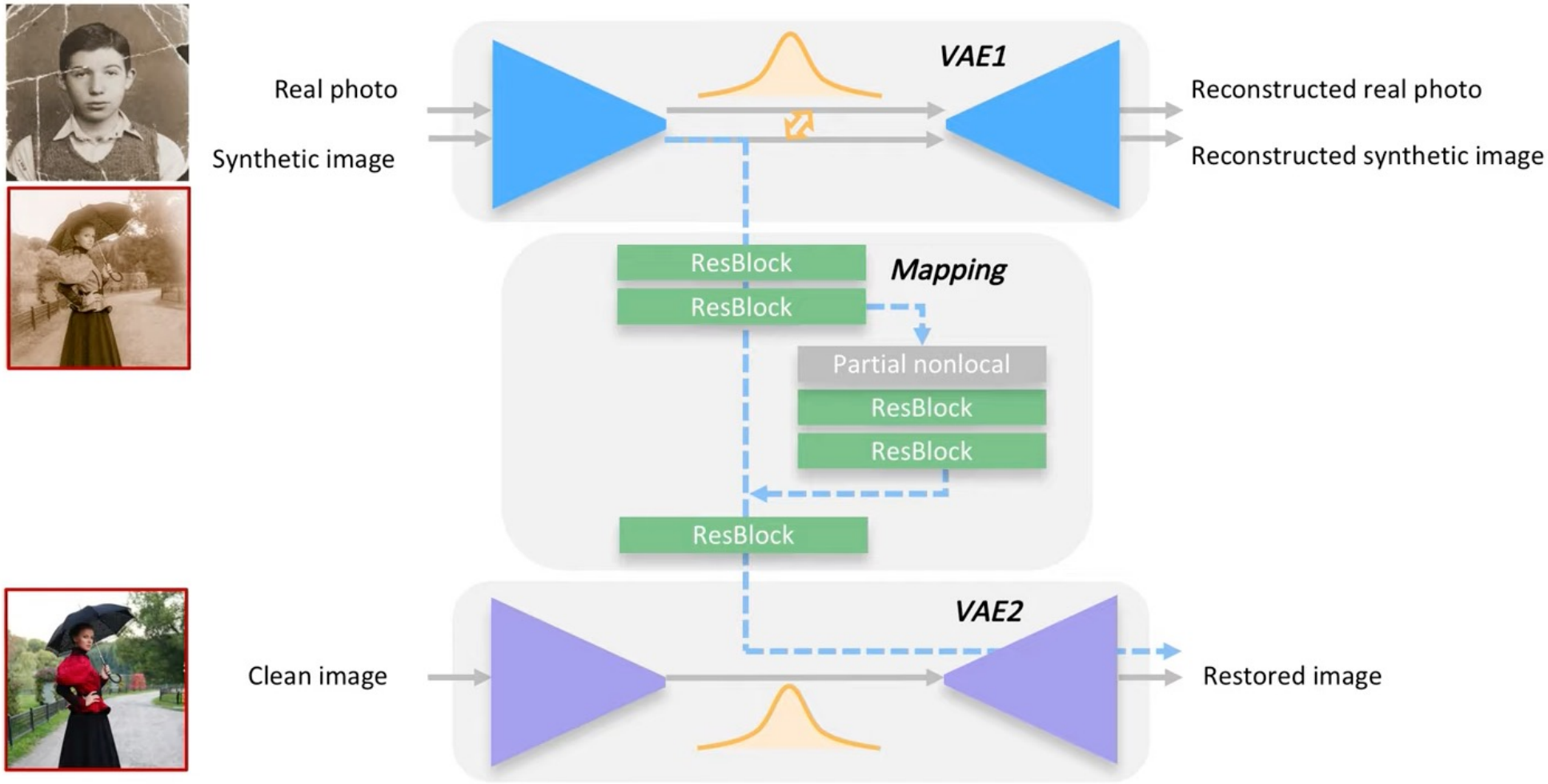
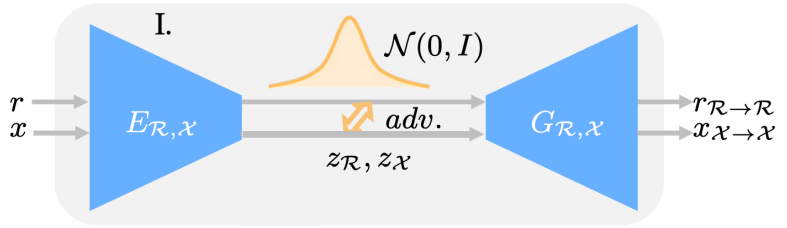


Photo Restoration [5/11]

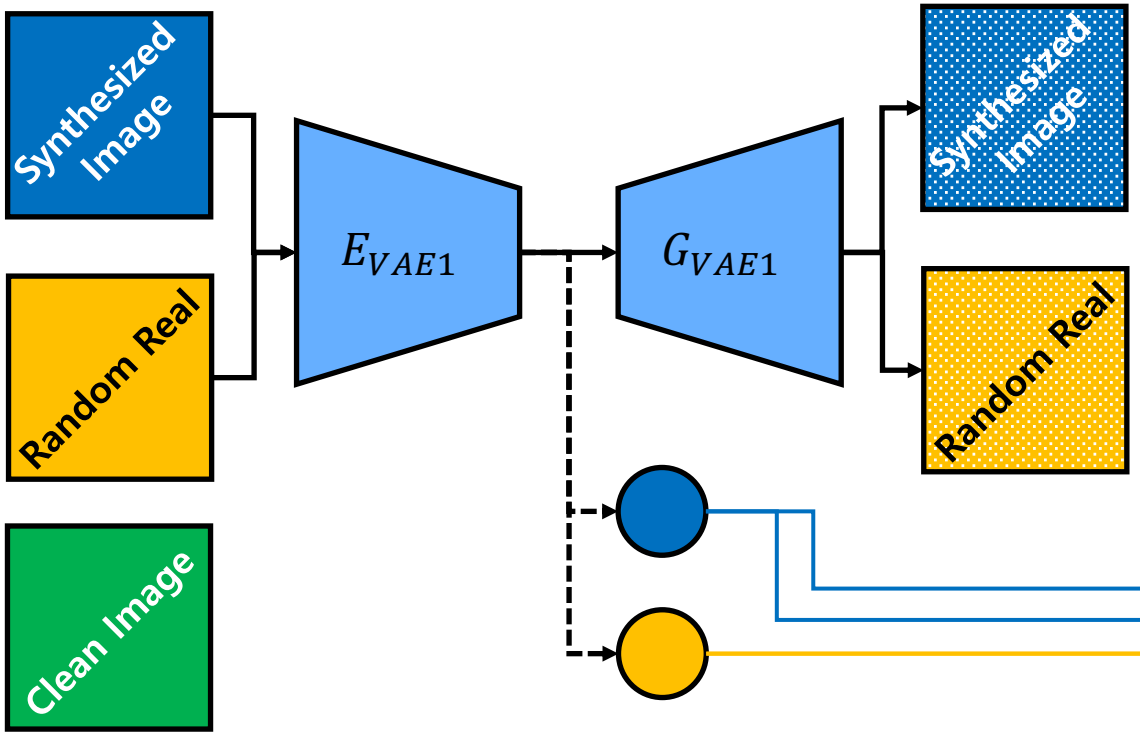
Training – VAE 1

* Mao, Xudong, et al. "Least squares generative adversarial networks." Proceedings of the IEEE international conference on computer vision (CVPR). 2017.



$$\min_D V_{\text{LSGAN}}(D) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x})-1)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))+1)^2\right] \quad (8)$$

$$\min_G V_{\text{LSGAN}}(G) = \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})))^2\right].$$



$$\mathcal{L}_{\text{VAE}_1}(r) = 2\text{KL}(E_{\mathcal{R},\mathcal{X}}(z_r|r)||\mathcal{N}(0,I)) + \alpha\mathbb{E}_{z_r\sim E_{\mathcal{R},\mathcal{X}}(z_r|r)}\left[\|G_{\mathcal{R},\mathcal{X}}(r_{\mathcal{R}\to\mathcal{R}}|z_r) - r\|_1\right] + \mathcal{L}_{\text{VAE}_1,\text{GAN}}(r) \quad (2)$$

