a large amount of synthetic images.

The same framework, however, does not apply to old photo restoration. First, the degradation process of old photos is rather complex, and there exists no degradation model that can realistically render the old photo artifact. Therefore, the model learned from those synthetic data generalizes poorly on real photos. Second, old photos are plagued with a compound of degradations and inherently requires different strategies for repair: unstructured defects that are spatially homogeneous, *e.g*., film grain and color fading, should be restored by utilizing the pixels in the neighborhood, whereas the structured defects, *e.g*., scratches, dust spots, etc., should be repaired with a global image context.

To circumvent these issues, we formulate the old photo restoration as a triplet domain translation problem. Different from previous image translation methods [13], we leverage data from three domains (*i.e*., real old photos, synthetic images and the corresponding ground truth), and the translation is performed in latent space. Synthetic images and the real photos are first transformed to the same latent space with a shared variational autoencoder [14] (VAE). Meanwhile, another VAE is trained to project ground truth clean images into the corresponding latent space. The mapping between the two latent spaces is then learned with the synthetic image pairs, which restores the corrupted images to clean ones. The advantage of the latent restoration is that the learned latent restoration can generalize well to real photos because of the domain alignment within the first VAE. Besides, we differentiate the mixed degradation, and propose a partial nonlocal block that considers the long-range dependencies of latent features to specifically address the structured defects during the latent translation. In comparison with several leading restoration methods, we prove the effectiveness of our approach in restoring multiple degradations of real photos.

2. Related Work

Single degradation image restoration. Existing image degradation can be roughly categorized into two groups: unstructured degration such as noise, blurriness, color fading, and low resolution, and structured degradation such as holes, scratches, and spots. For the former unstructured ones, traditional works often impose different image priors, including non-local self-similarity [15, 16, 17], sparsity [18, 19, 20, 21] and local smoothness [22, 23, 24]. Recently, a lot of deep learning based methods have also been proposed for different image degradation, like image denoising [5, 6, 25, 26, 27, 28, 29], super-resolution [7, 30, 31, 32, 33], and deblurring [8, 34, 35, 36].

Compared to unstructured degradation, structured degradation is more challenging and often modeled as the "image painting" problem. Thanks to powerful semantic modeling ability, most existing best-performed inpainting methods are learning based. For example, Liu et al. [37] masked out the hole regions within the convolution operator and enforces the network focus on non-hole features only. To get better inpainting results, many other methods consider both local patch statistics and global structures. Specifically, Yu et al. [38] and Liu et al. [39] proposed to employ an attention layer to utilize the remote context. And the appearance flow is explicitly estimated in Ren et al. [40] so that textures in the hole regions can be directly synthesized based on the corresponding patches.

No matter for unstructured or structured degradation, though the above learning-based methods can achieve remarkable results, they are all trained on the synthetic data. Therefore, their performance on the real dataset highly relies on synthetic data quality. For real old images, since they are often seriously degraded by a mixture of unknown degradation, the underlying degradation process is much more difficult to be accurately characterized. In other words, the network trained on synthetic data only, will suffer from the domain gap problem and perform badly on real old photos. In this paper, we model real old photo restoration as a new triplet domain translation problem and some new techniques are adopted to minimize the domain gap.

Mixed degradation image restoration. In the real world, a corrupted image may suffer from complicated defects mixed with scratches, loss of resolution, color fading, and film noises. However, research solving mixed degradation is much less explored. The pioneer work [41] proposed a toolbox that comprises multiple light-weight networks, and each of them responsible for a specific degradation. Then they learn a controller that dynamically selects the operator from the toolbox. Inspired by [41], [42] performs different convolutional operations in parallel and uses the attention mechanism to select the most suitable combination of operations. However, these methods still rely on supervised learning from synthetic data and hence cannot generalize to real photos. Besides, they only focus on unstructured defects and do not support structured defects like image inpainting. On the other hand, Ulyanov et al. [43] found that the deep neural network inherently resonates with low-level image statistics and thereby can be utilized as an image prior for blind image restoration without external training data. This method has the potential, though not claimed in [43], to restore in-the-wild images corrupted by mixed factors. In comparison, our approach excels in both restoration performance and efficiency.

Old photo restoration. Old photo restoration is a classical mixed degradation problem, but most existing methods [1, 2, 3, 4] focus on inpainting only. They follow a similar paradigm *i.e*., defects like scratches and blotches are first identified according to low-level features and then inpainted by borrowing the textures from the vicinity. How-