mathematics

MDPI

Article

Domain Adaptation Based on Human Feedback for Enhancing Image Denoising in Generative Models

Hyun-Cheol Park [1], Dat Ngo [1] and Sung Ho Kang [2,*]

[1] Department of Computer Engineering, Korea National University of Transportation, 50, Daehak-ro, Daesowon-myeon, Chungju-si 27469, Republic of Korea; hc.park@ut.ac.kr (H.-C.P.); datngo@ut.ac.kr (D.N.)
[2] National Institute for Mathematical Sciences, 70, Yuseong-daero 1689 beon-gil, Yuseong-gu, Daejeon 34047, Republic of Korea
* Correspondence: runits@nims.re.kr

Abstract: How can human feedback be effectively integrated into generative models? This study addresses this question by proposing a method to enhance image denoising and achieve domain adaptation using human feedback. Deep generative models, while achieving remarkable performance in image denoising within training domains, often fail to generalize to unseen domains. To overcome this limitation, we introduce a novel approach that fine-tunes a denoising model using human feedback without requiring labeled target data. Our experiments demonstrate a significant improvement in denoising performance. For example, on the Fashion-MNIST test set, the peak signal-to-noise ratio (PSNR) increased by 94%, with an average improvement of 1.61 ± 2.78 dB and a maximum increase of 18.21 dB. Additionally, the proposed method effectively prevents catastrophic forgetting, as evidenced by the consistent performance on the original MNIST domain. By leveraging a reward model trained on human preferences, we show that the quality of denoised images can be significantly improved, even when applied to unseen target data. This work highlights the potential of human feedback for efficient domain adaptation in generative models, presenting a scalable and data-efficient solution for enhancing performance in diverse domains.

Keywords: generative adversarial network; human feedback; domain adaptation; unseen domain; denoising

MSC: 68T07

Academic Editors: Junlin Hu and Jie Wen

Received: 6 January 2025
Revised: 7 February 2025
Accepted: 10 February 2025
Published: 12 February 2025

Citation: Park, H.-C.; Ngo, D.; Kang, S.H. Domain Adaptation Based on Human Feedback for Enhancing Image Denoising in Generative Models. *Mathematics* **2025**, *13*, 598. https://doi.org/10.3390/math13040598

1. Introduction

Deep generative models have achieved remarkable success in image generation tasks [1–3]. In particular, generative adversarial networks (GANs) are widely known as a fundamental theory that demonstrates how to generate realistic images. Recently, GANs are also utilized for specific purposes such as image denoising [4–6], super-resolution [7,8], and style transfer [9–12]. These objectives involve training GANs using supervised learning with paired data sets aiming to learn the target distribution, which has been shown to yield successful results.

However, despite the impressive performance of these models within their training domain, they often encounter challenges when applied to unseen domains, resulting in subpar outputs. In the context of GANs based on image-to-image generation [13], which aim to preserve the original intrinsic characteristics while learning the target distribution, both successful and unsuccessful cases can emerge during testing on unseen domains. For

example, when presented with ten samples from an unseen domain, seven of them may yield successful translations, while the remaining three produce unsatisfactory results. This raises the question: should we simply discard these three failed samples, or is there a way to enhance and improve them to achieve better outcomes? Obtaining ground-truth data for the unseen domain could facilitate domain-specific training; however, in many practical scenarios, acquiring target data for unseen domains poses significant challenges. As an alternative, applying domain adaptation methods [14–21] can mitigate this issue, but even domain-adapted models may still yield failed results based on human preferences.

Our objective diverges from conventional domain adaptation approaches. As demonstrated in [14,19], domain adaptation focuses on training methods that aim to minimize the distinguishability between the source and target domain distributions in the latent space. This macro-level approach seeks to minimize the overall gap between source and target domains. However, at a micro-level, there remain opportunities for improvement in the generated results. Therefore, our goal is to address and rectify instances of failure within the output produced by the trained model.

Similarly, during the training of ChatGPT [22], it excels at generating high-quality language responses through extensive pre-training on vast data sets. Nevertheless, upon human evaluation, the generated sentences may exhibit a dichotomy: some appear naturally flowing, while others seem less fluent. To bridge this gap, ChatGPT [22] leverages human feedback [23] to enhance its ability to produce more seamlessly natural sentences. Furthermore, in the domain of aligning text-to-image models [24], the introduction of human feedback has demonstrated significant improvements in model performance. However, research on model refinement through human feedback in GANs is still scarce, and through our paper, we aim to showcase the potential of model refinement through the profound influence of human feedback.

Recently, drawing inspiration from the success of reinforcement learning human feedback [25] in language domains, we present an innovative approach for unseen domain adaptation based on human feedback. Analogous to how children learn from the feedback provided by their parents, we adopt a similar strategy. For instance, if a child learns how to remove noise from the background of a single image, they can subsequently apply denoising techniques to new images. While the quality of the denoised image may vary, receiving feedback from a parent can lead to improvement. Even if we cannot surpass our previous achievements, we can still imitate and learn from them. This approach shows promise in addressing the challenges of unsupervised unseen domain adaptation and opens new possibilities for model enhancement through the profound influence of human feedback. As we explore this innovative avenue, our aim is to make significant contributions to the field of AI and foster advancements in human-guided domain adaptation research.

To achieve this, we introduce a deep feedback network that utilizes human feedback to the adaptation of an unlabeled target domain. As illustrated in Figure 1, to replicate restricted learning circumstances, we conduct experiments on the denoising problem. Initially, we train the model using a restrictive training approach, focusing solely on denoising the digit '0' within the MNIST data set. Subsequently, we evaluate the model's performance on the Fashion-MNIST data set, which represents an unseen domain. It becomes evident that the pre-trained model, trained on MNIST, produces unintended results when applied to the unseen domain. To adapt to the unseen domain, we introduce a training method based on human feedback. Human feedback assesses the model's results in the unseen domain as either 'Good' or 'Bad'. The model is then fine-tuned using the gradient of these assessments. This approach shows promising potential for efficiently fine-tuning the model using feedback from generators trained in other domains.