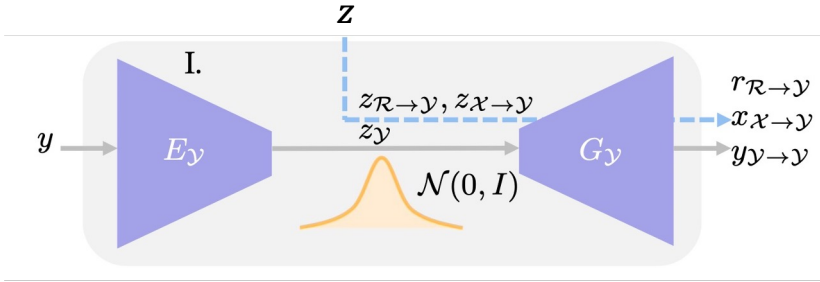
$$\mathcal{L}^{\text{latent}}_{\text{VAE}_1,\text{GAN}}(r,x) = \mathbb{E}_{x\sim\mathcal{X}}[D_{\mathcal{R},\mathcal{X}}(E_{\mathcal{R},\mathcal{X}}(x))^2] + \mathbb{E}_{r\sim\mathcal{R}}[(1-D_{\mathcal{R},\mathcal{X}}(E_{\mathcal{R},\mathcal{X}}(r)))^2]. \quad (3)$$

$$\min_{E_{\mathcal{R},\mathcal{X}},G_{\mathcal{R},\mathcal{X}}} \max_{D_{\mathcal{R},\mathcal{X}}} \mathcal{L}_{\text{VAE}_1}(r) + \mathcal{L}_{\text{VAE}_1}(x) + \mathcal{L}^{\text{latent}}_{\text{VAE}_1,\text{GAN}}(r,x). \quad (4)$$

Photo Restoration [6/11]

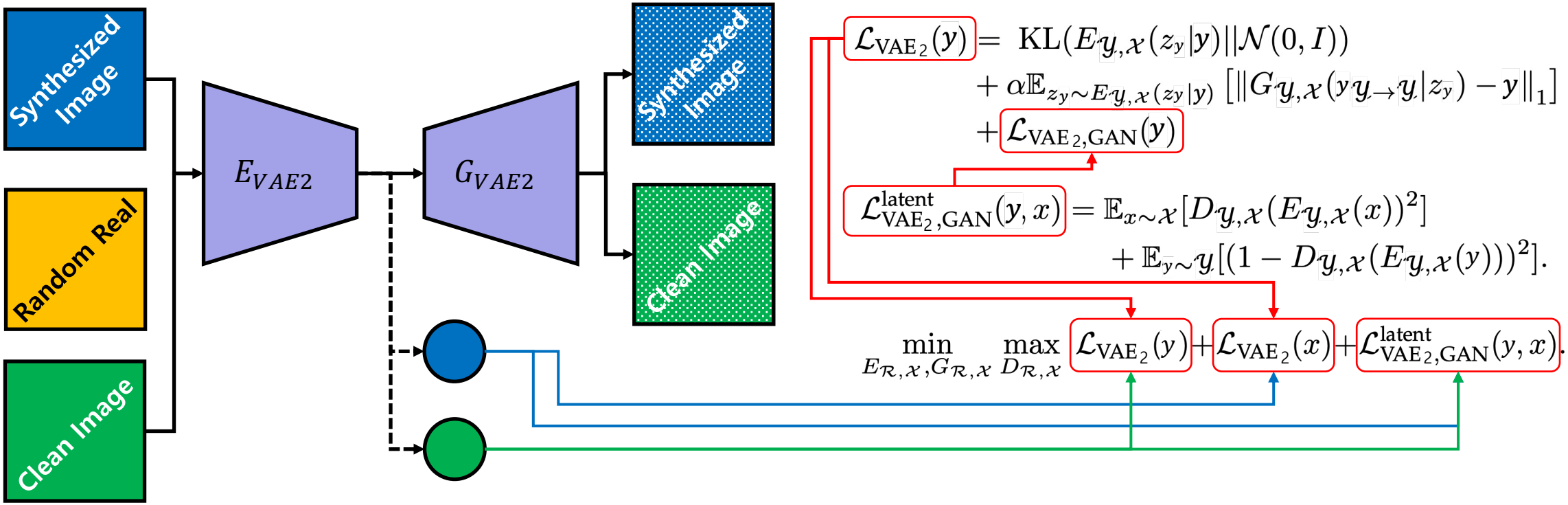
Training – VAE 2

* Mao, Xudong, et al. "Least squares generative adversarial networks." Proceedings of the IEEE international conference on computer vision (CVPR). 2017.



$$\min_D V_{\text{LSGAN}}(D) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\left[(D(\boldsymbol{x})-1)^2\right] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z}))+1)^2\right] \quad (8)$$

$$\min_G V_{\text{LSGAN}}(G) = \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\left[(D(G(\boldsymbol{z})))^2\right].$$



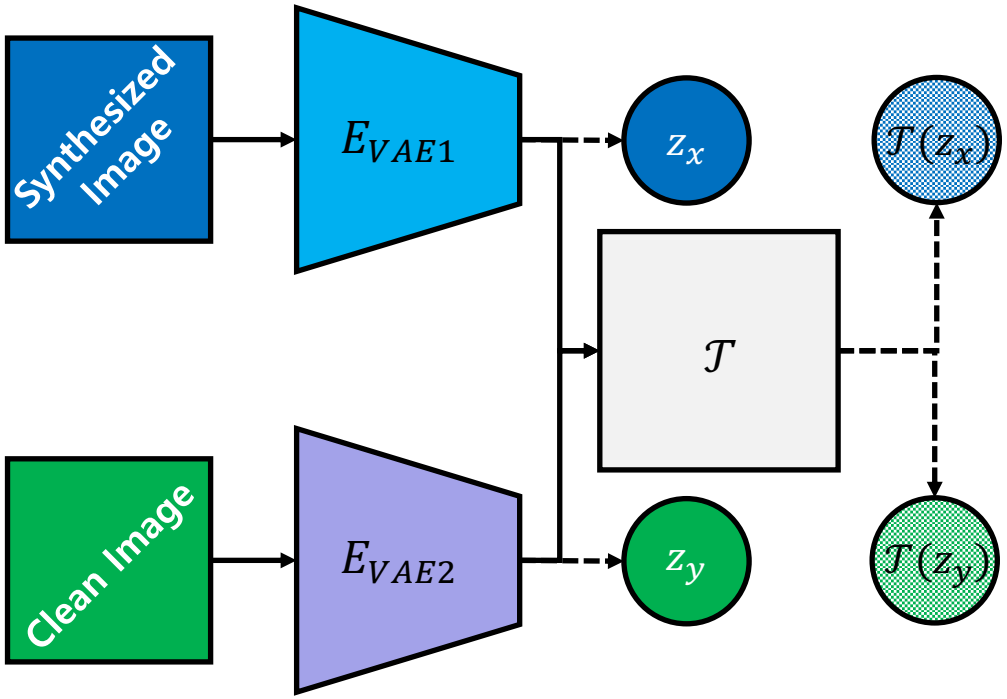
$$\mathcal{L}_{\text{VAE}_2}(y) = \text{KL}(E_{y,x}(z_y|y)||\mathcal{N}(0,I)) + \alpha\mathbb{E}_{z_y\sim E_{y,x}(z_y|y)}\left[\|G_{y,x}(y_{y\to y}|z_y) - y\|_1\right] + \mathcal{L}_{\text{VAE}_2,\text{GAN}}(y)$$

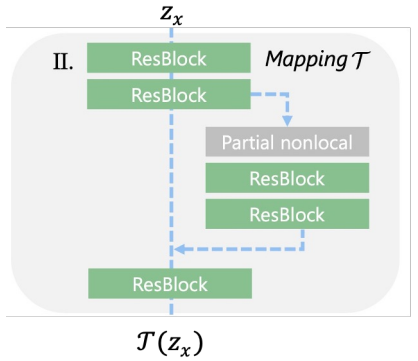
$$\mathcal{L}^{\text{latent}}_{\text{VAE}_2,\text{GAN}}(y,x) = \mathbb{E}_{x\sim\mathcal{X}}[D_{y,x}(E_{y,x}(x))^2] + \mathbb{E}_{y\sim\mathcal{Y}}[(1-D_{y,x}(E_{y,x}(y)))^2].$$

$$\min_{E_{\mathcal{R},\mathcal{X}},G_{\mathcal{R},\mathcal{X}}} \max_{D_{\mathcal{R},\mathcal{X}}} \mathcal{L}_{\text{VAE}_2}(y) + \mathcal{L}_{\text{VAE}_2}(x) + \mathcal{L}^{\text{latent}}_{\text{VAE}_2,\text{GAN}}(y,x).$$

Photo Restoration [7/11]

