Conditional Diffusion Process for Inverse Halftoning

Anonymous Author(s)
Affiliation
Address
email

Abstract

Inverse halftoning is a technique used to recover realistic images from ancient prints (*e.g.*, photographs, newspapers, books). The rise of deep learning has led to the gradual incorporation of neural network designs into inverse halftoning methods. Most of existing inverse halftoning methods adopt the U-net architecture, which uses an encoder to encode halftone prints, followed by a decoder for image reconstruction. However, the mainstream U-net architecture has poor generalization ability in practical applications. Specifically, when there is a large gap between the halftone patterns of the training and test data, the reconstructed continuous-tone images have obvious artifacts. This is an important issue in practical applications, since the algorithms for generating halftones are ever-evolving. Even for the same algorithm, different parameter choices will result in different halftone dithering patterns. In this paper, we propose the first generative halftoning method in the literature, which regards the black pixels in halftones as physically moving particles, and makes the randomly distributed particles move under some certain guidance through the reverse diffusion process, so as to obtain the desired halftone patterns. In particular, we propose a Conditional Diffusion model for image Halftoning (CDH), which consists of a halftone dithering process and an inverse halftoning process. By changing the initial state of the diffusion model, our method can generate visually plausible halftones with different dithering patterns under the condition of image gray level and Laplacian prior. To avoid introducing redundant patterns and undesired artifacts, we propose a meta-halftone guided network to incorporate blue noise guidance in the diffusion process. In this way, halftone images subject to more diverse distributions are fed into the inverse halftoning model, which helps the model to learn a more robust mapping from the halftone distribution to continuous-tone distribution, thereby improving the generalization ability to unseen samples. Quantitative and qualitative experimental results demonstrate that the proposed method achieves state-of-the-art results.

1 Introduction

Halftoning refers to the task of simulating the brightness change of a continuous-tone image by changing the size or frequency of halftone dots (such as ink dots). In the last century, halftoning technology has been widely used in the printing industry to store precious image data through old newspapers, books, photographs, *etc*. At the same time, inverse halftoning technology emerged to recover stored continuous-tone image from vintage materials. The goal of the inverse halftoning technique is to minimize the loss of information in the restoration process, so that the restored print has the highest possible visual quality.

Traditional inverse halftoning methods are mainly based on image filtering (*e.g.*, wavelet domain filtering (Xiong et al., 1999), edge-preserving filtering (Kite et al., 2000), SUSAN filtering (Siddiqui and Bouman, 2007)), bilateral filtering (Sun et al., 2014) and statistical learning methods (*e.g.*, least

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

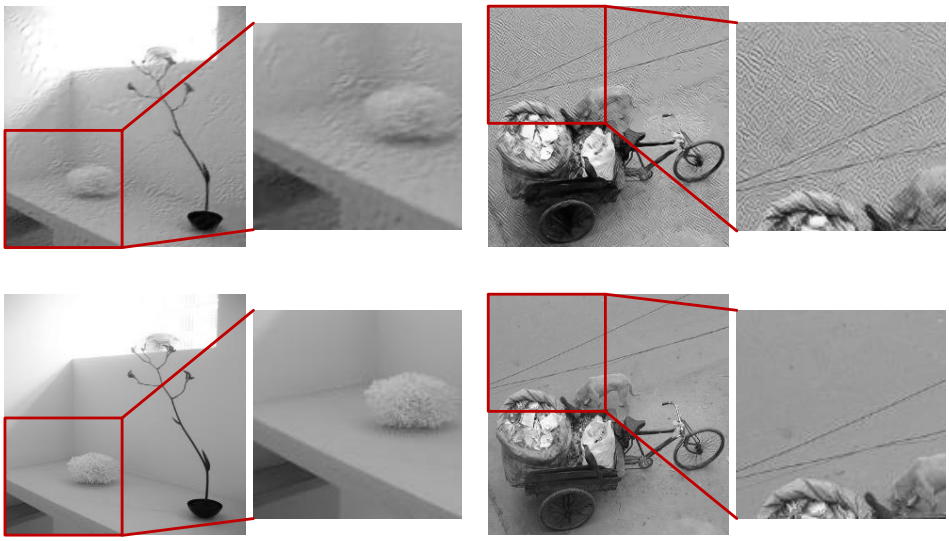

Figure 1: Illustration of the presence of artifacts in restored images by previous inverse halftoning method (Xia and Wong, 2018), where domain shifts exist in the training and test data. The top row are the restored results for some testing images, and the bottom row are the corresponding ground-truth continuous-tone images.

mean square filtering (Chen and Hang, 1997), maximum a posteriori (Stevenson, 1997), look-up table (Mese and Vaidyanathan, 2001; Chung and Wu, 2005), dictionary learning (Zhang et al., 2018)). With the revival of deep learning technology, inverse halftoning methods based on deep neural networks have made great progress and attracted more and more attention (Hou and Qiu, 2017; Xiao et al., 2017; Kim and Park, 2018). The most representative method is the U-net (Ronneberger et al., 2015) based architecture, which has an encoder to learn the hidden encoding of the halftone printing, followed by a decoder to reconstruct the image (Xia and Wong, 2018; Gao et al., 2019).