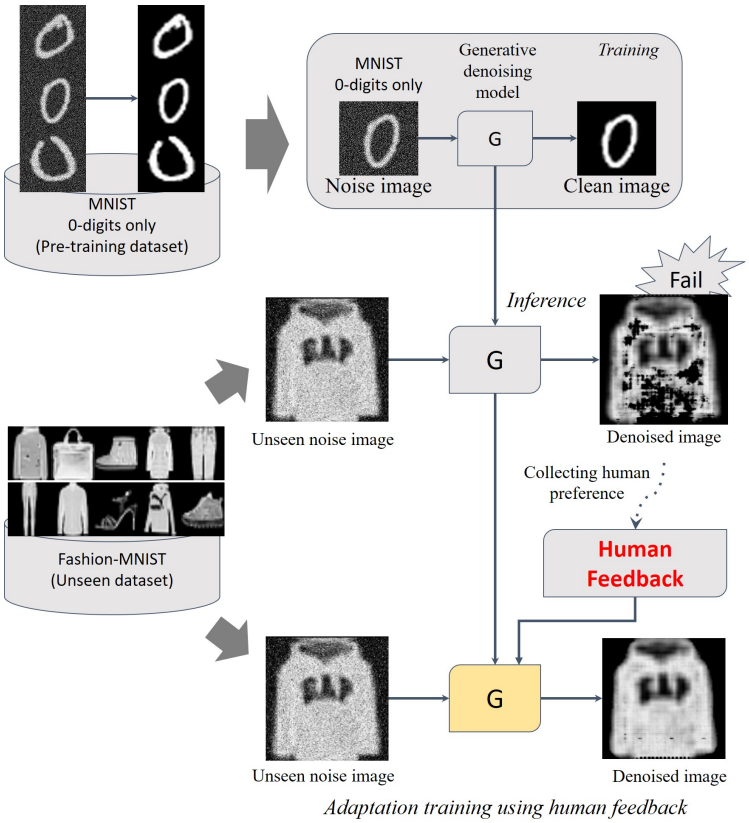

Figure 1. Overview of adaptation training: Step 1 involves pre-training the generator, Step 2 uses a reward model trained with human feedback, and Step 3 fine-tunes the generator for domain adaptation.

We can summarize our main contributions as follows

- We propose an adaptation method for the domain of image-based generative models through human feedback.
- We perform domain adaptation while maintaining the quality of the generated image using an optional loss function with a reward model using human feedback.
- We show that the model can be adapted by human feedback, even in the absence of labeled target data.

2. Methods

Our overall process consists of three steps. First, the denoising model is pre-trained in the basis domain, serving as the fundamental ability for denoising. Next, the reward model is trained using human feedback. To train the reward model, humans manually annotate denoised images as either *Good* or *Bad*. Finally, the basis generator is re-trained using the reward model. Even if the generator produces denoised images of low quality, it will be trained to prioritize good results based on the provided human feedback.

2.1. Pre-Training Basis Domain for Denoising

In this step, we focused on creating an intentional class-biased generator. The model is trained to acquire the fundamental ability of denoising using simple images, as shown in Step 1 of Figure 2. The architecture of the model consists of generative adversarial networks (GANs). We employed the pix2pix [26] model as our baseline, which relies on paired training. To train the model, a paired data set is required, consisting of both clean and noisy images. For our paired training data set, we used only 0 digits in the MNIST data set. To create a pair, 0-digit images were selected and combined with synthesized noise.

Consider the synthesized noise of image z, which is a 2D image represented as $z \in R^{m \times n}$. It is composed of both the original image and noise, denoted as x and n, respectively:

$$z = x + n. \tag{1}$$

We assume that the clean image is selected from the source domain. Therefore, the synthesized noise image z and the original image x are treated as paired data. For convenience notation, the source and unseen domain data are denoted as z_s and z_u, respectively.

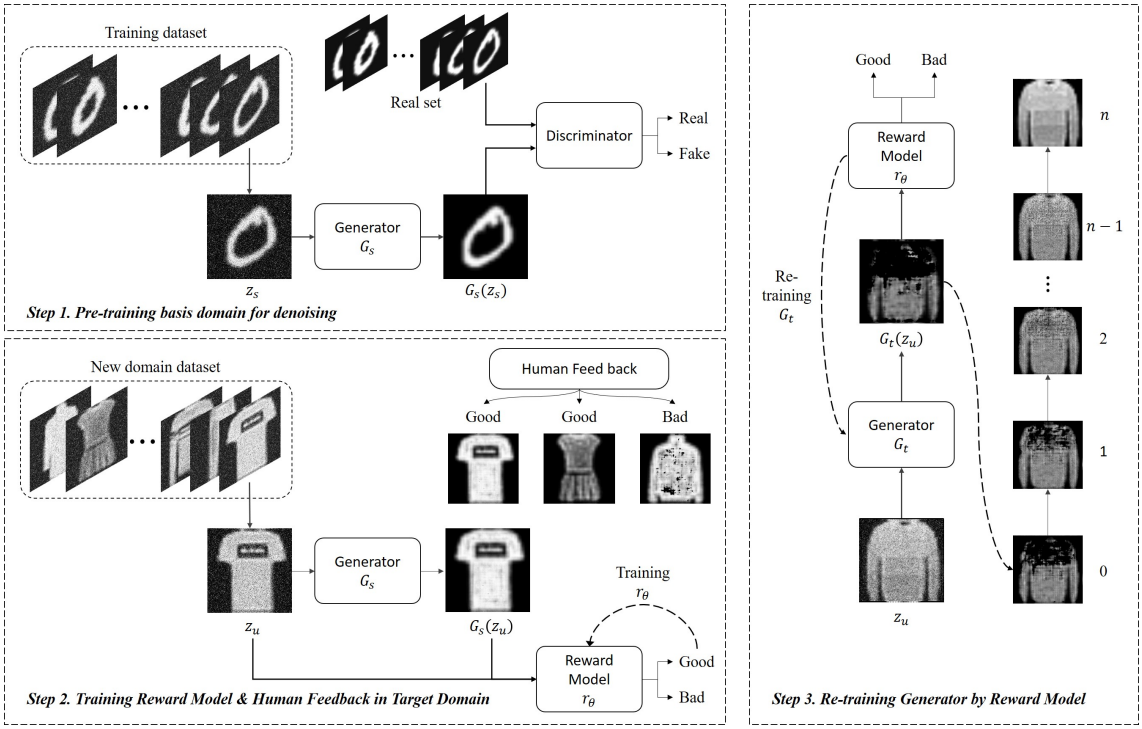

Figure 2. The training data set '0' is sourced from the MNIST data set, while the new domain data set is the Fashion-MNIST data set. The G_s model in Step 2 is trained on the MNIST data set during Step 1. Subsequently, the G_s model in Step 3 is fine-tuned using the reward model based on human feedback.