Training – Mapping Network





$$\mathcal{L}_{\mathcal{T}}(x,y) = \overset{60}{\lambda_1}\mathcal{L}_{\mathcal{T},\ell_1} + \mathcal{L}_{\mathcal{T},\text{GAN}} + \overset{10}{\lambda_2}\mathcal{L}_{\text{FM}} \tag{5}$$

$$\mathcal{L}_{\mathcal{T},\ell_1} = \mathbb{E}\,\|\mathcal{T}(z_x) - z_y)\|_1$$

$$\mathcal{L}^{\text{latent}}_{\text{VAE}_1,\text{GAN}}(r,x) = \mathbb{E}_{x\sim\mathcal{X}}[D_{\mathcal{T}}(\mathcal{T}(z_x))^2] + \mathbb{E}_{r\sim\mathcal{R}}[(1 - D(z_y)^2]$$

LPIPS with L1

$$\mathcal{L}_{\text{FM}} = \mathbb{E}\left[\sum_i \frac{1}{n^i_{D_{\mathcal{T}}}}\|\phi^i_{D_{\mathcal{T}}}(x_{\mathcal{X}\to\mathcal{Y}}) - \phi^i_{D_{\mathcal{T}}}(y_{\mathcal{Y}\to\mathcal{Y}})\|_1 + \sum_i \frac{1}{n^i_{\text{VGG}}}\|\phi^i_{\text{VGG}}(x_{\mathcal{X}\to\mathcal{Y}}) - \phi^i_{\text{VGG}}(y_{\mathcal{Y}\to\mathcal{Y}})\|_1\right], \tag{6}$$

$\phi^i_{D_{\mathcal{T}}}$ (ϕ^i_{VGG}) : feature map of the i-th layer

$n^i_{D_{\mathcal{T}}}$ (n^i_{VGG}) : number of activation in i-th layer

Photo Restoration [8/11]

Non-local Block (Inpainting with global and local features)

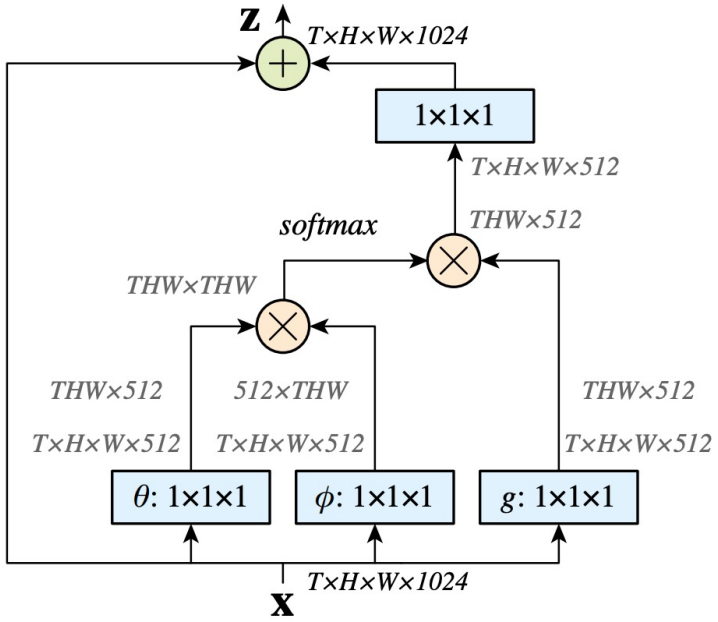


Figure 2. A spacetime **non-local block**. The feature maps are shown as the shape of their tensors, *e.g.*, $T\times H\times W\times 1024$ for 1024 channels (proper reshaping is performed when noted). "$\otimes$" denotes matrix multiplication, and "$\oplus$" denotes element-wise sum. The softmax operation is performed on each row. The blue boxes denote $1\times1\times1$ convolutions. Here we show the embedded Gaussian version, with a bottleneck of 512 channels. The vanilla Gaussian version can be done by removing θ and ϕ, and the dot-product version can be done by replacing softmax with scaling by $1/N$.

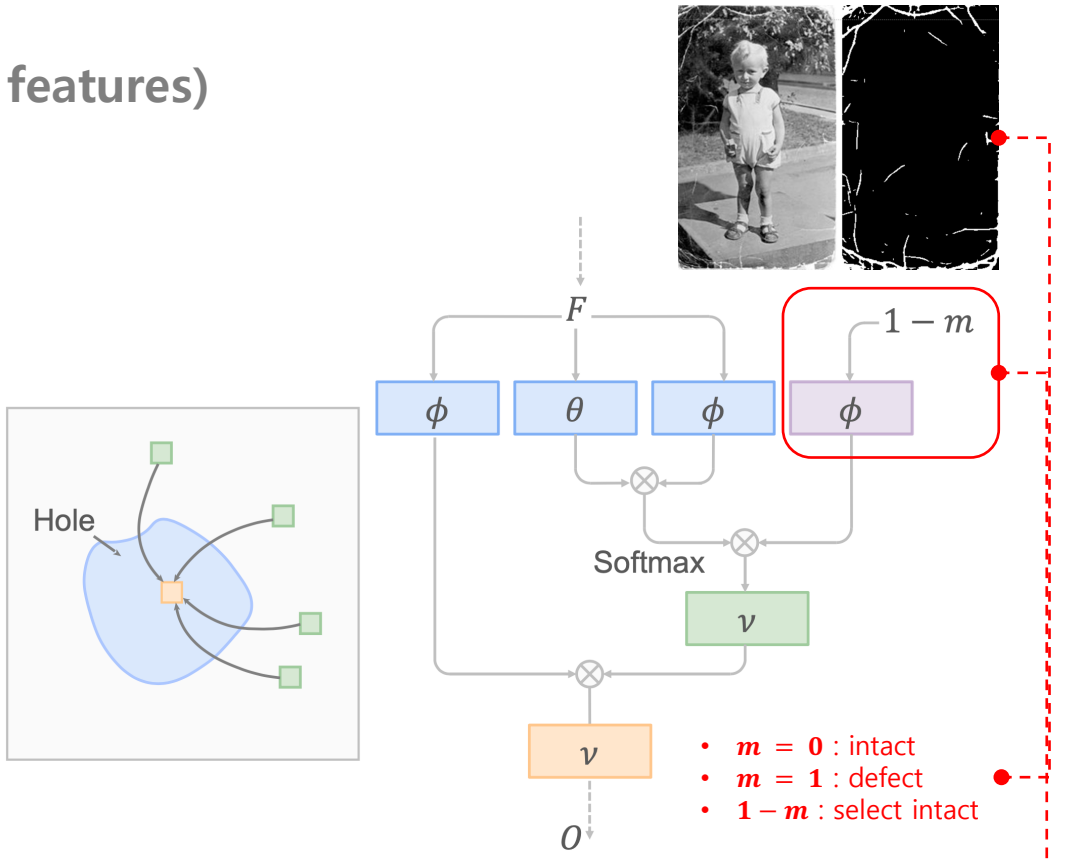


Fig. 4: **Partial nonlocal block**. Left shows the principle. The pixels within the hole areas are inpainted by the context pixels outside the corrupted region. Right shows the detailed implementation.

Photo Restoration [9/11]