However, the U-net architecture has poor generalization ability in practical applications. Specifically, when there is a large gap between the halftone patterns of the training and test data, the reconstructed continuous-tone images often have obvious artifacts. This is an important issue in practical applications, since the algorithms used for generating halftones evolve with time. Even for the same algorithm, different parameter choices will result in different halftone dithering patterns. Taking frequency modulation (FM) halftoning as an example, we select 9 classical error diffusion processes with different dither patterns, namely Floyd-Steinberg Dithering, Jarvis-Judice-Ninke Dithering, Stucki Dithering, Atkinson Dithering, Burkes Dithering, Sierra Dithering, and several of their variants (Lau and Arce, 2018). We train the U-net model on the images generated by 5 of the algorithms, and test on the halftone images obtained with the rest 4 algorithms. The experimental results are shown in Figure 1. Artifacts can be clearly observed in the restored halftones.

To address this problem, in this paper, we propose a Conditional Diffusion model for image Halftoning (CDH), as shown in Figure 2, which consists of a halftone dithering process and an inverse halftoning process. We regard the black pixels in halftones as physically moving particles, and make the randomly distributed particles move under some certain guidance through the reverse diffusion process, so as to obtain the desired halftone distribution. Specifically, for the halftone dithering process, we train a conditional diffusion model to generate halftones with different dithering patterns under the condition of image gray level and Laplacian prior. By changing the initial state of the diffusion model, it can simulate different point processes to generate diverse halftone images. To avoid introducing redundant patterns and undesired artifacts during halftone generation, we propose a meta-halftone guided network to incorporate the blue noise guidance into the diffusion process. For inverse halftoning, we train an inverse halftoning diffusion model to learn the mapping function from the halftone distribution to the continuous-tone distribution. In this way, halftones subject to more diverse distributions are input to the inverse halftoning model, which helps the model to learn a more robust mapping to perform inverse halftoning and improve the generalization ability to unseen samples.

Our contributions are summarized as follows:

- This is the first work to propose a generative halftoning method, which regards the black pixels in halftones as physically moving particles, and makes the randomly distributed particles move under some certain guidance through the reverse diffusion process, so as to obtain the desired halftone dithering patterns.
- To avoid introducing redundant patterns and undesired artifacts during halftone generation, we propose a meta-halftone guided network to incorporate the blue noise guidance into the halftone diffusion process.
- To obtain better generalization ability, we use the x_0 state of halftone dithering diffusion as the condition for inverse halftoning diffusion, so that the inverse halftoning diffusion model benefits from a wider range of dithering patterns and learns a more robust mapping.

We conduct experiments on the dataset consisting of 9 halftoning algorithms, and quantitative and qualitative experiments demonstrate that the proposed method achieves state-of-the-art results. We release the code of this paper anonymously to facilitate further research[1].

2 Related Work

Inverse Halftoning. Traditional inverse halftoning methods are mainly based on image filtering (*e.g.*, edge-preserving filtering (Kite et al., 2000), wavelet domain filtering (Xiong et al., 1999), bilateral filtering (Sun et al., 2014), SUSAN filtering (Siddiqui and Bouman, 2007)) and statistical learning methods (*e.g.*, look-up table (Chung and Wu, 2005; Mese and Vaidyanathan, 2001), least mean square filtering (Chen and Hang, 1997), maximum a posteriori (Stevenson, 1997), dictionary learning (Zhang et al., 2018)). For example, Sun *et al.* (Sun et al., 2014) proposed to use an anisotropic Gaussian filter and an edge-preserving filter for inverse halftoning. With the revival of deep learning technology, some researchers have tried to use deep neural networks to accomplish inverse halftoning, *e.g.*, U-net based models (Hou and Qiu, 2017; Xiao et al., 2017; Gao et al., 2019), residual learning based models (Xia and Wong, 2018) and contextual learning based models (Kim and Park, 2018). For example, Xia *et al.* (Xia and Wong, 2018) proposed a progressively residual based U-net that synthesizes the global tone and subtle details to generate inverse halftones. However, these encoder-decoder based architectures suffer from unacceptable artifacts in the generated images when faced with domain generalization problems.