The generator is trained to produce samples of good quality from input noise variables p_n. To train the model on the source domain, the final loss is defined as follows:

$$L_{step_1}(G_s, D) = L_{GAN}(G_s, D) + L_{pixel-wise}(G_s) \tag{2}$$

where the samples $G_s(z_s)$ are obtained when $z_s \sim p_n$ follows a distribution that represents good quality in the source domain. In other words, The generator G_s is trained to learn the mapping from the noise image z to the clean image x, denoted as $G_s : z_s \to x$. The objective of the generator is to estimate the distribution of x, denoted as $G_s(z_s) \approx x$. To achieve this, the GAN consists of an adversarial discriminator D, which distinguishes between 'Real' and 'Fake' images. 'Real' refers to the original image x, while 'Fake' corresponds to the generated image $G_s(z_s)$ produced by the generator. Both the generator G_s and the discriminator D are trained adversarially. The objective function can be expressed as follows:

$$\min_{G_s} \max_{D} L_{GAN}(G_s, D) = \mathbb{E}_{z_s, x} \left[\log D(z_s, x)\right] + \mathbb{E}_{z_s \sim p_n(z_s)} \left[\log\left(1 - D(z_s, G_s(z_s))\right)\right]. \tag{3}$$

where $G_s(z_s)$ represents the generation of a clean image from a noisy image z_s. The discriminator D is responsible for classifying between the real and fake distributions. In order to induce a mistake in D, G_s aims to minimize Equation (3). On the other hand, D maximizes the objective function to distinguish between real and generated images.

In our study, we tackle the problem of denoising while preserving the underlying morphological structures. Traditional GAN [1] frameworks approximate the target distribution during training. However, in the context of image processing, the generated images may inadvertently alter the essential morphological characteristics of the originals [26–28]. To mitigate this issue and ensure the preservation of morphological structures, an auxiliary loss term is incorporated into the objective function:

$$L_{pixel-wise}(G_s) = \mathbb{E}_{z_s,x}\left[\|x - G_s(z_s)\|_1\right]. \tag{4}$$

Similar to the [26] approach, the auxiliary loss employs the $L1$ distance between the target image x and the generated image $G_s(z_s)$.

2.2. Human Feedback and Training Reward Model

The integration of human feedback has demonstrated high adaptability across various domains [24,25,29]. This valuable information is used to train a reward model, which acts as a substitute for human assessment and enhances the model's performance. In this section, we provide a detailed description of how human feedback is gathered. In the previous section, we presented the basic denoising GAN model using pix2pix [26], which we referred to as the supervised denoising model (SDM). Human feedback is obtained through manual assessments of the SDM results from unseen domain samples z_u. The assessments are categorized as 'Good' if the image was clean and 'Bad' if the image contained noise or collapse (see Step 2 in Figure 2).

The assessments 'Good' and 'Bad' are utilized as ground truth labels ($y_r = 0, 1$) to train the reward model. The reward model, denoted as r_θ, follows the same architecture as the discriminator in the SDM. The loss function for r_θ is as follows:

$$L_{reward}(\hat{G}_s, r_\theta) = \min_{r_\theta} \mathbb{E}_{z_u \sim p_n(z_u)} - [y_r \log r_\theta(\hat{G}_s(z_u), z_u) + (1 - y_r) \log\left(1 - r_\theta(\hat{G}_s(z_u), z_u)\right)]. \tag{5}$$

where $\hat{G}_s$ is a frozen denoising generator model using the source domain. During the training of the reward model, $\hat{G}_s$ remains untrainable and is solely used to generate denoised images. r_θ assesses these denoised images and is trained using y_r labels. Notably, the reward model can be trained to capture human preferences, as the y_r labels are collected through human feedback.

2.3. Objective

In this section, we present the final formulation of the loss function, which consists of auxiliary terms. Each auxiliary term includes reward loss, consistency loss, and regularization loss, used to train G_t. Here, G_t represents the adapted model, which is fine-tuned from G_s in the unseen domain. Thus, the architecture and initial parameters of G_t are the same as those of G_s.

Reward Loss L_r: The primary objective of generator G_t is to generate denoised images that are assessed by the reward model as '*Good*' (0, indicating clean images). The minimization of L_r aims to train generator G_t to generate clean images. In other words, the reward loss

L_r trains G_t to map from the distribution of '*Bad*' quality images (distribution j) to the distribution of '*Good*' quality images (distribution k), $G_t : j \rightarrow k$.

$$L_r(G_t) = \mathbb{E}_{z_u \sim p_n(z_u)} \left[-\log \left(1 - \hat{r_\theta}(G_t(z_u), z_u) \right) \right]. \tag{6}$$

where $\hat{r_\theta}$ is a reward model trained on human feedback that has fixed parameters. Thus, $\hat{r_\theta}$ only assesses the quality of the generated image from G_t and the input image z_u.

In this context, by fine-tuning G_t from G_s using L_r loss, G_t is able to closely approximate the $x \sim p_{data}$ distribution represented by r_θ in an unseen domain. However, relying solely on L_r loss for training G_t may lead to over-fitting and the risk of distorting the morphological information of the original images. To alleviate this problem, we describe 'Regularization Loss' and 'Consistency Loss' as follows.

Consistency Loss L_p**:** As the model learns from new data, there is a potential issue of the performance of past good results deteriorating due to parameter updates. This is commonly referred to as the problem of catastrophic forgetting. To control this issue, it is necessary to compare the outcomes of the initial parameters with the current results. We present a novel compensatory term, denoted as L_p, which facilitates a comparison between the outputs of the initial frozen generator, $\hat{G}_s$, and the target generator G_t. The primary objective of L_p is to minimize the pixel-wise L_1 loss between the outcomes generated by $\hat{G}_s$ and G_t, thereby ensuring that the current model preserves crucial insights acquired from the initial generator throughout the training procedure. By incorporating this approach, we effectively address the issue of neglecting important details and consequently witness a notable enhancement in the overall performance of the current model.

$$L_p(G_t) = \mathbb{E}_{z_u \sim p_n(z_u)} \left[\sigma(\hat{r_\theta}(\hat{G}_s(z_u), z_u)) \| \hat{G}_s(z_u) - G_t(z_u) \|_1 \right]. \tag{7}$$

$$\sigma(r) = \begin{cases} 0 & \text{if } r \geq \epsilon \\ 1 & \text{if } r < \epsilon \end{cases} \tag{8}$$

where σ is the step function, and r denotes the result of the reward. ϵ is a threshold value ranging from 0 to 1.