Non-local Block

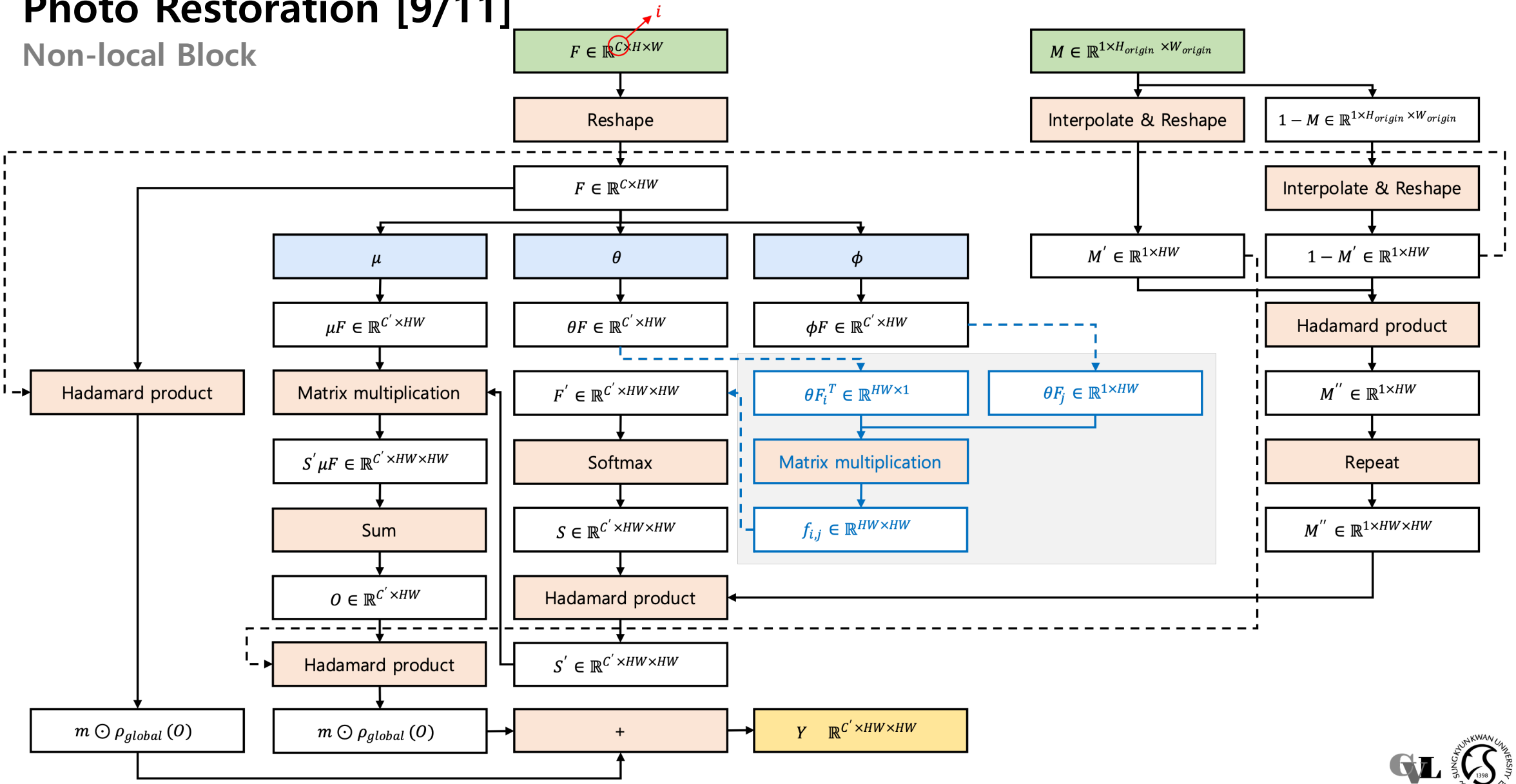
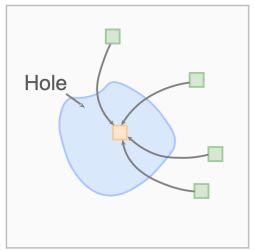
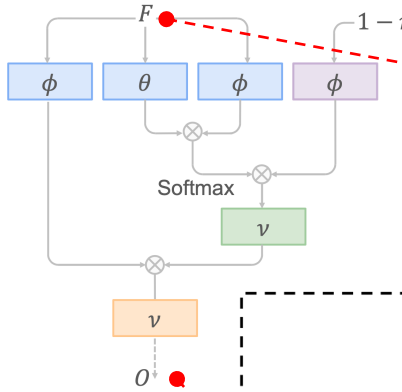


Photo Restoration [10/11]

Non-local Block





$$s_{i,j} = (1 - m_j) f_{i,j} / \sum_{\forall k} (1 - m_k) f_{i,k}, \quad (7)$$

$$f_{i,j} = \exp(\theta(F_i)^T \cdot \phi(F_j)) \quad (8)$$

$$O_i = \nu \left(\sum_{\forall j} s_{i,j} \mu(F_j) \right), \quad (9)$$

$$F_{fuse} = (1 - m) \odot \rho_{\text{local}}(F) + m \odot \rho_{\text{global}}(O), \quad (10)$$

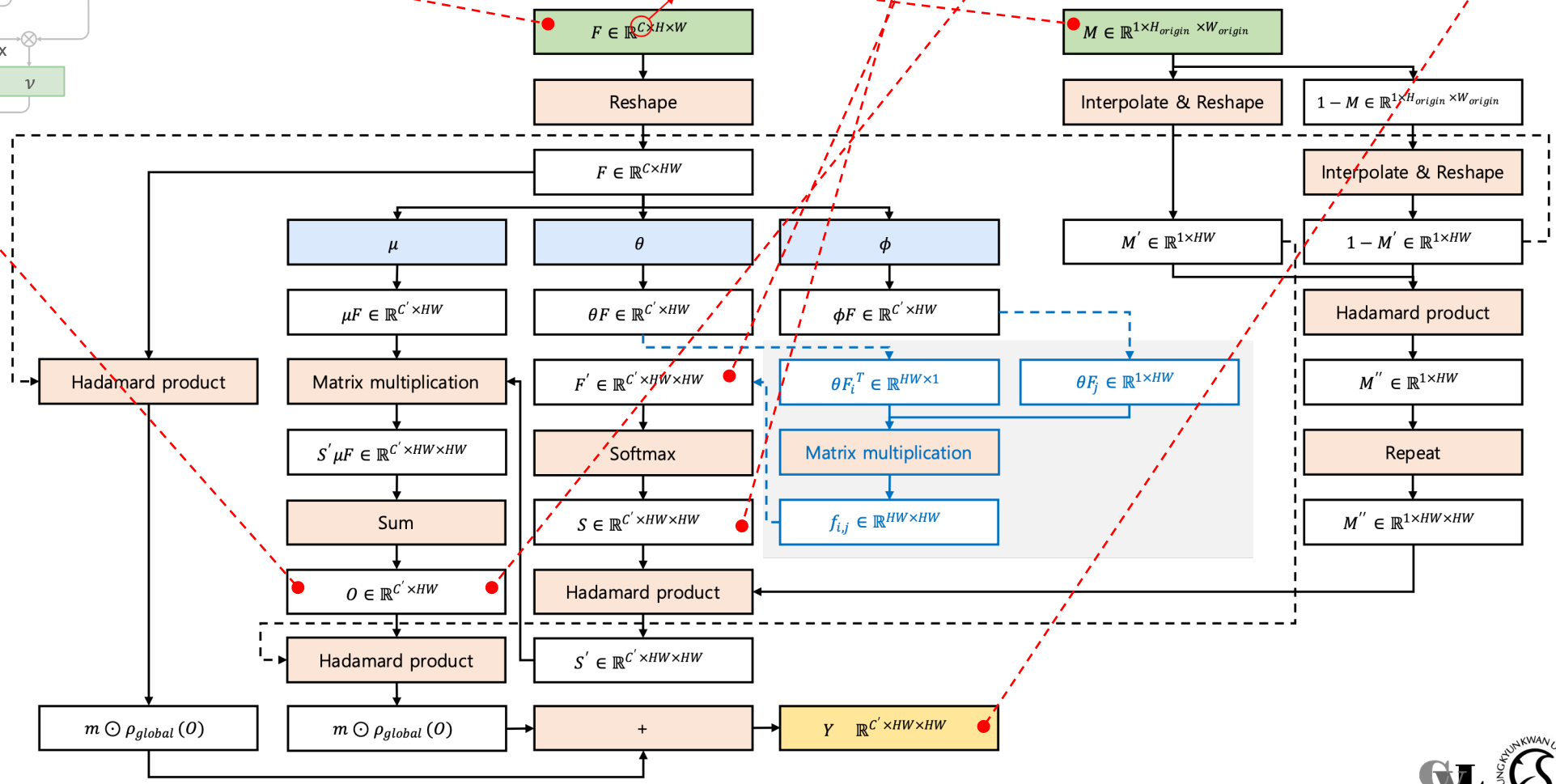


Photo Restoration [11/11]