Diffusion Models. Diffusion models are a class of deep generative models developed from non-equilibrium thermodynamics (Sohl-Dickstein et al., 2015). They define a Markovian process for the diffusion steps, incrementally adding random noises to the original data, and then learn the reverse diffusion process to resample the data from noises (Sohl-Dickstein et al., 2015; Ho et al., 2020; Luo and Hu, 2021). Diffusion models are closely related to score-based generative models, which generate samples by Langevin dynamics based on estimated gradients of the data distribution (Song and Ermon, 2019, 2020; Song et al., 2020b, 2021). Some improved techniques are proposed to help the diffusion model converge to a lower negative log-likelihood or to speed up sampling (Nichol and Dhariwal, 2021; Dhariwal and Nichol, 2021; Song et al., 2020a). For example, Song *et al.* (Song et al., 2020a) generalizes denoising diffusion probabilistic models via a class of non-Markovian processes, which can correspond to deterministic generative processes and gives rise to implicit models that produce high quality samples much faster. However, these diffusion models cannot be directly applied in halftone generation, since they do not take into account the blue noise properties of halftone prints, making it difficult to generate visually pleasing dithering patterns.

3 Method

3.1 Halftone Dithering Diffusion Conditioned on Gray-scale and Laplacian Prior

Traditional halftone dithering methods are mainly based on techniques of amplitude modulation (Blatner et al., 1993; Campbell et al., 1966) and frequency modulation (Eschbach, 1997; Floyd, 1976; Meşe and Vaidyanathan, 1999). With the development of deep learning, some modern approaches

[1] https://code-website.wixsite.com/anonymous

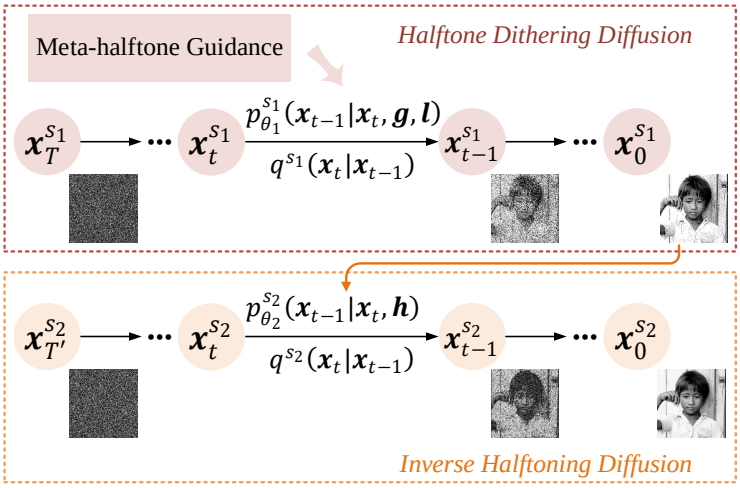

Figure 2: Illustration of the proposed Conditional Diffusion model for image Halftoning (CDH), which consists of a halftone dithering diffusion process and an inverse halftoning diffusion process.

try to generate halftone dithering images using convolutional networks Kim and Park (2018); Xia et al. (2021). However, these methods are limited to specific dithering patterns, and cannot achieve flexible generative halftones. To address this problem, considering that the halftone prints are only composed of black and white pixels, we regard the black pixels in halftones as particles, and make these particles move from the random Gaussian distribution to the halftone distribution under the reverse diffusion process, resulting in a generative halftoning model.

Given the distribution $q^{s_1}(\boldsymbol{x}_0^{s_1})$ of halftone prints, the diffusion process q^{s_1} is a Markovian noising process (Ho et al., 2020) that gradually adds noise to $\boldsymbol{x}_0^{s_1}$ to obtain $\boldsymbol{x}_{1:T}^{s_1}$, where s_1 denotes the halftone dithering diffusion stage. Specifically, at each step t, the diffusion step adds the random Gaussian noise with a β_t-controlled variance:

$$q^{s_1}(\boldsymbol{x}_{1:T}^{s_1}|\boldsymbol{x}_0^{s_1}) = \prod_{t=1}^{T} q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_{t-1}^{s_1}), \tag{1}$$

$$q^{s_1}(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{s_1}) = \mathcal{N}(\boldsymbol{x}_t^{s_1}; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}^{s_1}, \beta_t \mathbf{I}), \tag{2}$$

where $\beta_t \in (0,1)$, $t = 1, ..., T$. With the reparameterization trick (Kingma and Welling, 2013), we can sample $\boldsymbol{x}_t^{s_1}$ from any time step t in a closed form: $\boldsymbol{x}_t^{s_1} = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0^{s_1} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, that is

$$q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_0^{s_1}) = \mathcal{N}(\boldsymbol{x}_t^{s_1}; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0^{s_1}, (1-\bar{\alpha}_t)\mathbf{I}), \tag{3}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$. In this way, we can directly derive $\boldsymbol{x}_t^{s_1}$ by $q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_0^{s_1})$ without repeatedly applying the Markovian process q^{s_1} and calculating $q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_{t-1}^{s_1})$.

