where $\phi^i_{D_\mathcal{T}}$ (ϕ^i_{VGG}) denotes the i^{th} layer feature map of the discriminator (VGG network), and $n^i_{D_\mathcal{T}}$ (n^i_{VGG}) indicates the number of activations in that layer.

3.2. Multiple degradation restoration

The latent restoration using the residual blocks, as described earlier, only concentrates on local features due to the limited receptive field of each layer. Nonetheless, the restoration of structured defects requires plausible inpainting, which has to consider long-range dependencies so as to ensure global structural consistency. Since legacy photos often contain mixed degradations, we have to design a restoration network that simultaneously supports the two mechanisms. Towards this goal, we propose to enhance the latent restoration network by incorporating a global branch as shown in Figure 3, which composes of a nonlocal block [49] that considers global context and several residual blocks in the following. While the original block proposed in [49] is unaware of the corruption area, our nonlocal block explicitly utilizes the mask input so that the pixels in the corrupted region will not be adopted for completing those area. Since the context considered is a part of the feature map, we refer to the module specifically designed for the latent inpainting as *partial nonlocal block*.

Formally, let $F \in \mathbb{R}^{C\times HW}$ be the intermediate feature map in M (C, H and W are number of channels, height and width respectively), and $m \in \{0,1\}^{HW}$ represents the binary mask downscaled to the same size, where 1 represents the defect regions to be inpainted and 0 represents the intact regions. The affinity between i^{th} location and j^{th} location in F, denoted by $s_{i,j} \in \mathbb{R}^{HW\times HW}$, is calculated by the correlation of F_i and F_j modulated by the mask $(1-m_j)$, *i.e.*,

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

where,

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

gives the pairwise affinity with embedded Gaussian. θ and ϕ project F to Gaussian space for affinity calculation. According to the affinity $s_{i,j}$ that considers the holes in the mask, the partial nonlocal finally outputs

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

which is a weighted average of correlated features for each position. We implement the embedding functions θ, ϕ, μ and ν with 1×1 convolutions.

We design the global branch specifically for inpainting and hope the non-hole regions are left untouched, so we fuse the global branch with the local branch under the guidance

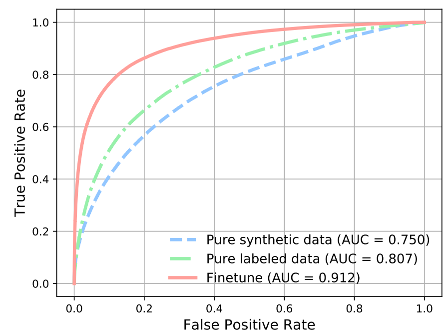

Figure 4: **ROC curve for scratch detection of different data settings.**

of the mask, *i.e.*,

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

where operator $\odot$ denotes Hadamard product, and ρ_{local} and ρ_{global} denote the nonlinear transformation of residual blocks in two branches. In this way, the two branches constitute the latent restoration network, which is capable to deal with multiple degradation in old photos. We will detail the derivation of the defect mask in Section 4.1.

4. Experiment

4.1. Implementation

Training Dataset We synthesize old photos using images from the Pascal VOC dataset [50]. In order to render realistic defects, we also collect scratch and paper textures, which are further augmented with elastic distortions. We use layer addition, lighten-only and screen modes with random level of opacity to blend the scratch textures over the real images from the dataset. To simulate large-area photo damage, we generate holes with feathering and random shape where the underneath paper texture is unveiled. Finally, film grain noises and blurring with random amount are introduced to simulate the unstructured defects. Besides, we collect 5,718 old photos to form the images old photo dataset.

Scratch detection To detect structured area for the parital nonlocal block, We train another network with Unet architecture [51]. The detection network is first trained using the synthetic images only. We adopt the focal loss [52] to remedy the imbalance of positive and negative detections. To further improve the detection performance on real old photos, we annotate 783 collected old photos with scratches, among which we use 400 images to finetune the detection network. The ROC curves on the validation set in Figure 4 show the effectiveness of finetuning. The area under the curve (AUC) after finetuning reaches 0.91.

Training details We adopt Adam solver [53] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate is set to 0.0002 for the first 100 epochs, with linear decay to zero thereafter.