High-Resolution Processing

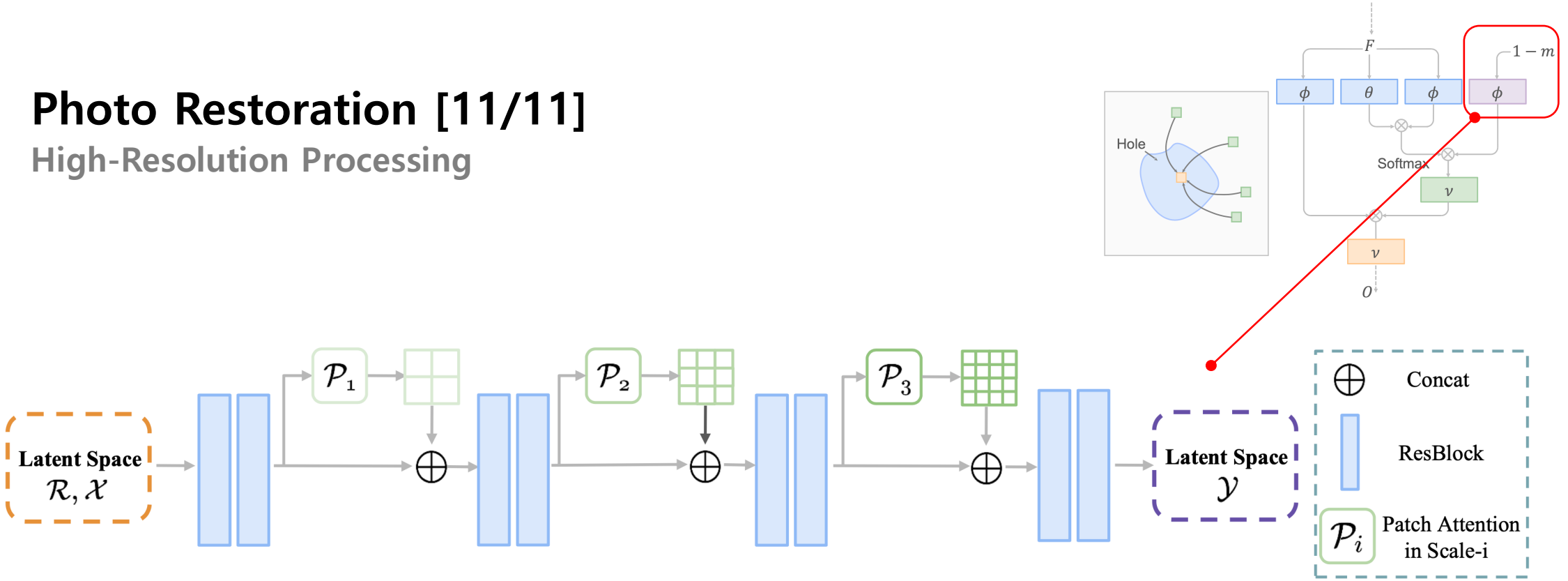


Fig. 5: **Multi-scale patch-based fusion.** The new proposed latent space mapping network vastly reduces the memory cost of partial nonlocal block and enable the processing of high-resolution photos.

Loffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning (ICML), 2015.
Park, Taesung, et al. "Semantic image synthesis with spatially-adaptive normalization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2019.

Face Enhancement

To restore the blind spot

$$x \in H \quad y \in W \quad c \in C$$

$$\underbrace{\gamma_{x,y,c}(r_f^i)}_{\text{Learnable parameter}} \frac{h_{x,y,c} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \underline{\beta_{x,y,c}(r_f^i)}, \quad (12)$$

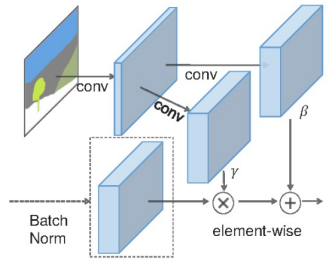


Figure 2: In SPADE, the mask is first projected onto an embedding space, and then convolved to produce the modulation parameters γ and β. Unlike prior conditional normalization methods, γ and β are not vectors, but tensors with spatial dimensions. The produced γ and β are multiplied and added to the normalized activation element-wise.

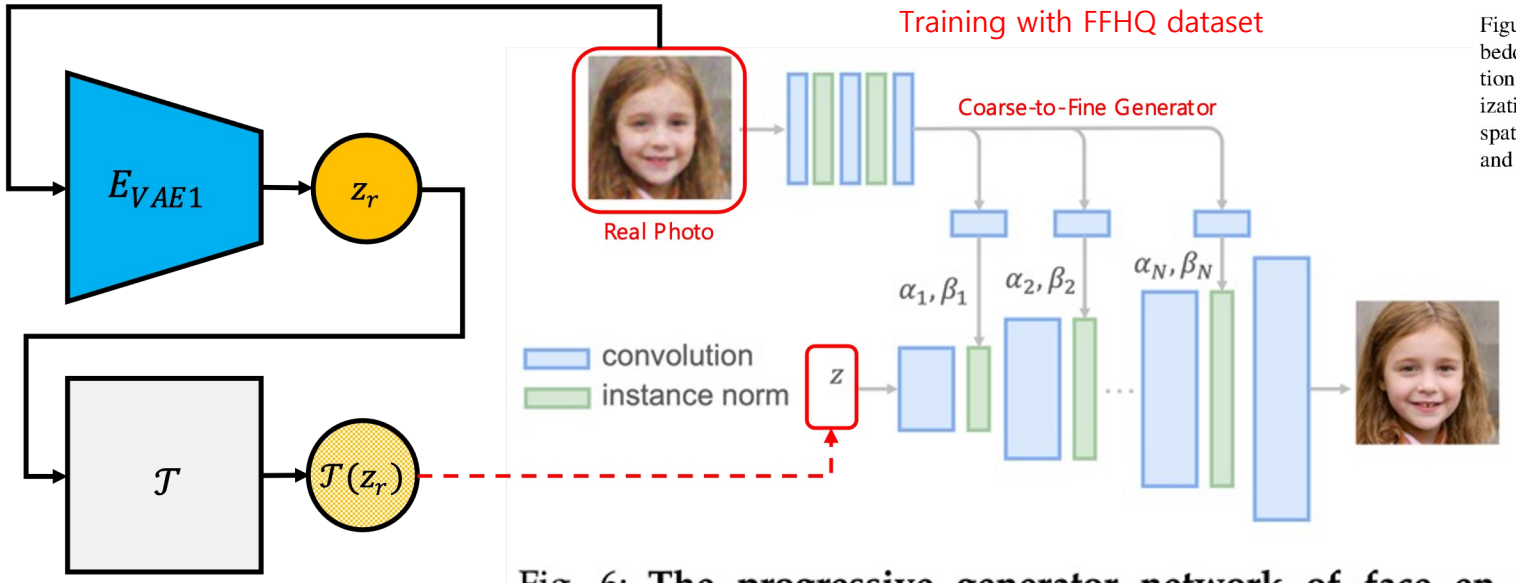


Fig. 6: **The progressive generator network of face enhancement.** Starting from a latent vector z, the network up-samples the feature map by deconvolution progressively. The degraded face will be injected into different resolutions in a spatial condition manner.

$$\mathcal{L}_{\text{perc}}^{\text{face}} = \mathbb{E}\left[\sum_i \frac{1}{n_{\text{VGG}}^i} \|\phi_{\text{VGG}}^i(G_f(z, r_f)) - \phi_{\text{VGG}}^i(r_c)\|_1\right], \quad (13)$$

r_f: degraded face
r_c: ground truth

$$\mathcal{L}_{\text{GAN}}^{\text{face}}(z, r_f, r_c) = \mathbb{E}_{z \sim \mathcal{Z}, r_f \sim \mathcal{R}_f}[D_f(G_f(z, r_f))^2] + \mathbb{E}_{r_c \sim \mathcal{R}_c}[(1 - D_f(r_c))^2]. \quad (14)$$

Experiments

Everingham, Mark, et al. "The pascal visual object classes challenge: A retrospective." International journal of computer vision. 2015.
Simard, Patrice Y., and David W. Steinkraus. "Elastic distortions for automatic generation of labeled data." U.S. Patent No. 7,418,128. 26 Aug. 2008.