In the halftone dithering scenario, we would like to obtain the halftone sample $\boldsymbol{x}_0^{s_1}$ via the reverse diffusion process $q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1})$ through a random Gaussian distribution, *i.e.*, the state $\boldsymbol{x}_T^{s_1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. However, considering that the particles cannot directly form the halftone distributions through the reverse diffusion process, we use the image gray level $\boldsymbol{r}$ and the Laplacian prior $\boldsymbol{l}$ as conditions to guide the reverse diffusion process. Formally, the reverse halftoning diffusion process is defined on the distribution $p_{\theta_1}^{s_1}(\boldsymbol{x}_{0:T}^{s_1}|\boldsymbol{r}, \boldsymbol{l})$, which is a Markov chain with start state $p^{s_1}(\boldsymbol{x}_T^{s_1}) = \mathcal{N}(\boldsymbol{x}_T^{s_1}; \mathbf{0}, \mathbf{I})$:

$$p_{\theta_1}^{s_1}(\boldsymbol{x}_{0:T}^{s_1}|\boldsymbol{r}, \boldsymbol{l}) = p^{s_1}(\boldsymbol{x}_T^{s_1}) \prod_{t=1}^{T} p_{\theta_1}^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1}, \boldsymbol{r}, \boldsymbol{l}), \tag{4}$$

$$p_{\theta_1}^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1}, \boldsymbol{r}, \boldsymbol{l}) = \mathcal{N}(\boldsymbol{x}_{t-1}^{s_1}; \boldsymbol{\mu}_{\theta_1}^{s_1}(\boldsymbol{x}_t^{s_1}, \boldsymbol{r}, \boldsymbol{l}, t), \boldsymbol{\Sigma}_{\theta_1}^{s_1}(\boldsymbol{x}_t^{s_1}, \boldsymbol{r}, \boldsymbol{l}, t)). \tag{5}$$

Different from previous methods like using class labels (Dhariwal and Nichol, 2021) or shape latents (Luo and Hu, 2021) as conditions for the diffusion model, we use the pixel-wise image priors

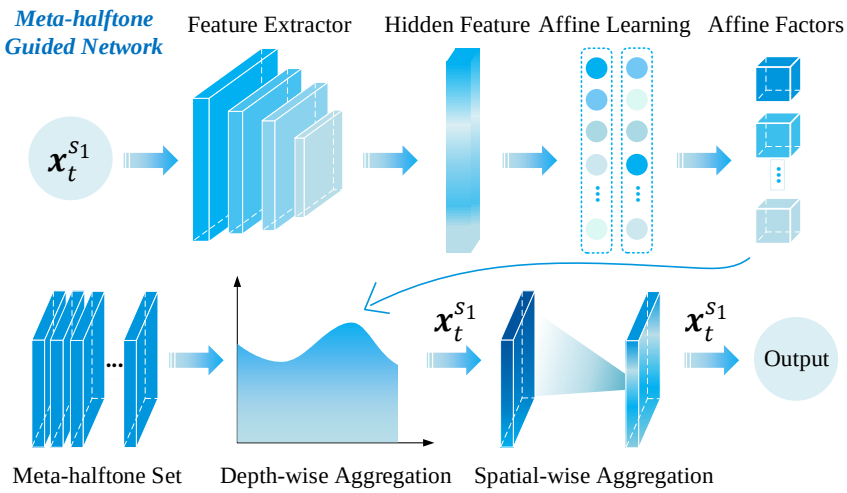

Figure 3: Illustration of the proposed Meta-halftone Guided Network.

as conditions. The conditions are performed in a simple way, *i.e.*, we concatenate the state vector $\boldsymbol{x}_t^{s_1}$ at diffusion step t with $\boldsymbol{r}$ and $\boldsymbol{l}$ along the channel dimension to provide pixel-wise guidance:

$$\boldsymbol{x}_t^{s_1} := \boldsymbol{x}_t^{s_1} \oplus \boldsymbol{r} \oplus \boldsymbol{l}, \tag{6}$$

where $\oplus$ denotes the concatenate operation. Due to the properties of halftones, local dithering patterns are more important than global semantic information for generating high-quality halftones. We also observe in the experiments that excessive introduction of global information can be harmful to dithering results.

