Bringing Old Photos Back to Life

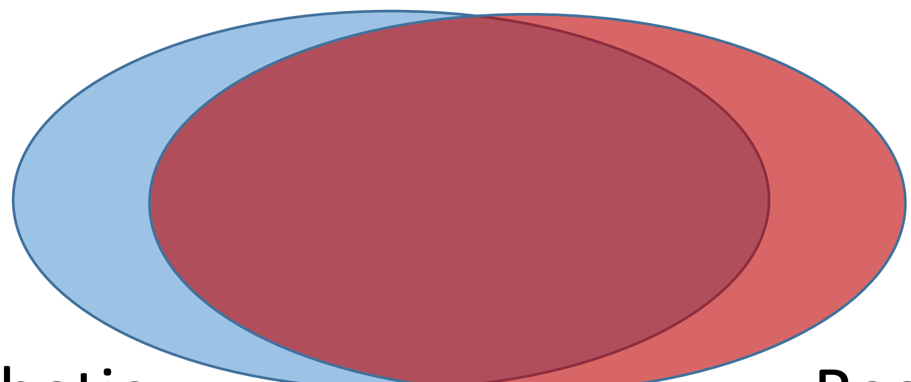
Ziyu Wan[1*], Bo Zhang[2], Dongdong Chen[3], Pan Zhang[4], Dong Chen[2], Jing Liao[1†], Fang Wen[2]
[1]City University of Hong Kong [2]Microsoft Research Asia [3]Microsoft Cloud + AI
[4]University of Science and Technology of China

CVPR 2020

Background

- Data types
  - Synthetic data (Paired):
    Original image + ground truth
    Necessity during training
  - Real data (Unpaired):
    Just original image
    For testing or daily using
- Hypothesis
  Two data types with similar or the same distribution

?