Data Preparation

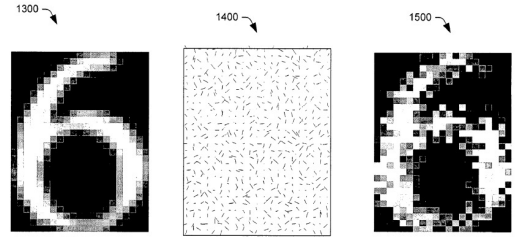
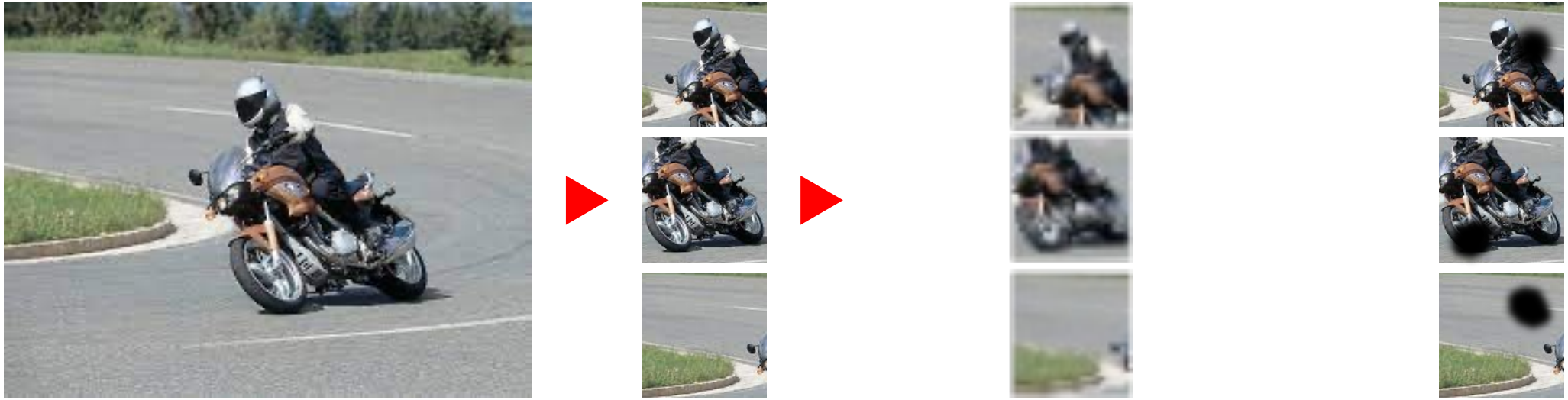


FIG. 13 FIG. 14 FIG. 15



Pascal VOC

Random Crop

Unstructured Degradation

- Gaussian white noise
- Gaussian blur
- JPEG compression
- Downsampling and upsampling

Structured Degradation (using Elastic Distortions)

- 62 scratchtexture images
- 55 paper texture images
- Random holes
- Film grain noise

+ 783 manual annotated samples

Defect Region Detection

For mask generation

Fig. 9: **Defect region detection results on real photos.**

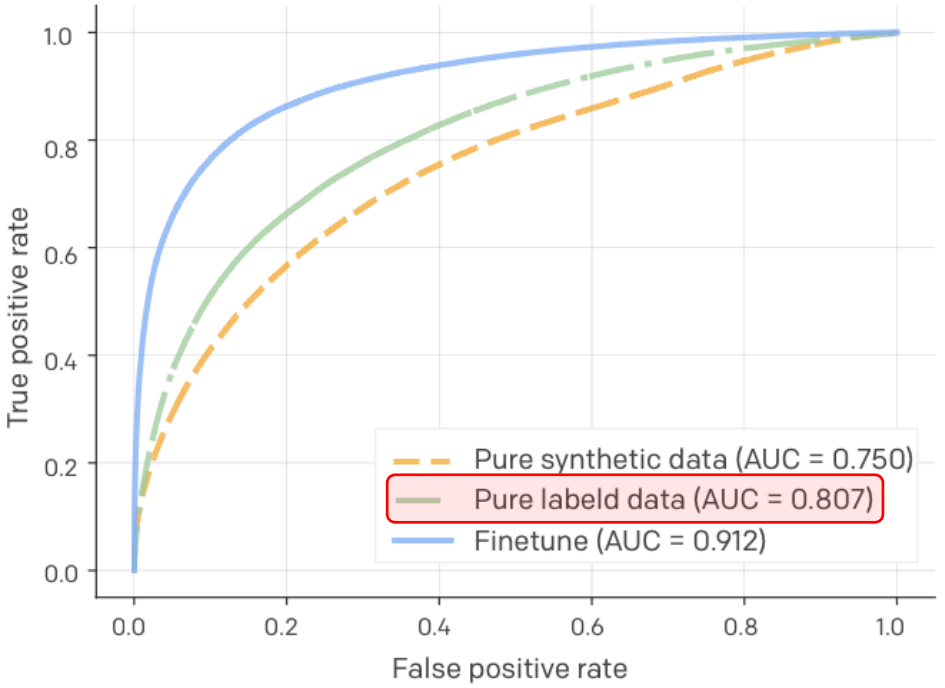


Fig. 8: **ROC curve for scratch detection of different data settings.** Combining both synthetic structured degradations and a small amount of labeled data, the scratch detection network could achieve great results.

Pure synthetic: synthetic data from Pascal VOC
Pure labeled: manually annotated real degraded photo
Finetune: finetune with manually annotated data on pretrained model (w/ synthetic data)

Wan, Ziyu, et al. "Bringing old photos back to life." proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). 2020.
Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." Advances in neural information processing systems (NeurIPS). 2017.

Quantitative Comparison

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Input	12.92	0.49	0.59	306.80
DIP [53]	22.59	0.57	0.54	194.55
Sequential [27], [70]	22.71	0.60	0.49	191.98
Attention [52]	**24.12**	**0.70**	0.33	208.11
Pix2pix [71]	22.18	0.62	**0.23**	135.14
Ours w/o partial nonlocal	23.14	0.68	0.26	143.62
Ours* [72] (CVPR 2020)	23.33	0.69	0.25	**134.35**
Ours	**23.45**	**0.71**	**0.24**	**130.64**

TABLE 2: **Quantitative results on the DIV2K dataset.** ↑ indicates that a higher score denotes a good image quality. We highlight the best two scores for each measure.

- LPIPS: feature distance
- FID: feature distribution distance

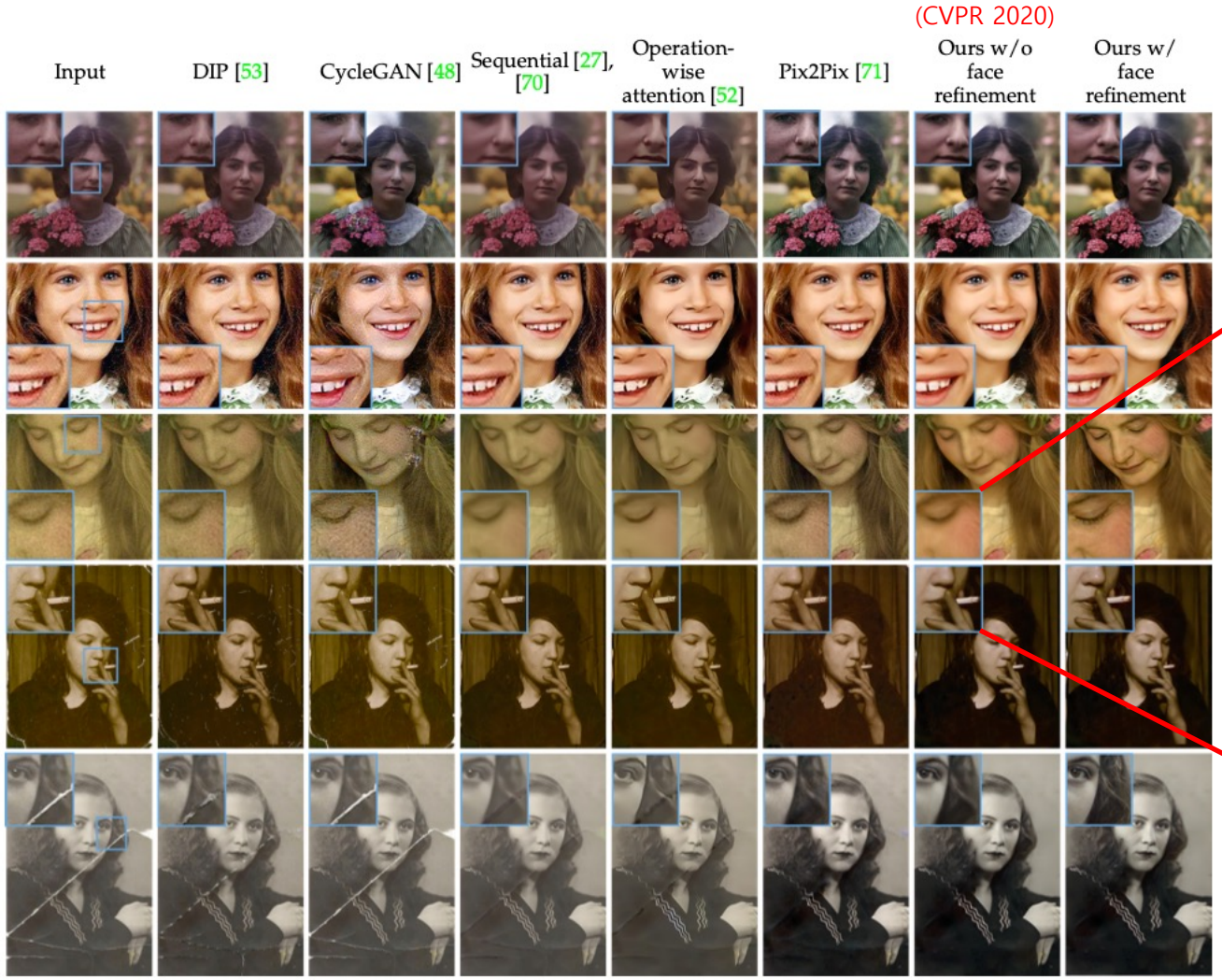
$$FID = \|\mu_x - \mu_y\|^2 - \|\sigma_x - \sigma_y\|^2$$

