CS 594 Deep Generative Models

Old Photo Restoration using Diffusion

Judhajit Roy

What is Old Photo Restoration?

- Old Photo Restoration is the task of restoring photos that suffer from degradation.
- Degradations like film grain, color fading and scratches.

Reasons for degradation of old photos

- Old photo prints deteriorate when kept in poor environmental condition, causing the photo content to be permanently damaged.
- Fortunately, we can now digitalize the photos and invite a skilled specialist for restoration.
- However, manual retouching is usually laborious and time consuming, which leaves piles of old photos impossible to get restored.

- Before Deep Learning the attempts to restore used techniques to automate localizing defects
- Local defects such as scratches are restored by filling damaged areas with inpainting techniques
- But these techniques can only focus on missing content
- Spatially-uniform defects such as film grain, sepia effect and color fading still remained unaddressed

- Deep learning techniques like CNNs helps us learn the mapping for specific task using large amount of synthetic images
- But this framework is not effective in old photo restoration which has a lot of components and is complex
- The model learned from these synthetic images generalizes poorly on real photos
- Also, the degradations (structured and unstructured) in these photos need to be addressed by different strategies:
    - Local pixels in neighborhood
    - Global image context

Bringing Old Photos Back to Life

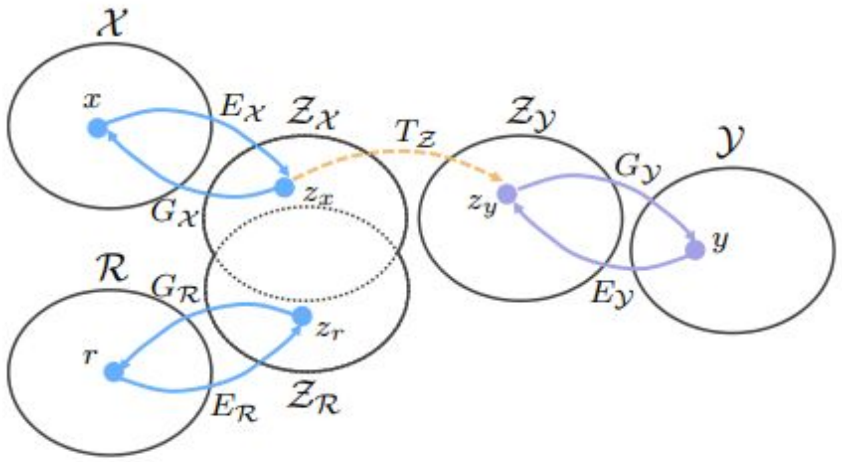
- Old photo restoration is done by formulating the problem as a triplet domain translation
  - Real old photos
  - Synthetic images
  - Corresponding ground truth
- Translation is performed in latent space
- Synthetic images and the real old photos are transformed into same latent space
- Ground truth is transformed into different latent space using another VAE

Foundation of the Problem

- Image restoration is the process of **recovering an image from a degraded version**
- It can be categorized into two levels:
  - Single degradation image restoration
  - Mixed or multi degradation image restoration
- Single degradation image restoration, is restoring any of the **structured or unstructured degradation**
- In the real world, we need Mixed degradation image restoration, that restores images with a mixture of these effects

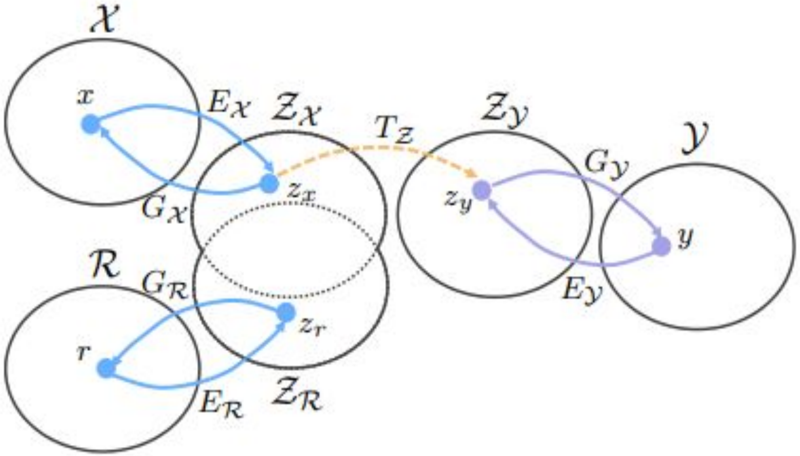
Translation method with three domains

- Three domains are r ∈ R, x ∈ X and y ∈ Y
- Where x and y are paired by data synthesizing i.e x is degraded from y
- r is the real old image
- $E_R : R \rightarrow Z_R$
- $E_X : X \rightarrow Z_X$
- $E_Y : Y \rightarrow Z_Y$
- And, $Z_R \approx Z_X$