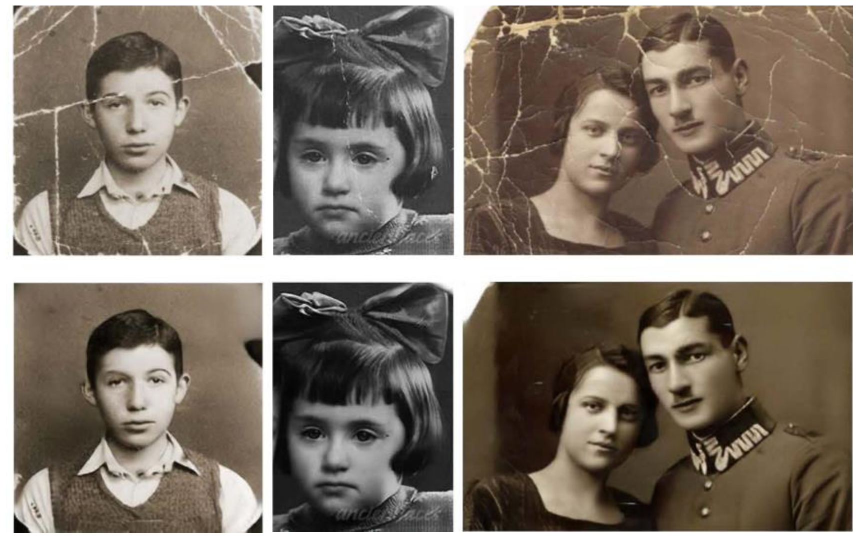
Synthetic

Real

Difference:
Font,
Language,
Illumination …

Introduction

- Artifacts in old photos

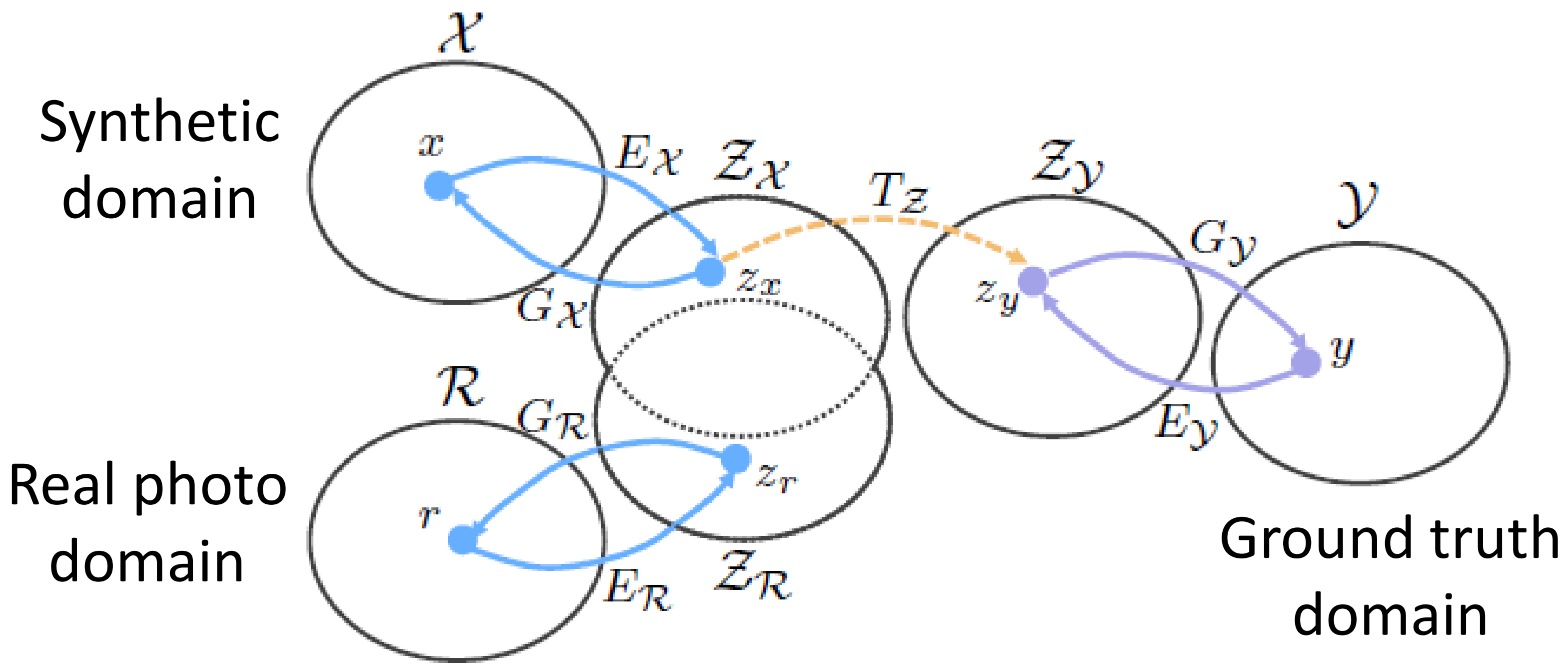


  - Unstructured defects: noise and blur
  - Image restoration
  - Structured defects: scratch and blemish
  - Image inpainting
- Challenge of supervised learning
  - The degradation of old photos is complex
  - Fail to generalize between synthetic and real photos