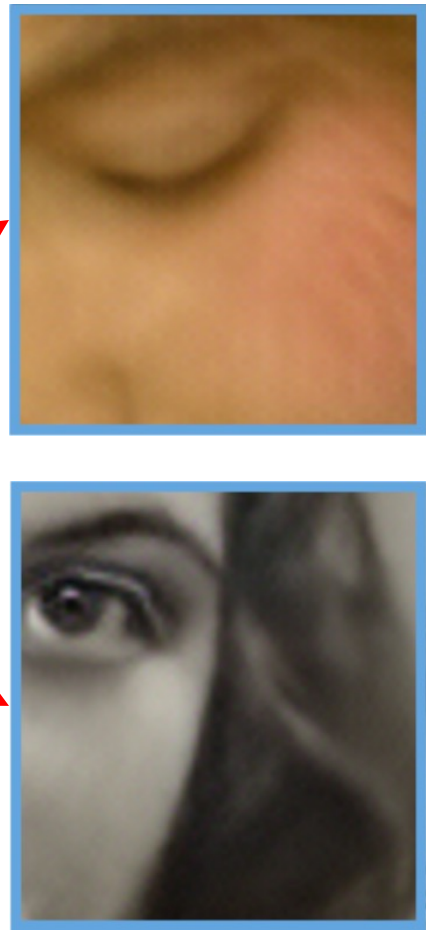
Method	Top 1	Top 2	Top 3	Top 4	Top 5
DIP [53]	2.54	8.49	19.26	39.09	74.22
CycleGAN [48]	4.24	8.21	19.54	28.32	50.42
Sequential [27], [70]	4.81	18.13	47.87	79.60	94.61
Attention [52]	6.79	21.24	49.85	73.08	88.38
Pix2Pix [71]	16.14	60.90	73.65	86.68	94.90
Ours	**65.43**	**83.00**	**89.80**	**93.20**	**97.45**

TABLE 3: **User study results.** The percentage (%) of each method is selected as the top K ($K = 1 - 5$) by users.

Qualitative Comparison [1/2]

(CVPR 2020)





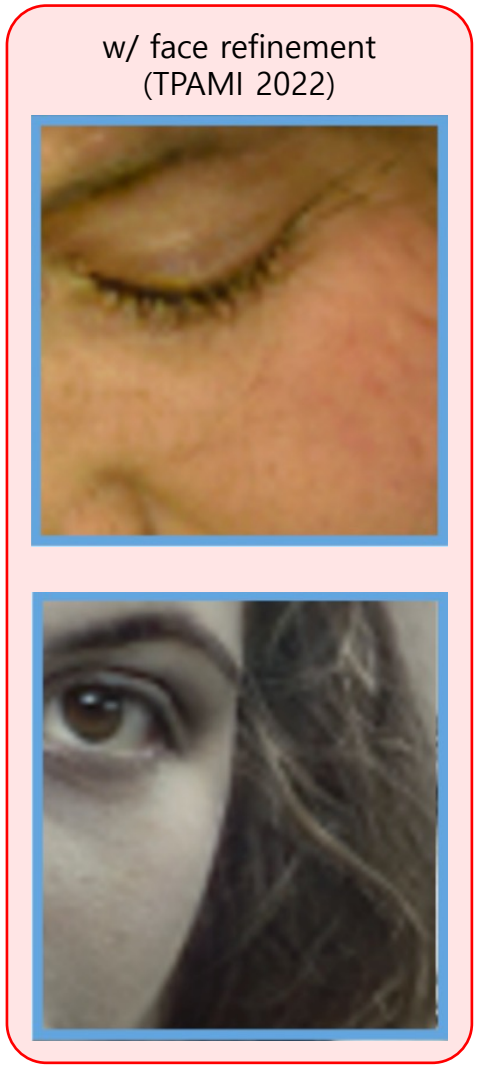


Fig. 10: **Qualitative comparison against state-of-the-art methods.** It shows that our method can restore both unstructured and structured degradation and our recovered results are significantly better than other methods.

Qualitative Comparison [2/2]

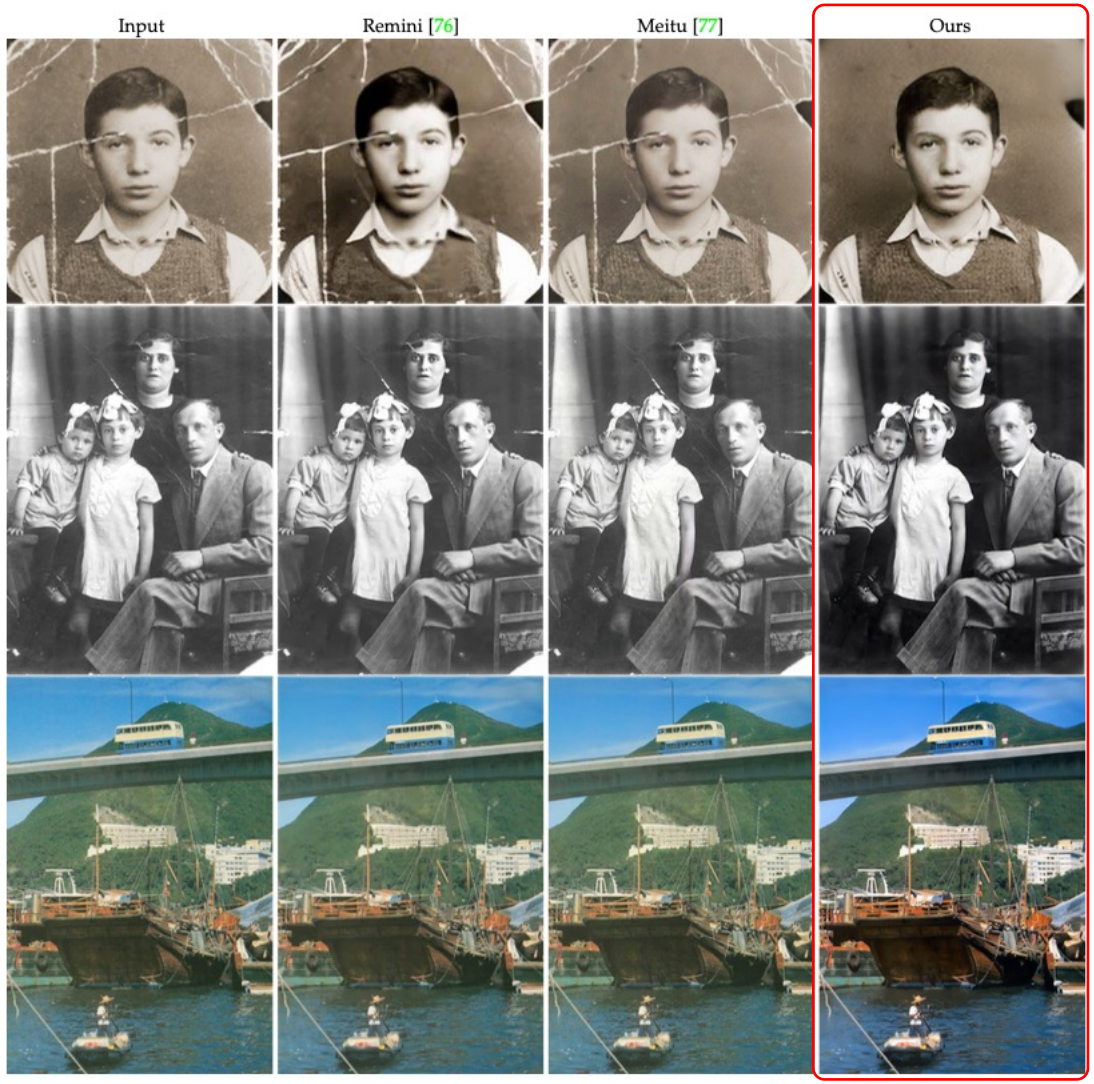


Fig. 11: **Qualitative comparisons against commercial tools.** Remini Photo Enhancer [76], Meitu [77] and our full pipeline results are included. The last two rows are high-resolution restoration (736x1024 and 640x960).