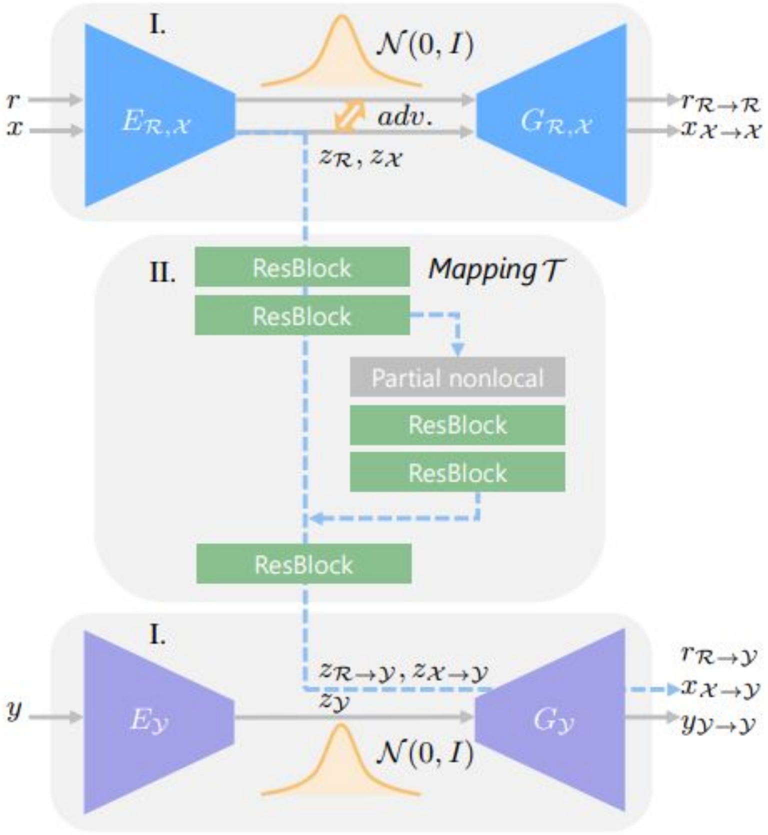
Translation method with three domains

- The image restoration is learned in latent space
- By, learning the translation from Z_X to Z_Y
- T_Z : Z_X to Z_Y
- Z_Y can be further reversed to Y
- Through generator $G_Y : Z_Y \rightarrow Y$
- Thus, by learning this translation real world old photos r can be restored by sequentially performing,
- $r_{R \rightarrow Y} = G_Y \circ T_Z \circ E_R(r)$



Shared VAE Architecture

- First Stage, r and x are trained on VAE1 and y is trained on VAE2
- The domain gap between Zr and Zx is closed by jointly training an adversarial discriminator
- Finally, the mapping is learned that restores the corrupted images to clean ones in the latent space