Motivation

- Triplet domain transform

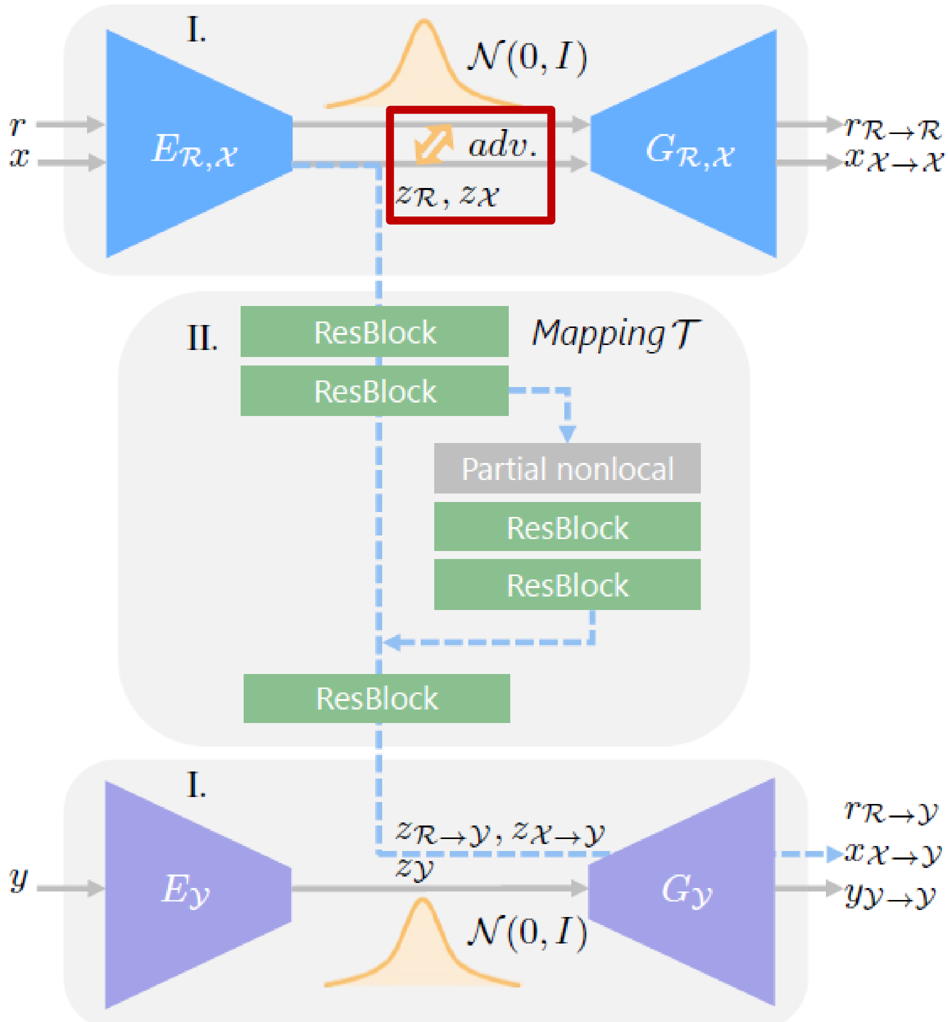
  Insight: domain gap is closed in latent space



Training: $\mathcal{X} \rightarrow \mathcal{Y} = G_{\mathcal{Y}} \circ T_{\mathcal{Z}} \circ E_{\mathcal{X}}(x)$

Testing: $\mathcal{R} \rightarrow \mathcal{Y} = G_{\mathcal{R}} \circ T_{\mathcal{Z}} \circ E_{\mathcal{X}}(r)$

Overview



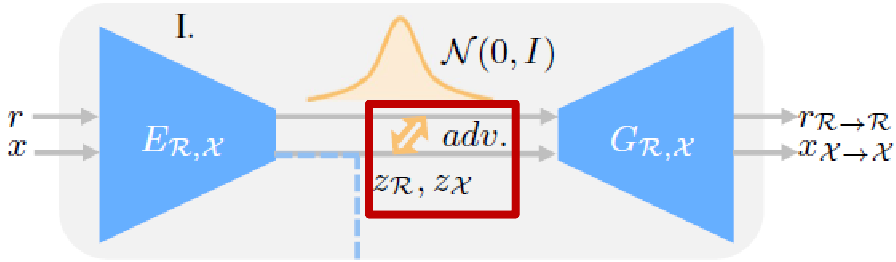
Representation learning of degraded images using VAE

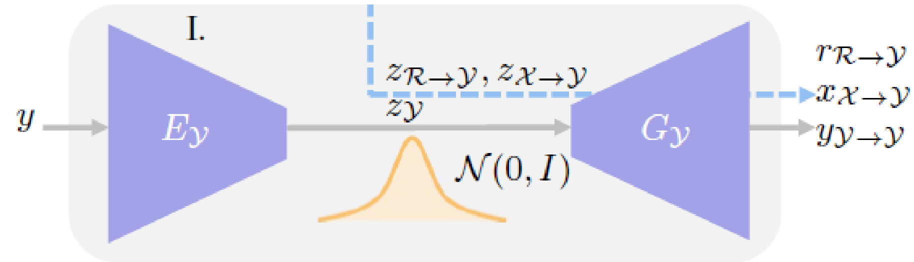
Alignment in latent space

Restoration of old photos through latent mapping

Representation learning of clean images using VAE

Representation learning





Training VAE (x, y with a similar loss)