We would like to learn the model $p_{\theta_1}^{s_1}$ to approximate the conditional probabilities $q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1})$. With Bayes theorem, it has

$$q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1}, \boldsymbol{x}_0^{s_1}) = q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_{t-1}^{s_1}, \boldsymbol{x}_0^{s_1}) \frac{q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_0^{s_1})}{q^{s_1}(\boldsymbol{x}_t^{s_1}|\boldsymbol{x}_0^{s_1})}, \tag{7}$$

and $q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1}, \boldsymbol{x}_0^{s_1})$ can be represented as a Gaussian distribution:

$$q^{s_1}(\boldsymbol{x}_{t-1}^{s_1}|\boldsymbol{x}_t^{s_1}, \boldsymbol{x}_0^{s_1}) = \mathcal{N}(\boldsymbol{x}_{t-1}^{s_1}; \tilde{\boldsymbol{\mu}}_t^{s_1}(\boldsymbol{x}_t^{s_1}, \boldsymbol{x}_0^{s_1}), \tilde{\beta}_t \mathbf{I}), \tag{8}$$

$$\tilde{\boldsymbol{\mu}}_t^{s_1}(\boldsymbol{x}_t^{s_1}, \boldsymbol{x}_0^{s_1}) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0^{s_1} + \frac{\sqrt{\alpha}_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t^{s_1}, \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t. \tag{9}$$

The simplified objective function (Ho et al., 2020) is used to train the diffusion model:

$$L^{s_1} = \mathbb{E}_{t\sim[1,T], \boldsymbol{x}_0^{s_1}\sim q^{s_1}(\boldsymbol{x}_0^{s_1}), \boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[||\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\theta_1}^{s_1}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0^{s_1} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)||^2\right]. \tag{10}$$

3.2 Meta-halftone Guided Network

Blue noise is critical for generating high-quality halftones. It avoids noticeable low-frequency visual artifacts in the resulting halftones by forcing random pixel dithering (Lau and Arce, 2018). Some traditional methods have been proposed to derive blue noise dither patterns, such as simulated annealing based methods (Sullivan et al., 1991), void-and-cluster techniques (Ulichney, 1993), power spectrum manipulation algorithms (Yao and Parker, 1994), and dither pattern ordering methods (Lau et al., 1999). However, most of these approaches are based on statistical methods or manually designed blue-noise dithering matrices, which are difficult to directly apply to the learning-based neural frameworks. Some researchers have tried to inject blue noise properties into the L1 norm for neural learning objectives (Xia et al., 2021), however, since the calculation of the L1 norm requires known halftone generation results, it cannot be used in generative models. So far, how to incorporate the blue-noise dithering properties into halftone generative models remains an unexplored problem.

To address this issue, we propose a meta-halftone guided network (as shown in Figure 3), which introduces blue noise guidance into halftone dithering diffusion process and avoids dithering results containing artifacts or redundant patterns. Formally, for step t of the halftone dithering diffusion process, we consider the state vector $\boldsymbol{x}_t^{s_1}$ and take it as input to the meta-halftone guided network. We first feed $\boldsymbol{x}_t^{s_1}$ into a feature extraction network $\mathbf{E}$ to obtain the extracted hidden feature $\boldsymbol{f}(\boldsymbol{x}_t^{s_1})$, *i.e.*, $\boldsymbol{f}(\boldsymbol{x}_t^{s_1}) = \mathbf{E}(\boldsymbol{x}_t^{s_1})$. We use pre-trained VGG (Simonyan and Zisserman, 2015) as the network $\mathbf{E}$ in the experiment, other networks such as InceptionNet (Szegedy et al., 2016) or ResNet (He et al., 2016) are also available. Next, we define the meta-halftone set $\mathcal{M}$, which consists of a set of meta-halftone vectors $\boldsymbol{m}_i (1 \leq i \leq k)$:

$$\mathcal{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, ..., \boldsymbol{m}_k], \tag{11}$$

where k represents the number of meta halftones. We obtain $\boldsymbol{m}_i$ by the halftone dithering diffusion states of a set of low-frequency images. Specifically, we construct k images $\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3 ..., \mathbf{I}_k$ with relatively more low-frequency components (*e.g.*, images with constant gray levels), and obtain the halftones $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, ..., \mathbf{H}_k$ accordingly using conventional halftone algorithms (*e.g.*, Floyd-Steinberg Algorithm (Lau and Arce, 2018)). We first train the model for a certain step using the U-net architecture alone, which is common in (Ho et al., 2020; Dhariwal and Nichol, 2021), without adding the meta-halftone guided network. We have

$$p_{\theta_1}^{s_1}(\boldsymbol{x}_{t-1}^{h_i} | \boldsymbol{x}_t^{h_i}, \boldsymbol{r}^{h_i}, \boldsymbol{l}^{h_i}) = \mathcal{N}(\boldsymbol{x}_{t-1}^{h_i}; \boldsymbol{\mu}_{\theta_1}^{s_1}(\boldsymbol{x}_t^{h_i}, \boldsymbol{r}^{h_i}, \boldsymbol{l}^{h_i}, t), \boldsymbol{\Sigma}_{\theta_1}^{s_1}(\boldsymbol{x}_t^{h_i}, \boldsymbol{r}^{h_i}, \boldsymbol{l}^{h_i}, t)), \tag{12}$$