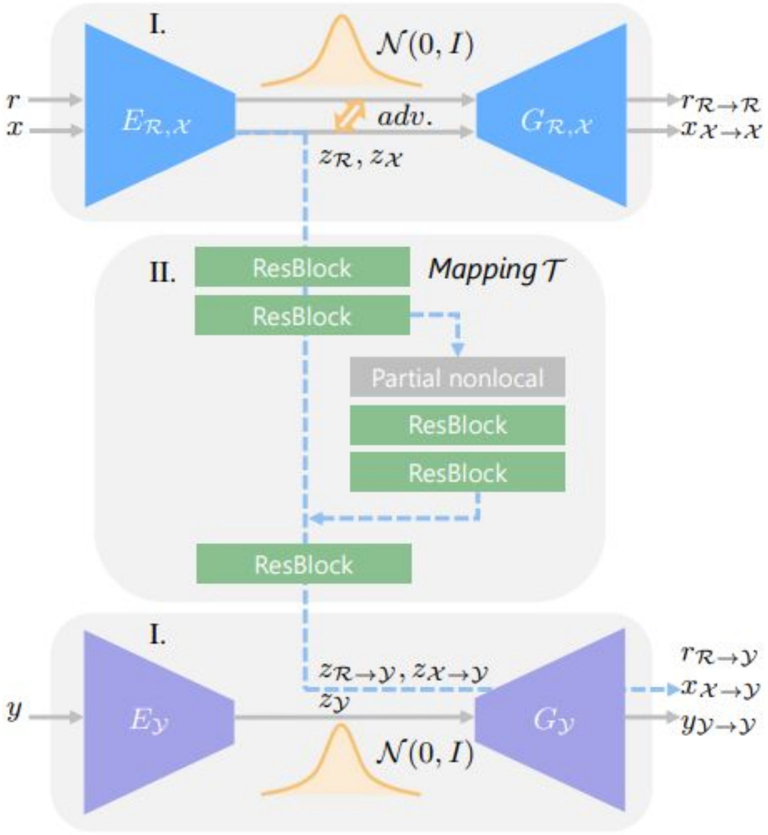
Objective functions

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

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

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

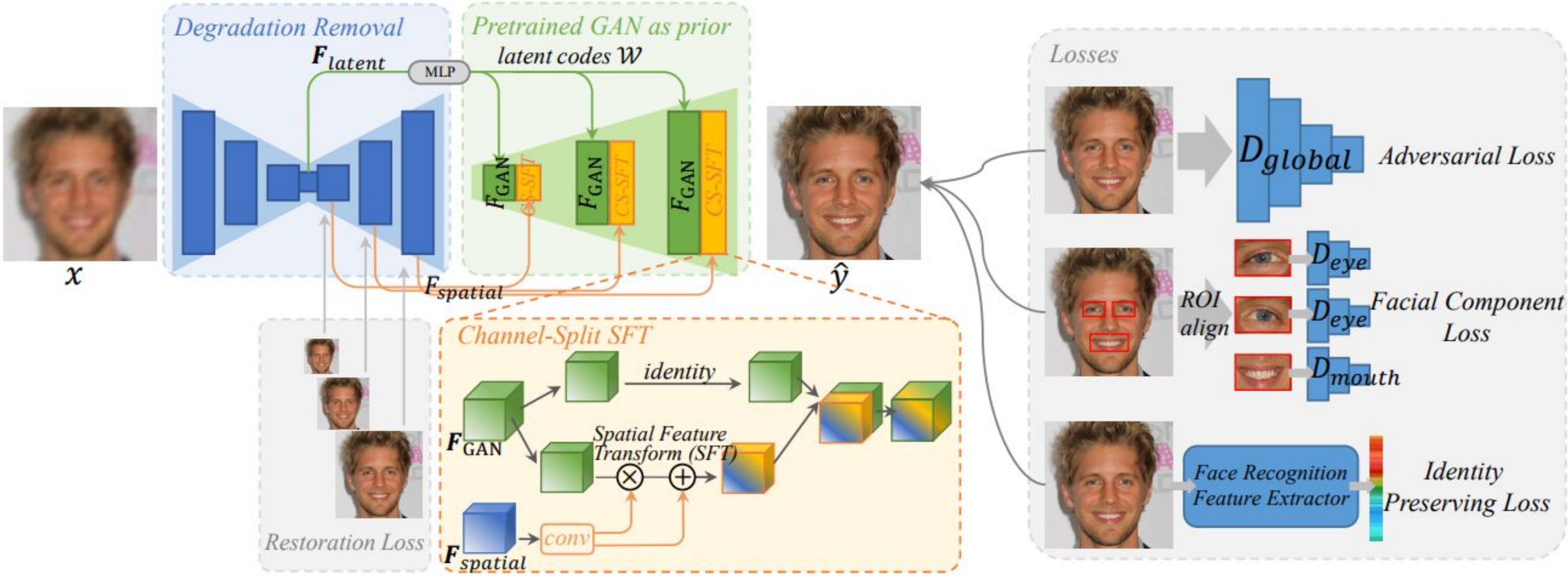
$$\mathcal{L}_{\mathcal{T}}(x,y) = \lambda_1\mathcal{L}_{\mathcal{T},\ell_1} + \mathcal{L}_{\mathcal{T},\text{GAN}} + \lambda_2\mathcal{L}_{\text{FM}}$$



Towards Real-World Blind Face Restoration with Generative Facial Prior (GFP-GAN)

- The task of Blind face restoration relies on facial priors, such as facial geometry prior or reference prior, to restore realistic and faithful details
- They leverage Generative Facial Prior (GFP) for real-world blind face restoration
- Which is the prior implicitly encapsulated in pretrained face Generative Adversarial Network (GAN) models such as StyleGAN
- Additionally, it also contains a degradation removal module