$$\mathcal{L}_{\mathrm{VAE}_1}(r) = \mathrm{KL}(E_{\mathcal{R},\mathcal{X}}(z_r|r)||\mathcal{N}(0,I))$$ KL-divergence penalization

$$+\alpha \mathbb{E}_{z_r \sim E_{\mathcal{R},\mathcal{X}}(z_r|r)}\left[\|G_{\mathcal{R},\mathcal{X}}(r_{\mathcal{R}\to\mathcal{R}}|z_r) - r\|_1\right]$$ L1 loss

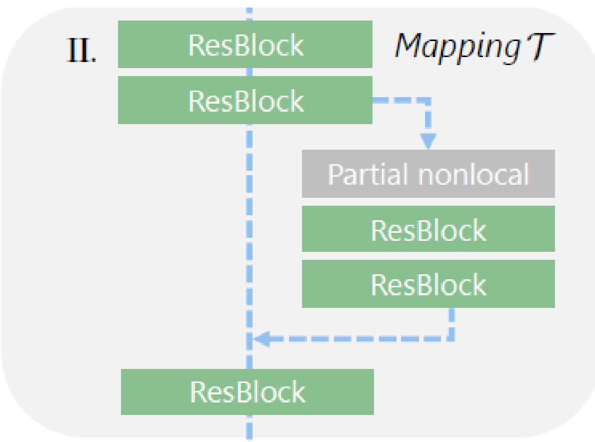
$$+\mathcal{L}_{\mathrm{VAE}_1,\mathrm{GAN}}(r)$$ Address over-smooth issue

Latent adversarial loss

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

Latent mapping

Mapping in low-dimensional latent space is much easier to learn than in the high-dimensional image space



Training mapping fixed VAE

$$\mathcal{L}_{\mathcal{T}}(x, y) = \lambda_1 \mathcal{L}_{\mathcal{T},\ell_1} + \mathcal{L}_{\mathcal{T},\mathrm{GAN}} + \lambda_2 \mathcal{L}_{\mathrm{FM}}$$

Latent space loss

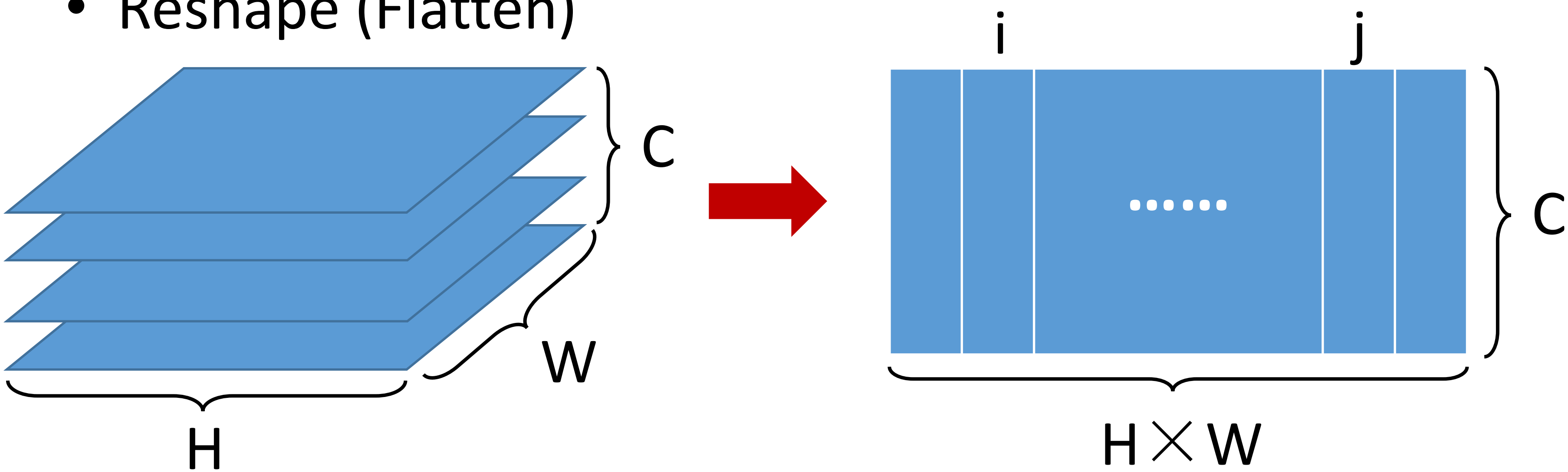
$$\mathcal{L}_{\mathcal{T},\ell_1} = \mathbb{E} \left\| \mathcal{T}(z_x) - z_y) \right\|_1$$