Regularization Loss L_n**:** We employ a regularization loss term to address the issues of over-fitting and mode collapse. In existing methods, the difference in cosine similarity of feature vectors in the latent space has been compared [21,30]. However, in our approach, we intuitively compare the outputs of the model from the past and the current training stages to suppress excessive variations caused by the model's learning. L_n calculates the pixel-wise L_1 loss between the results of the current generator and the $(n-i)th$ generator $\hat{G}_t$. The generator $\hat{G}_t$ copies weights from G_t every N steps and then freezes them.

$$L_n(G_t) = \mathbb{E}_{z_u \sim p_n(z_u)} \left[\| \hat{G}_t(z_u) - G_t(z_u) \|_1 \right]. \tag{9}$$

The final loss used to train G_t is as follows:

$$\min_{G_t} L(G_t) = L_r(G_t) + \alpha L_p(G_t) + (1 - \alpha) L_n(G_t). \tag{10}$$

where, α controls the relative balance of the two L_p and L_n losses. In the ablation study, we analyze the impact of each auxiliary loss on the final loss. Figure 3 represents the flow diagram of the more specific final objective loss functions we designed.

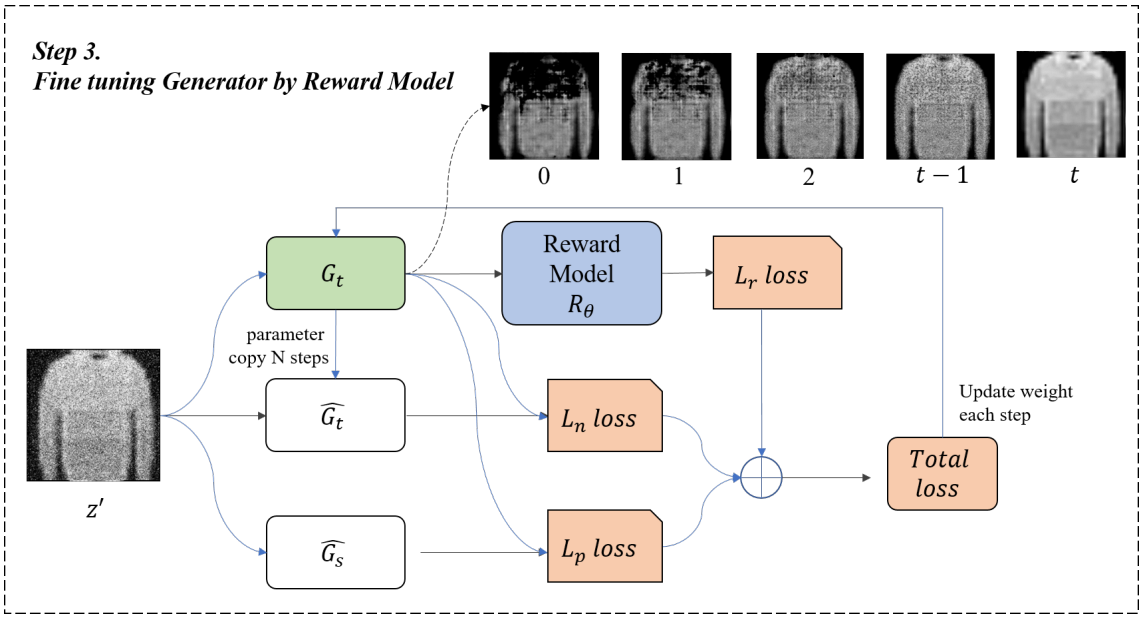

Figure 3. Flow diagram of the final objective loss functions: L_r drives domain adaptation, while L_p prevents catastrophic forgetting, and L_n ensures regularization.

3. Experiments

3.1. Data Sets

We utilized two data sets in our experiments: MNIST [31] and Fashion-MNIST [32]. MNIST is a grayscale image data set consisting of 10 classes representing digits from 0 to 9. Each MNIST image has dimensions of 28 × 28 pixels. The data set comprises a training set of 60,000 images and a test set of 10,000 images.

In the experiments, the MNIST data set serves as the source domain for training the initial denoising generator. Specifically, only the '0' digit is used for restrictive training on the source domain. The training set consists of 6000 samples, and the validation set contains 1000 samples. The MNIST images are resized to a size of 256 × 256 pixels using bicubic interpolation. To train the initial denoising generator, a pair of data points is required, consisting of clean and noisy images. The original MNIST data set is used as the clean image counterpart, while the noisy images are created by introducing artifact noise in the form of salt-and-pepper noise combined with Gaussian noise. The salt-and-pepper noise was applied with equal proportions of salt and pepper pixels at 50%, while the Gaussian noise was added with a mean of 0 and a standard deviation of 0.05, perturbing the pixel intensity values. This approach ensures a controlled and reproducible noise level, typical for evaluating image denoising models [33].

Fashion-MNIST is a data set consisting of images representing 10 types of fashion items. It also includes a training set of 60,000 images and a test set of 10,000 images. Fashion-MNIST is employed to evaluate the model's performance and train adaptive learning. The images in Fashion-MNIST are resized to 256 × 256 pixels and similarly augmented with noise, as conducted with the MNIST data set.

3.2. Training Setting

Pre-training for denoising: "pix2pix" [26] is employed as the baseline model in this experiment. The main objective of most GANs is to establish a mapping $G : Z \rightarrow X$. "pix2pix" demonstrated the training approach for pixel-wise mapping between input and output images. Consequently, the generator of "pix2pix" can effectively learn the transformation from the noise space Z to the clean space X. In this experiment, we trained a denoising model, denoted as G_s, using the MNIST data set. G_s was specifically trained using

a set of 1000 image pairs consisting of clean digits and their corresponding noisy versions. The clean images used in the training process were specifically selected to represent the digit '0'. For optimization, we employed the Adam solver [34] with a batch size of 10, a learning rate of 0.0002, and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The denoising model was trained for 200 epochs.

Inference and human feedback: In this paper, our proposed method demonstrates the adaptability of a pre-trained model to a target domain through human feedback. To gather human feedback, the pre-trained generate model G_s is used to infer results in the target domain, which are then manually assessed by human evaluators. In our experiments, we employ Fashion-MNIST as the target domain data set, and we collect human feedback for the 10,000 test images in this data set.

Training for reward model by human feedback: The reward model, denoted as r_θ, is utilized in the auxiliary loss term. The architecture of the reward model is designed to be the same as the discriminator of the "pix2pix" model. The hyperparameters used for training r_θ remain consistent with the pre-training setting.