Ablation Study

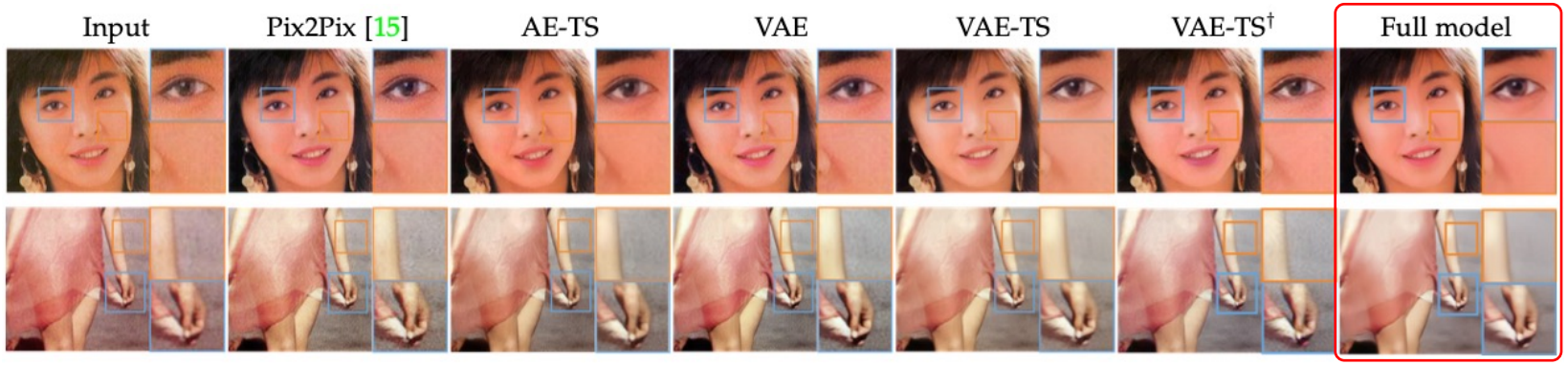


Fig. 12: **Ablation study for latent translation with VAEs.** By involving feature translation, VAE and feature-level adversarial loss, the domain gap between synthetic degradations and real-world defects could be narrowed better, leading to better restoration results step by step. VAE-TS†: Remove the real old data for two-stage VAE.

Annotation	Neural Network	Training
AE-TS	Two Auto-Encoder	Two-stage
VAE	Two Variational AE	Simultaneously
VAE-TS	Two VAE	Two-stage
VAE-TS†	Two VAE	Two-stage
Full model	Two VAE + Face refine	Two-stage

Fig. 13: **Ablation study of the partial nonlocal block.** Partial nonlocal better inpaints the structured defects.

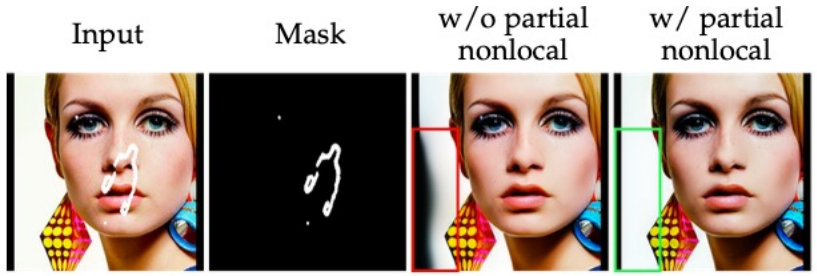


Fig. 14: **Ablation study of the partial nonlocal block.** Partial nonlocal does not touch the non-hole regions as this operation is aware of the corruption area.

Method	Pix2Pix	AEs-TS	VAEs	VAEs-TS	VAE-TS†	full model
Wasserstein ↓	1.837	1.432	1.048	0.765	1.027	**0.581**
BRISQUE ↓	25.549	24.547	23.949	23.396	23.850	**23.016**

TABLE 4: **Ablation study of latent translation with VAEs.** We provide some quantitative comparisons here to demonstrate the superior performance of the full model. Our full method achieves the best results on both distribution distance and BRISQUE metric.

Effect of Multi-Scale Patch-based Fusion

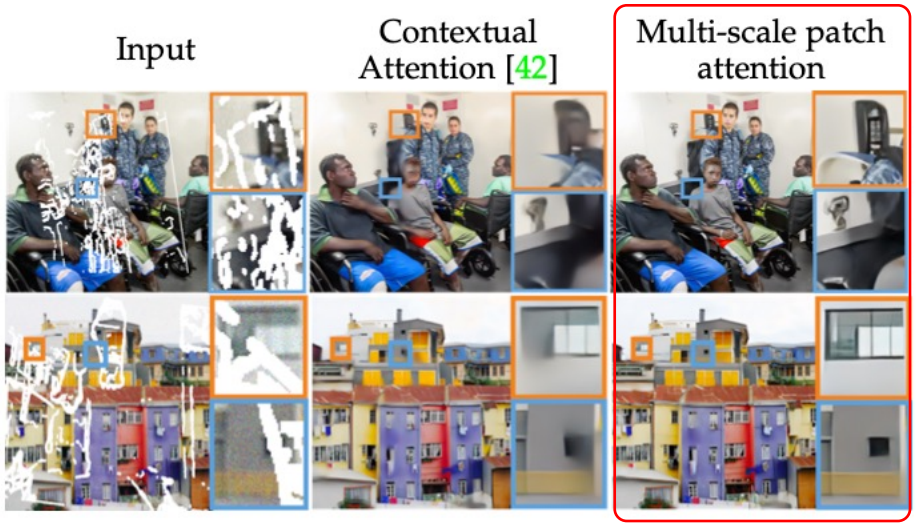


Fig. 15: **Qualitative comparisons between contextual attention and ours on high-resolution (1024x1024) photos.**

Method	PSNR ↑	SSIM ↑	LPIPS ↓	FID ↓
Remove Global	23.14	0.68	0.26	143.62
Nonlocal [64]	23.19	0.68	0.26	141.49
Contextual Attention [42]	23.25	0.69	0.25	138.27
Partial Nonlocal (CVPR 2020)	23.33	0.69	0.25	134.35
Multi-scale Patch Fusion	**23.45**	**0.71**	**0.24**	**130.64**

TABLE 5: **Quantitative comparisons on the DIV2K dataset (256x256). Bold**: Best. Underline: Second best.

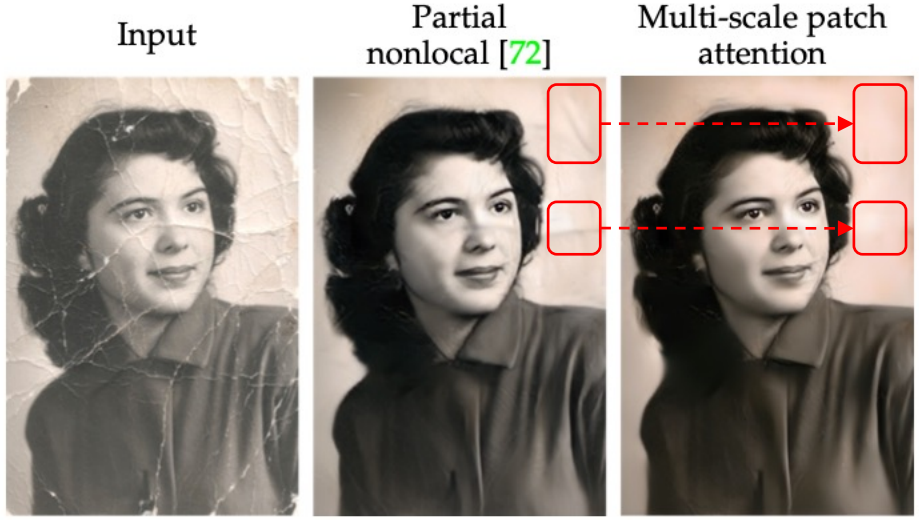


Fig. 16: **Qualitative comparisons between two designed global branches.** The multi-scale patch attention could better resolve the structured defects.

Summary

Purpose
- Unpaired old (degraded) photo restoration

Contributions
- Domain alignment: well generalization on real photos by training synthetic data
- Partial nonlocal block: reconstruction of blind spots / corruption area
- Multi-scale patch-based fusion: reduction of memory cost and better performance
- Coarse-to-fine generator: reconstruction of the high-resolution face

Limitations
- Anisotropic pollution is not totally restored.
- Shading along the folds of the photo hindrance the restoration process.