GFP-GAN Architecture

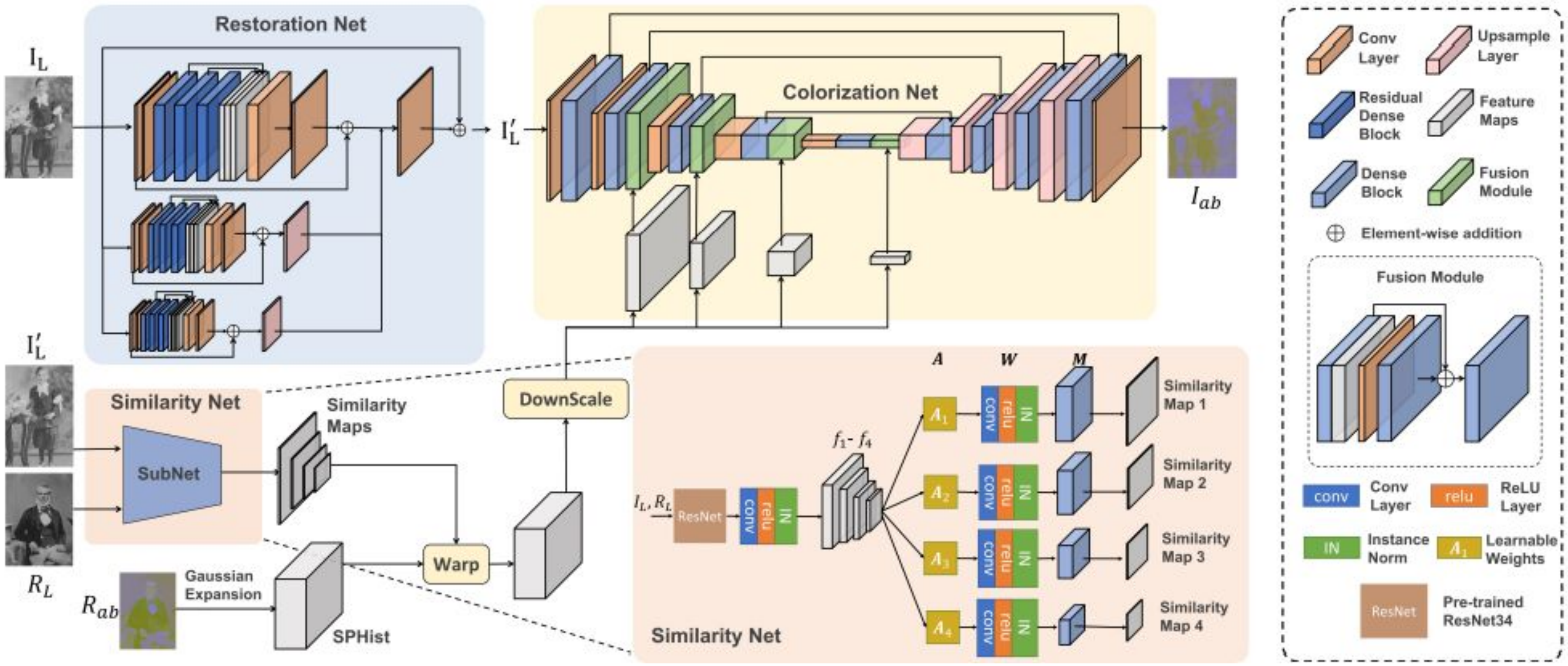


- Degradation removal followed by pretrained face GAN prior
- They are bridged by latent code mapping and Channel-Split Spatial Feature Transform (CS-SFT) layers

Pik-Fix: Restoring and Colorizing Old Photos

- It is a novel reference-based end-to-end learning framework that is able to both repair and colorize old and degraded pictures
- The framework is divided into several sub-networks
- A restoration sub-network that conducts restoration from degradations
- A similarity sub-network that performs color histogram matching and color transfer
- A colorization subnet that learns to predict the chroma elements of images

Pik-Fix Architecture



Sub-Networks

Restoration Sub-Net

- Restores degradations like physical defects (cracks,tears) and capture effects (blur, noise)
- It is Residual Dense Network that is multi-layered to preserve the broader range of distortions

Similarity Sub-Net

- The similarity sub-net is designed to project the reference image features onto the feature space of the input picture.
- It consists of a pre-trained ResNet34 from which the feature maps are retrieved from both input and reference pictures

Colorization Sub-Net

- Consists of guiding process to colorize the picture using the color prior in the reference picture
- The prior is a space-preserving color histogram (SPHist)
- The network is a U-Net with dense blocks in the encoder

Summarizing the models

Bringing Old Photos Back to Life (Shared VAE):

- The domain gap is reduced between old photos and synthetic images, and the translation to clean images is learned in latent space.
- This method suffers less from generalization issue compared with prior methods

GFP-GAN:

- The GFP-GAN framework that leverages the rich and diverse generative facial prior for the challenging blind face restoration task.
- It is limited to restoration of photos with faces

Pik-Fix:

- It is the first end-to-end system that is able to simultaneously restore and colorize old photos.
- Uses 3 subnetworks, each handles a specific degradation but is trained holistically

Project Idea

- Going back to the first model that represents the images in latent space.
- This latent space can be exploited using the latent diffusion model for old photo restoration.
- Latent diffusion model has to capabilities for inpainting and super-resolution thus it seems to be capable enough to solve this problem.
- Starting with one of the degradations like blur and noise and then move onto other degradations