Adaptive training by human feedback: Note that the adaptive training process implements Equation (10), utilizing the same set of hyperparameters as mentioned above. It is important to note that G_t has trainable parameters, whereas $\hat{G}_s$, $\hat{G}_t$, and $\hat{r_\theta}$ are untrainable parameters. The constant ϵ in Equation (8) is set to 0.2, and the constant α in Equation (10) is set to 0.9. In the ablation study, we examine the influence of L_p and L_n as α is varied.

3.3. Evaluation

We evaluate the quality of the denoised images using the metrics of PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure). PSNR is a widely used metric for evaluating denoising models. It measures the quality of the denoised image by comparing it to the original (clean) image. Higher PSNR values indicate better denoising performance. PSNR can be calculated using the mean squared error (MSE) between the denoised image and the original image. SSIM is another popular metric that quantifies the similarity between the denoised image and the original image. It takes into account not only pixel-level differences but also structural information, such as luminance, contrast, and structure. Higher SSIM values indicate better preservation of structural details.

3.4. Results of Domain Adaptive Denoising by Human Feedback

Comparison of evaluation metrics: In this section, we examine the results of domain adaptive denoising. Our intuition is that, even when presented with unseen data from a target domain, if we provide human feedback to a supervised learning model, the model can adapt to the data effectively. Note that our human feedback is not ground-truth for a denoised image; it shows human preference, consisting of '*Good*' and '*Bad*'. In Table 1, '$G_s(z)$ vs. x' represents the denoising results of the model before adaptive learning using human feedback. '$G_t(z)$ vs. x' shows the denoising outcomes of the adapted model based on human feedback. The results obtained from the MNIST data set indicate the performance in the pre-trained domain, while the results from the Fashion-MNIST data set reflect the performance on unseen data. Therefore, we can observe the adaptation progress between the initial model G_s and the updated model G_t. In our experiments, both G_s and G_t demonstrated a significant improvement in PSNR measured on the Fashion-MNIST test set, with an overall increase of 94% over the entire 10K data set. The statistical analysis of the PSNR improvement revealed a mean increase of 1.61 ± 2.78 dB (MAX: 18.21, MIN: 0.0001). Figure 4a shows the images with the most significant increase in PSNR, along with the corresponding metrics between each generator output image and the ground truth image. In addition, for the remaining 6% of cases, there was a decrease in PSNR

values, with the statistical analysis showing a mean decrease of 0.12 ± 0.18 dB (MAX: 2.75, MIN: 0.0001) (See Figure 4b). Figure 5 shows the boxplot of PSNR for each generator G_s and G_t on the experimental data set. The G_t images from the same data set exhibit higher PSNR values, indicating improved image quality after adaptation. Particularly noteworthy is that after adaptation, the PSNR and SSIM values of the MNIST test set (10K) from the G_t generator, corresponding to the source domain, show little to no variation or even slight improvement (see Figure 5a). This demonstrates the prevention of the catastrophic forgetting issue for the source domain even after adaptation to the target domain. Furthermore, we apply the G_t model tuned on the Fashion-MNIST test set with the reward model to the Fashion-MNIST training set (60k). This demonstrates that when the reward model is trained in a new domain, it can effectively work without requiring additional training.

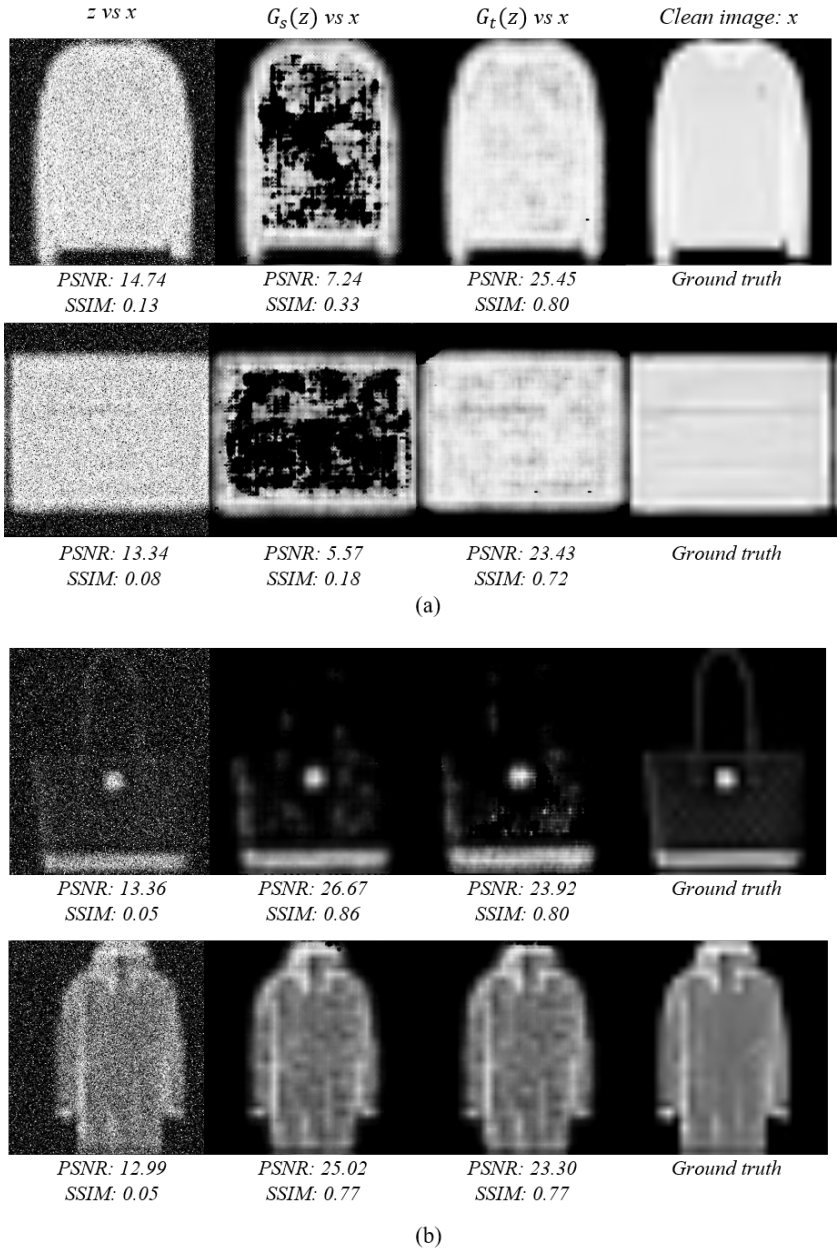

Figure 4. Visual results for adaptation. The PSNR and SSIM values for each image are calculated with respect to the ground truth. G_s represents the model pre-trained on MNIST, while G_t represents the model fine-tuned from G_s using human feedback. (**a**) Sample images with the most significant increase in PSNR from G_s and G_t output. (**b**) Most decreased PSNR images.