where the superscript h_i denotes the constructed halftone data from $\mathbf{H}_i \in \{\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_k\}$, and $\boldsymbol{r}^{h_i}$ and $\boldsymbol{l}^{h_i}$ can be derived from $\mathbf{I}_i \in \{\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_k\}$. The superscript s_1 of $x_{t-1}^{h_i}, x_t^{h_i}$ are omitted here for brevity. We can get $\boldsymbol{\epsilon}_{\theta_1}^{s_1}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0^{h_i} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t)$ by parameterization of $\mu_{\theta_1}^{s_1}$:

$$\mu_{\theta_1}^{s_1}(\boldsymbol{x}_t^{h_i}, t) = \frac{1}{\sqrt{\alpha_t}}\big(\boldsymbol{x}_t^{h_i} - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_{\theta_1}^{s_1}(\boldsymbol{x}_t^{h_i}, t)\big). \tag{13}$$

We use $\boldsymbol{m}_i$ in Equation 11 to represent the model prediction and construct the meta-halftone set $\mathcal{M}$ accordingly:

$$\boldsymbol{m}_i = \boldsymbol{\epsilon}_{\theta_1}^{s_1}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0^{h_i} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t), 1 \leq i \leq k. \tag{14}$$

Next, the model learns the affine relationship between the hidden feature $\boldsymbol{f}(\boldsymbol{x}_t^{s_1})$ and the meta-halftone set $\mathcal{M}$ through the affine learning layer, and then obtains the affine factors $\boldsymbol{g}(\boldsymbol{x}_{t\leftarrow i}^{s_1})$:

$$\boldsymbol{g}(\boldsymbol{x}_{t\leftarrow i}^{s_1}) = \boldsymbol{w}_{g\leftarrow i} \cdot \boldsymbol{f}(\boldsymbol{x}_t^{s_1}) + \boldsymbol{b}_{g\leftarrow i}, \tag{15}$$

where $\boldsymbol{w}_{g\leftarrow i}$ and $\boldsymbol{b}_{g\leftarrow i}$ represent learnable weights and biases, respectively. With the calculated $\boldsymbol{g}(\boldsymbol{x}_{t\leftarrow i}^{s_1})$, we perform the depth-wise aggregation of the meta-halftone set $\mathcal{M}$:

$$\widetilde{\boldsymbol{m}} = \sum_{i=1}^{k} \boldsymbol{m}_i \cdot \boldsymbol{g}(\boldsymbol{x}_{t\leftarrow i}^{s_1}). \tag{16}$$

Besides depth information, spatial information is also important for the guidance of meta-halftoning. We perform the spatial-wise aggregation of $\widetilde{\boldsymbol{m}}$ and $\boldsymbol{x}_t^{s_1}$, and denote the result as $\widetilde{\boldsymbol{m}}'$:

$$\widetilde{\boldsymbol{m}}' = \sum \boldsymbol{d} * \left(\widetilde{\boldsymbol{m}} \oplus \boldsymbol{x}_t^{s_1}\right), \tag{17}$$

where $*$ denotes the spatial convolution and $\boldsymbol{d}$ represents the convolution kernel. The output of the meta-halftone guided network $\boldsymbol{o}$ is determined by the meta-halftone guidance $\widetilde{\boldsymbol{m}}'$ and $\boldsymbol{x}_t^{s_1}$:

$$\boldsymbol{o} = \widetilde{\boldsymbol{m}}'\boldsymbol{x}_t^{s_1} + \boldsymbol{x}_t^{s_1}, \tag{18}$$

and $\boldsymbol{o}$ is subsequently fed into the traditional U-net architecture as in previous work (Dhariwal and Nichol, 2021; Nichol and Dhariwal, 2021). It is worth noting that the proposed meta-halftone guided network has the advantage of being very lightweight, containing only $81K$ training parameters.

3.3 Inverse Halftoning Diffusion Conditioned on $\boldsymbol{x}_0^{s_1}$

The inverse halftoning diffusion process is substantially similar to the halftone dithering diffusion in Section 3.1, except that the diffusion conditions are different. The goal of the inverse halftoning process is to learn a mapping from the halftone distribution to the continuous-tone distribution. For the inverse halftoning diffusion process, we deal with the process $q^{s_2}(\boldsymbol{x}_{1:T'}^{s_2}|\boldsymbol{x}_0^{s_2})$ that gradually adds noise to the data $q^{s_2}(\boldsymbol{x}_0^{s_2})$ from a continuous-tone distribution, and the process $q^{s_2}(\boldsymbol{x}_{t-1}^{s_2}|\boldsymbol{x}_t^{s_2}, \boldsymbol{x}_0^{s_2})$ of gradually denoising from Gaussian noise to obtain desired samples.