Feature matching loss

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

Partial nonlocal block

- Reshape (Flatten)



- Pairwise affinity with Gaussian

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

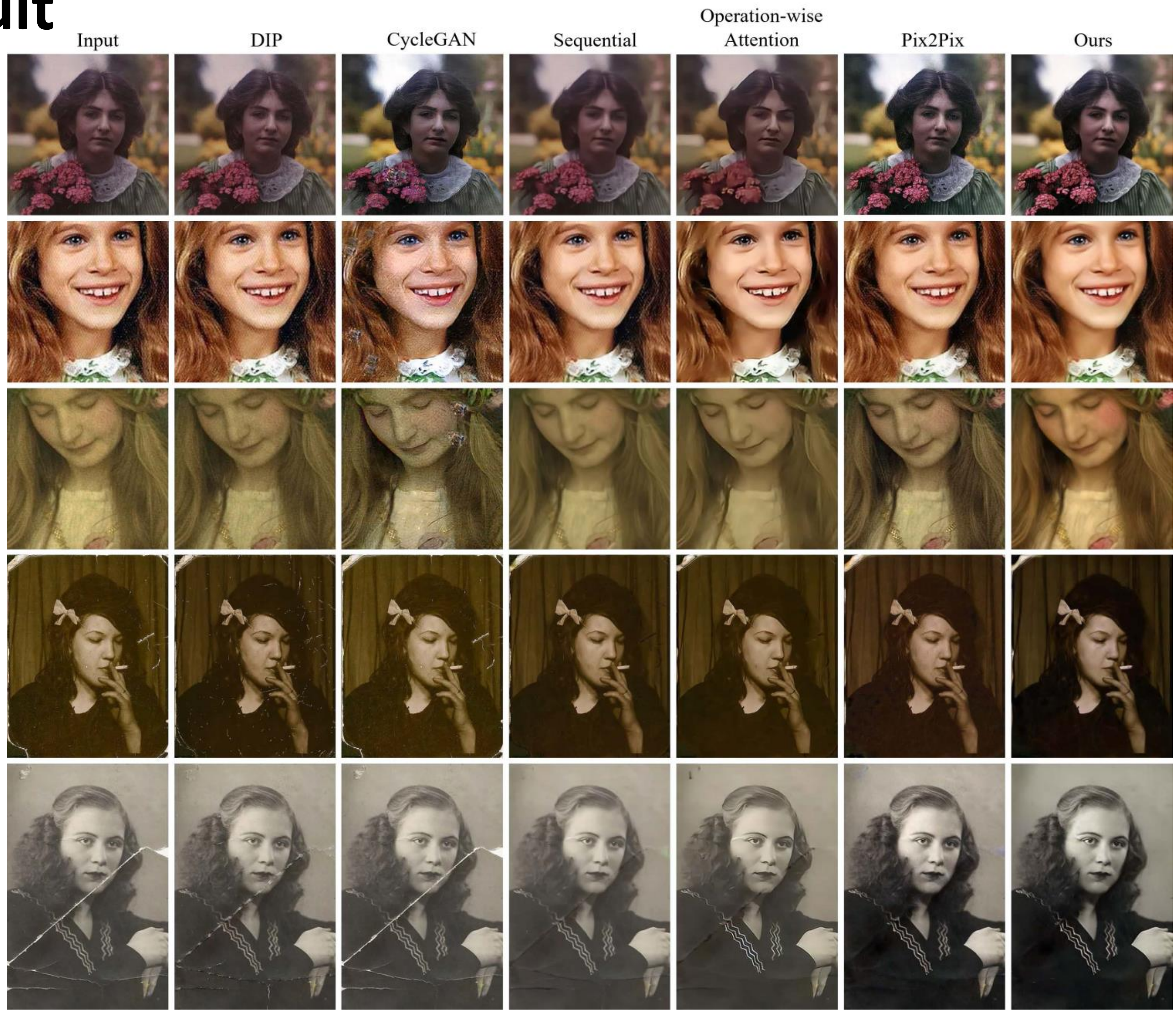
- Correlation between two channel vector i, j