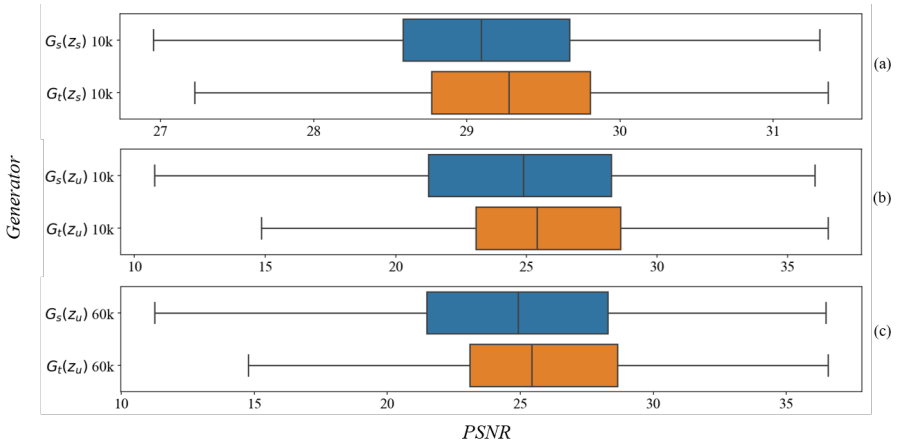

Figure 5. Boxplot of PSNR for each generator G_s and G_t on the experimental data set. Even after fine-tuning G_t on unseen data, we observe that G_t produces results without PSNR degradation in the pre-training domain. This finding demonstrates the effectiveness of our proposed method, which utilizes human feedback to mitigate catastrophic forgetting. (**a**) MNIST test set (10k). (**b**) Fashion-MNIST test set (10k). (**c**) Fashion-MNIST train set (60k).

Table 1. Result of fine tuning using human feedback. Each row corresponds to the outcomes under different conditions of the loss function. The first row represents our proposed results. The second row shows results without L_p loss, the third row shows results without L_n loss, and the fourth row shows results using only L_r loss. The fifth row presents results from the model trained on the source domain, and the last row displays the baseline results between noisy and clean images.

	MNIST Test (10k)		**Fashion-MNIST Test (10k)**		**Fashion-MNIST Train (60k)**	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
$\boldsymbol{G_t(z)}$ vs. x	29.36 ± 0.92	0.95 ± 0.01	25.68 ± 3.91	0.84 ± 0.11	25.75 ± 3.86	0.84 ± 0.10
$G_t(z)$ vs. x/wo L_p loss	24.20 ± 0.65	0.66 ± 0.08	25.00 ± 2.44	0.68 ± 0.10	25.07 ± 2.37	0.69 ± 0.10
$G_t(z)$ vs. x/wo L_n loss	29.10 ± 0.98	0.95 ± 0.01	25.30 ± 4.35	0.83 ± 0.12	25.41 ± 4.25	0.83 ± 0.12
$G_t(z)$ vs. x/only L_r loss	20.66 ± 0.69	0.82 ± 0.03	17.97 ± 2.15	0.58 ± 0.13	18.03 ± 2.14	0.58 ± 0.12
$G_s(z)$ vs. x	29.26 ± 1.04	0.94 ± 0.01	24.18 ± 5.57	0.80 ± 0.16	24.27 ± 5.52	0.80 ± 0.16
Baseline source (z vs. x)	14.72 ± 0.06	0.12 ± 0.01	13.23 ± 0.13	0.07 ± 0.13	13.23 ± 0.13	0.07 ± 0.02

Visual evaluation: Figure 4 illustrates the improvement in denoising and restoration, particularly in addressing image collapse. Notably, G_s trained on the '0' digits of MNIST, exhibits instances where the results suffer from image collapse in several images, indicating a lack of adaptation. However, our approach effectively enhances the image quality by leveraging human feedback, as demonstrated by the results obtained with G_t.

3.5. Ablation Study

To validate the effectiveness of each loss term of our method, we conduct comprehensive ablation studies for the loss term.

Effect of L_p term: The Lp term compares the image quality between G_s and G_t and is the loss function between images that are well evaluated by human feedback based on the reward function. We examine the effect of the L_p loss on the quality of the output. Typically, the constant alpha of L_p is fixed at 0.9. To evaluate the effect of excluding the L_p term, we vary the alpha value to 0, resulting in the loss equation becoming $L(G_t) = L_r(G_t) + L_n(G_t)$. Performing the adaptation without an L_p term exhibits low quantitative performance, as demonstrated in the second row of Table 1. Additionally, Figure 6d,e depict the anomaly texture created in the image for reference.