Unlike Section 3.1, we use a model $p_{\theta_2}^{s_2}(\boldsymbol{x}_{t-1}^{s_2}|\boldsymbol{x}_t^{s_2}, \boldsymbol{h})$ to estimate $q^{s_2}(\boldsymbol{x}_{t-1}^{s_2}|\boldsymbol{x}_t^{s_2})$, which is conditioned on the halftone distribution $\boldsymbol{h}$ (*i.e.*, sampling $\boldsymbol{x}_0^{s_1}$ state from halftone dithering diffusion model):

$$p_{\theta_2}^{s_2}(\boldsymbol{x}_{0:T'}^{s_2}|\boldsymbol{h}) = p^{s_2}(\boldsymbol{x}_T^{s_2}) \prod_{t=1}^{T'} p_{\theta_2}^{s_2}(\boldsymbol{x}_{t-1}^{s_2}|\boldsymbol{x}_t^{s_2}, \boldsymbol{h}), \tag{19}$$

$$p_{\theta_2}^{s_2}(\boldsymbol{x}_{t-1}^{s_2}|\boldsymbol{x}_t^{s_2}, \boldsymbol{h}) = \mathcal{N}(\boldsymbol{x}_{t-1}^{s_2}; \boldsymbol{\mu}_{\theta_2}^{s_2}(\boldsymbol{x}_t^{s_2}, \boldsymbol{h}, t), \boldsymbol{\Sigma}_{\theta_2}^{s_2}(\boldsymbol{x}_t^{s_2}, \boldsymbol{h}, t)). \tag{20}$$

Note that the sampling speed of the halftone dithering diffusion model is relatively fast, usually only $\sim$10s per sample (see Section 4.1 for details). Pixel-wise guidance is adopted to the diffusion state $\boldsymbol{x}_t^{s_2}$ and halftone conditions $\boldsymbol{h}$:

$$\boldsymbol{x}_t^{s_2} := \boldsymbol{x}_t^{s_2} \oplus \boldsymbol{h}. \tag{21}$$

The rest parts (such as the training objective) are similar to the halftone dithering diffusion process, and we omit them here to save space.

4 Experiments

4.1 Experimental Setup

Datasets. In order to better evaluate the generalization ability of different methods, we construct a dataset with relatively strong domain shifts. The domain shifts mainly arise from two aspects: the halftone algorithm and the image semantic. We construct the training set and validation set based on the UTKFace dataset (Zhang et al., 2017) and the test set based on the VOC2012 dataset (Everingham et al., 2010). We collect 9 different halftone dithering algorithms, namely Floyd-Steinberg Dithering, Jarvis-Judice-Ninke Dithering, Stucki Dithering, Atkinson Dithering, Burkes Dithering, Sierra Dithering, and several of their variants (Lau and Arce, 2018). We use 5 of them in the training set and validation set, and the remaining 4 algorithms are used as the test set. There are $7,857$ images in the training set, 400 images in the validation set, and 400 images in the test set.

Baselines. We choose the baselines that are commonly used in inverse halftoning tasks, which include the U-net based methods, generative adversarial network based methods, and super-resolution based methods: **PRL** (Xia and Wong, 2018) proposes an inverse halftoning network with progressively residual learning, which synthesizes the global tone and subtle details from halftone images in a progressive manner. **ESRGAN** (Wang et al., 2018) is used as baselines in the previous inverse halftoning task, which introduces the residual-in-residual dense block to generate realistic textures and avoid unpleasant artifacts. **GLEAN** (Chan et al., 2021) is applied to the super-resolution task, using pre-trained generative adversarial networks to solve ill-posed problems in image restoration. We first smooth the halftone images with Gaussian kernels of different sizes, and then use GLEAN for image restoration. **RealESRGAN** (Wang et al., 2021) is extended to a practical restoration application, which introduces a high-order degradation modeling process to better simulate complex real-world degradations and employs a U-net discriminator with spectral normalization to increase discriminator capability.

Evaluation Metrics. Following previous methods (Xia and Wong, 2018), we use two metrics to evaluate different methods: Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). PSNR expresses the ratio between the power of the peak signal and the power of noise, and SSIM evaluates the image quality in terms of luminance, contrast and structure.

Implementation Details. The image size for halftone dithering and inverse halftoning is 256×256. The number of channels for the diffusion model is 128, and we use 3 residual blocks. The linear noise schedule is adopted throughout diffusion. For the halftone dithering diffusion model, the learning rate is set to 0.0001, and k is set to 20. We use 800 diffusion steps in the training phase. During the

Table 1: Model performance comparison of the proposed method CDH and baseline methods in terms of PSNR and SSIM.