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

- Outputs

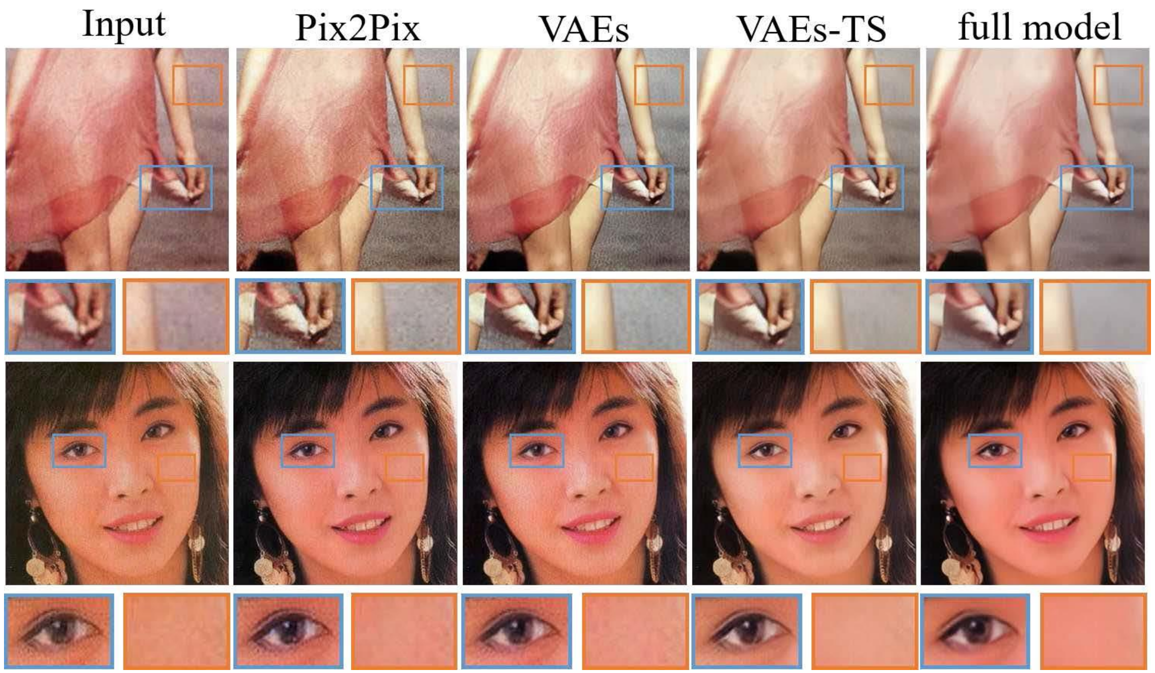
$$O_i = \nu \left(\sum_{\forall j} s_{i,j} \mu(F_j) \right)$$