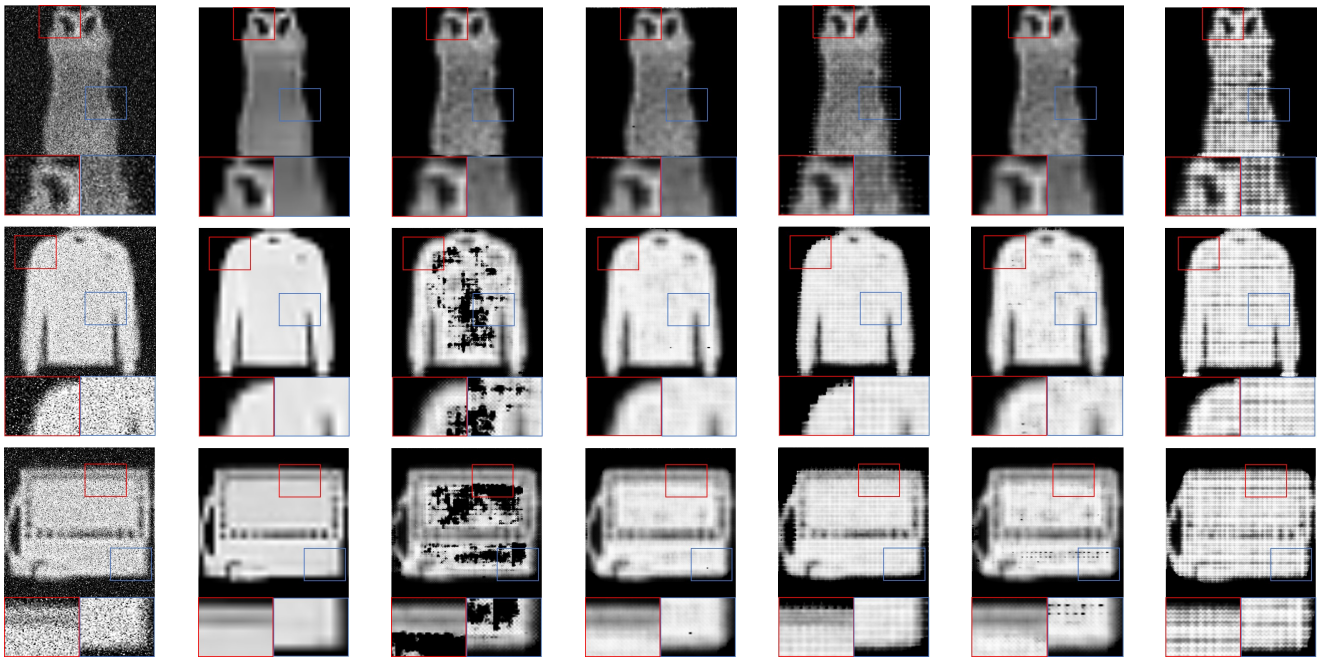

Figure 6. Comparison of image quality with and without the auxiliary loss. (**d**–**g**) Results with different auxiliary loss conditions. Each condition improves the image quality compared to (**c**), but there are noticeable differences in details such as texture and artifacts. (**a**) Input image with noise. (**b**) Ground truth. (**c**) Denoised images by G_s. (**d**) Denoised images by G_t. (**e**) Denoised images by G_t without the L_p term. (**f**) Denoised images by G_t without the L_n term. (**g**) Denoised images by G_t using only the L_r term.

Effect of L_n term: L_n represents the L_1 loss between the $(n-2)$th and (n)th iterations of $G(\hat{G}_t$ and $G_t)$. In terms of quantitative evaluation, it demonstrates comparable performance (Table 1, first and third row). However, in qualitative assessment, it becomes evident that there are limitations in generating the desired image to a satisfactory degree. (See Figure 6d–f). We also examine the effect of L_n loss on the output quality. The role of L_n is to restrict significant parameter changes from the previous model. Given that the function of L_n is to restrict parameter updates between the previous and current models, it becomes apparent that there are limitations in generating the desired image to a satisfactory extent in qualitative evaluation, leading to potential issues such as collapse.

Effect of L_r term: L_r represents the cross-entropy loss used to distinguish between '*Good*' and '*Bad*' cases. In the experiments where L_p and L_n are ablated, the adaptation relies solely on L_r, resulting in parameter updates exclusively driven by human feedback. Consequently, in the absence of pixel-wise losses, such as L_p and L_n, it is evident that image details and shapes are not preserved, as illustrated in Figure 6d–g.

4. Discussion and Conclusions

In this paper, we propose a novel method based on human feedback to address the domain adaptation problem of denoising generative models, particularly focusing on the condition of an unlabeled target domain. Unlike conventional approaches that aim to enhance a generative model's overall performance on the entire test data set, our method leverages human feedback to directly improve the quality of failed images in denoising tasks. While many existing approaches require a large amount of labeled data and may discard failed images, our approach fine-tunes the model based on human feedback, similar to the process used in ChatGPT [22] to select generated sentences of 'Good' quality.

This novel utilization of human feedback represents a promising avenue for enhancing generative models.

A comparison with existing domain adaptation techniques highlights the distinct features of the proposed approach. Conventional domain adaptation methods primarily focus on minimizing the distributional gap between the source and target domains [14–21]. While these approaches effectively address macro-level domain shifts, they often lack the ability to correct fine-grained failure cases in individual images. In contrast, the proposed method introduces a micro-level adaptation strategy by incorporating human feedback to specifically target and improve failure cases within the target domain. This fine-grained approach to domain adaptation has not been extensively explored in prior studies and provides a novel perspective on improving model performance beyond conventional techniques.

To assess the effectiveness of the proposed method, a performance comparison was conducted using PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure). The experimental results demonstrate a 94% improvement in PSNR in the target domain, indicating a substantial enhancement in performance compared to traditional domain adaptation techniques. This quantitative evaluation confirms the effectiveness of leveraging human feedback for fine-tuning generative models and highlights its potential in practical applications.

The originality of this study lies in its novel integration of human feedback into generative AI for domain adaptation. Rather than relying on large labeled datasets, as is common in existing domain adaptation approaches, the proposed method utilizes subjective human evaluations to adapt the model efficiently and effectively. This strategy not only reduces the dependency on extensive labeled datasets but also enhances the adaptability of the model to real-world scenarios where failure cases require specific adjustments. By employing human-guided model fine-tuning, this study introduces an innovative approach to improving the robustness and flexibility of generative AI models.

We proposed a novel human-guided domain adaptation approach for image denoising and demonstrated its effectiveness on MNIST and Fashion-MNIST datasets. Notably, the PSNR improved by 94% on the Fashion-MNIST test set, with an average increase of 1.61 ± 2.78 dB and a maximum of 18.21 dB, while preserving performance on the original MNIST domain. These results highlight the potential of human feedback in enhancing model adaptability and performance across domains.

Domain adaptation poses challenges, particularly regarding the issue of catastrophic forgetting. However, through our proposed adaptation approach, which incorporates selective loss functions and an ablation study based on decisions from a reward model trained with human feedback, we successfully mitigated the challenging issue of catastrophic forgetting. Our results align with related studies, demonstrating the effectiveness of our approach.

Despite its success, the study has limitations, including the possibility of overfitting to specific noise types and challenges in applying the method to more complex datasets. Future research will address these limitations by testing on high-resolution datasets and exploring methods to standardize and optimize human feedback collection.

In the context of real-world applications, the unseen data domain adaptation of deep generative models has always been a crucial research topic. In this paper, we demonstrate the adaptation of a model trained on the source domain to the label-less target domain, guided by human feedback. Through ablation study, we analyzed the loss functions and provided compelling evidence for the direction of domain adaptation research, particularly in the realm of image generation.