Result

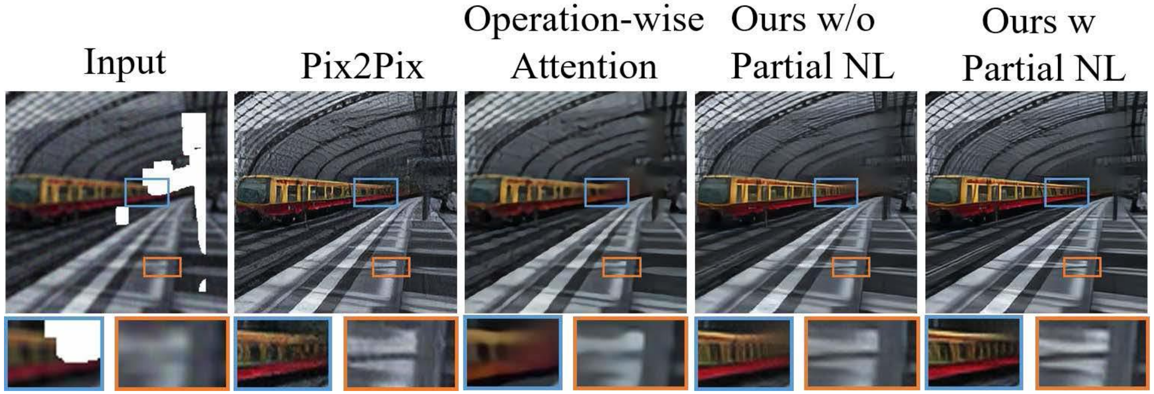

Input
DIP
CycleGAN
Sequential
Operation-wise Attention
Pix2Pix
Ours

Ablation study

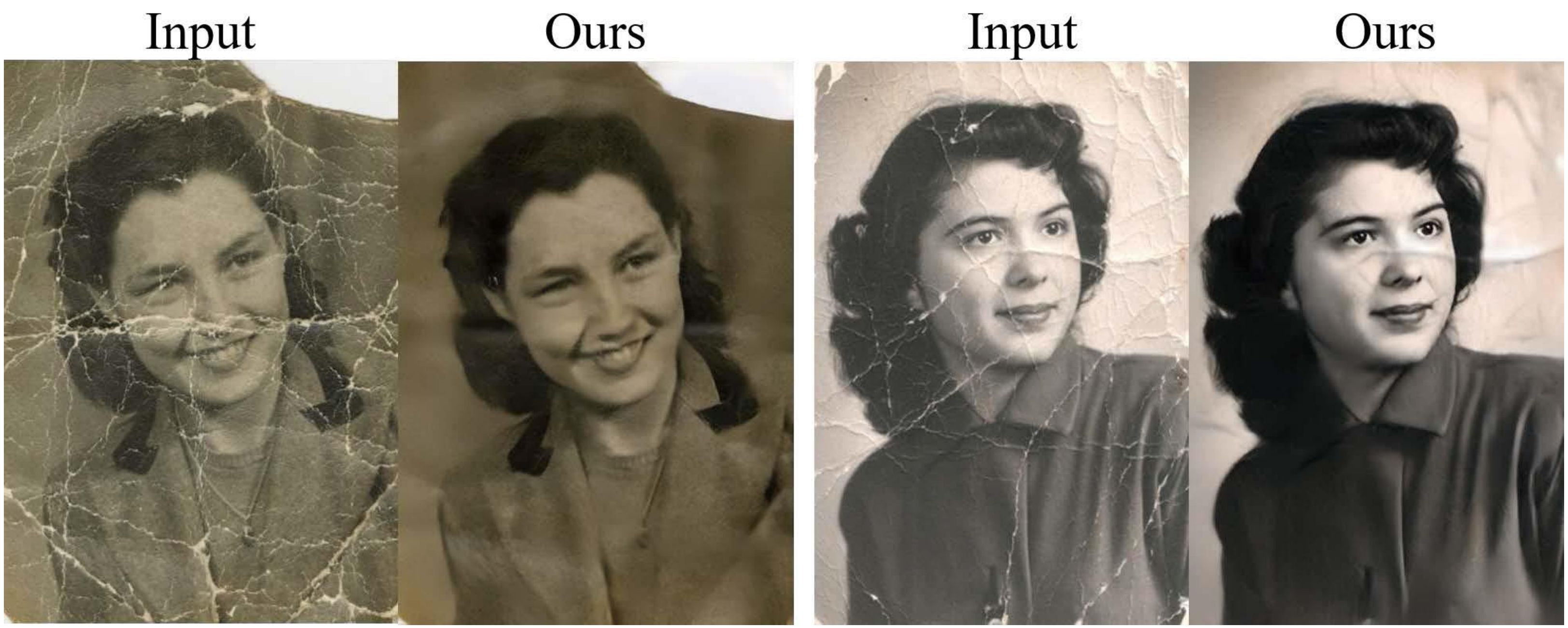
Latent translation with VAEs



Partial nonlocal block



Limitation



Fail to handle complex shading artifacts

Thank you!