Although the proposed method has been successfully applied to simple image datasets, such as MNIST and Fashion-MNIST, its performance may not be guaranteed when extended

to more complex images (e.g., medical imaging or natural photographs). Additionally, if feedback is inconsistent across users, the model's performance may deteriorate. To mitigate these issues, further exploration of feedback integration and standardization methods is required.

To improve the reliability and efficiency of human feedback in future work, we will explore several key strategies. Specifically, we aim to analyze the correlation between feedback reliability and performance improvement to develop an optimized feedback collection strategy that minimizes the amount of required feedback while maximizing its effectiveness. In addition, we will investigate cost-effective methods for large-scale feedback collection, such as utilizing crowdsourcing platforms or user-friendly interfaces, to enhance the practical feasibility of integrating human feedback into real-world applications. These efforts will be complemented by measures to ensure consistency and reliability, including the integration of multiple feedback sources and the establishment of clear evaluation criteria to minimize potential biases. To further reduce dependence on human feedback, we will also explore automated evaluation techniques, such as proxy model learning, which can either supplement or replace human input while preserving model adaptability and performance. These strategies will contribute to improving the scalability and robustness of the proposed approach, facilitating its application in more complex and diverse domains.

Furthermore, we will grapple with two things as follows: 1. Human preference: Our work also collects human feedback data by personal preference, similar to ChatGPT. Thus, the distribution of 'Good' quality can be different. This will be connected directly with the model's performance. 2. Model performance is dependent on pre-training: We assume that the SDM is over a certain level. However, if SDM does not work in an unseen domain, we can not collect human feedback. Human feedback has to be collected in the 'Good' and 'Bad' categories.

Author Contributions: Conceptualization, S.H.K. and H.-C.P.; methodology, H.-C.P.; software, S.H.K.; validation, H.-C.P., D.N., and S.H.K.; formal analysis, S.H.K.; investigation, D.N.; resources, D.N.; data curation, D.N.; writing—original draft preparation, H.-C.P.; writing—review and editing, S.H.K.; visualization, S.H.K.; supervision, S.H.K.; project administration, H.-C.P. and S.H.K.; funding acquisition, H.-C.P. and S.H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00338504) and by the National Institute for Mathematical Sciences (NIMS) funded by the Korean Government under Grant NIMS-B25910000.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
2. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial Feature Learning. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, 24–26 April 2017.
3. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
4. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
5. Tran, L.D.; Nguyen, S.M.; Arai, M. GAN-based noise model for denoising real images. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.
6. Vo, D.M.; Nguyen, D.M.; Le, T.P.; Lee, S.W. HI-GAN: A hierarchical generative adversarial network for blind denoising of real photographs. *Inf. Sci.* **2021**, *570*, 225–240. [CrossRef]

7. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projection networks for super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1664–1673.
8. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
9. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
10. Zhu, J.Y.; Zhang, R.; Pathak, D.; Darrell, T.; Efros, A.A.; Wang, O.; Shechtman, E. Toward multimodal image-to-image translation. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 465–476.
11. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Dualgan: Unsupervised dual learning for image-to-image translation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2849–2857.
12. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 700–708.
13. Tian, C.; Fei, L.; Zheng, W.; Xu, Y.; Zuo, W.; Lin, C.W. Deep learning on image denoising: An overview. *Neural Netw.* **2020**, *131*, 251–275. [CrossRef] [PubMed]
14. Volpi, R.; Morerio, P.; Savarese, S.; Murino, V. Adversarial feature augmentation for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5495–5504.
15. Wang, Y.; Wu, C.; Herranz, L.; Van de Weijer, J.; Gonzalez-Garcia, A.; Raducanu, B. Transferring gans: Generating images from limited data. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 218–234.
16. Kang, G.; Jiang, L.; Yang, Y.; Hauptmann, A.G. Contrastive adaptation network for unsupervised domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4893–4902.
17. Alanov, A.; Titov, V.; Vetrov, D.P. Hyperdomainnet: Universal domain adaptation for generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 29414–29426.
18. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3722–3731.
19. Lin, K.; Li, T.H.; Liu, S.; Li, G. Real photographs denoising with noise domain adaptation and attentive generative adversarial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 15–20 June 2019.
20. Chen, L.; Chen, H.; Wei, Z.; Jin, X.; Tan, X.; Jin, Y.; Chen, E. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7181–7190.
21. Kwon, G.; Ye, J.C. One-shot adaptation of gan in just one clip. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12179–12191. [CrossRef]
22. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
23. Christiano, P.F.; Leike, J.; Brown, T.; Martic, M.; Legg, S.; Amodei, D. Deep Reinforcement Learning from Human Preferences. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Volume 30.
24. Lee, K.; Liu, H.; Ryu, M.; Watkins, O.; Du, Y.; Boutilier, C.; Abbeel, P.; Ghavamzadeh, M.; Gu, S.S. Aligning text-to-image models using human feedback. *arXiv* **2023**, arXiv:2302.12192.
25. Stiennon, N.; Ouyang, L.; Wu, J.; Ziegler, D.; Lowe, R.; Voss, C.; Radford, A.; Amodei, D.; Christiano, P.F. Learning to summarize with human feedback. In Proceedings of the Advances in Neural Information Processing Systems, Online Conference, 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Volume 33; pp. 3008–3021.
26. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
27. Zhang, T.; Cheng, J.; Fu, H.; Gu, Z.; Xiao, Y.; Zhou, K.; Gao, S.; Zheng, R.; Liu, J. Noise adaptation generative adversarial network for medical image analysis. *IEEE Trans. Med. Imaging* **2019**, *39*, 1149–1159. [CrossRef] [PubMed]
28. Park, H.S.; Jeon, K.; Lee, S.H.; Seo, J.K. Unpaired-paired learning for shading correction in cone-beam computed tomography. *IEEE Access* **2022**, *10*, 26140–26148. [CrossRef]
29. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 27730–27744.

30. Zhu, P.; Abdal, R.; Femiani, J.; Wonka, P. Mind the gap: Domain gap control for single shot domain adaptation for generative adversarial networks. *arXiv* **2021**, arXiv:2110.08398.
31. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
32. Xiao, H.; Rasul, K.; Vollgraf, R. Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms. *arXiv* **2017**, arXiv:1708.07747.
33. Fu, B.; Zhao, X.; Song, C.; Li, X.; Wang, X. A salt and pepper noise image denoising method based on the generative classification. *Multimed. Tools Appl.* **2019**, *78*, 12043–12053. [CrossRef]
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.