Method	PSNR	SSIM
PRL (Xia and Wong, 2018)	22.82	0.698
ESRGAN (Wang et al., 2018)	20.20	0.428
$\text{ESRGAN}_{\text{lkernel}}$ (Wang et al., 2018)	22.47	0.645
GLEAN (Chan et al., 2021)	20.11	0.491
$\text{GLEAN}_{\text{lkernel}}$ (Chan et al., 2021)	21.74	0.607
RealESRGAN (Wang et al., 2021)	21.04	0.626
$\text{RealESRGAN}_{\text{lkernel}}$ (Wang et al., 2021)	20.94	0.619
CDH (Ours)	$\mathbf{24.24}_{\pm 0.4}$	$\mathbf{0.727}_{\pm 0.01}$

w/o Meta-halftone Guided Network | w/ Meta-halftone Guided Network | Groundtruth Dithering | w/o Meta-halftone Guided Network | w/ Meta-halftone Guided Network | Groundtruth Dithering

Figure 4: Experimental results of the meta-halftone guided network. We show the generated halftone dithering results w/ and w/o the meta-halftone guided network, as well as the halftone dithering groundtruth.

sampling phase, we use strided sampling schedule (Nichol and Dhariwal, 2021) to reduce the reverse diffusion process to 80 steps. For the inverse halftoning diffusion model, we set the learning rate to 0.0001 and use 4000 steps for the diffusion process. We run the experiments on A40 GPU clusters. The time required to train the halftone dithering diffusion model is about 4 hours, and the time to train the inverse halftoning diffusion model is about 10 hours. AdamW (Loshchilov and Hutter, 2018) is used to optimize the model.

4.2 Experimental Results

We evaluate the performance of the proposed method CDH and baseline methods in terms of PSNR and SSIM, and the experimental results are shown in Table 1. We evaluate the effect of different Gaussian kernel sizes on the baseline performance, where $\text{ESRGAN}_{\text{lkernel}}$, $\text{GLEAN}_{\text{lkernel}}$, $\text{RealESRGAN}_{\text{lkernel}}$ denote the use of Gaussian kernels of size 7, and ESRGAN, GLEAN, RealESRGAN represent the use of Gaussian kernels of size 3, respectively. The experimental results show that our method achieves the best performance on both metrics. Previous methods such as PRL (Xia and Wong, 2018), due to the lack of considering the generalization ability of the model on different halftoning algorithms, show sub-optimal performance on test data that has severe domain shifts. Different Gaussian kernel sizes affect the performance of image restoration baselines, such as (Wang et al., 2018; Chan et al., 2021; Wang et al., 2021). A larger Gaussian kernel erases the dot dithering pattern in the halftone to a greater extent, resulting in restored images with less halftone artifacts, but also loss of high frequency information. Our method performs better than the image restoration baselines with different Gaussian kernel sizes.

To verify the effectiveness of the proposed meta-halftone guided network, we conduct ablation studies to observe the effect on the generated halftone dithering qualities with and without the meta-halftone guided network. Experimental results are demonstrated in Figure 4. From the results shown in Figure 4, it can be observed that when the meta-halftone guided network is removed, the

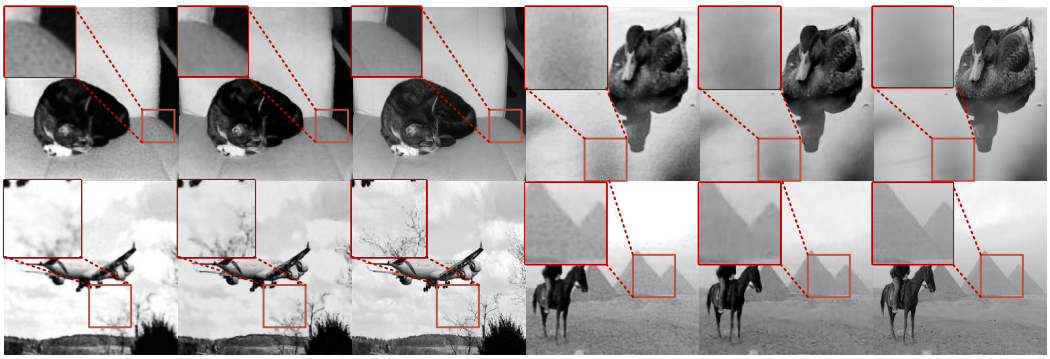

Figure 5: Illustration of the qualitative experimental results. We show the inverse halftoning results generated by the baseline and our method, as well as the groundtruth.

generated halftones have more obvious dithering artifacts, resulting in poorer visual quality. With the meta-halftone guided network, the visual quality of the generated halftones is improved and more similar to the groundtruth. The experimental results demonstrate the effectiveness of the proposed meta-halftone guided network.

4.3 Qualitative Results

To further verify the effectiveness of the proposed method, we also conduct qualitative experiments to compare the performance of our methods and baselines, which are illustrated in Figure 5. From the qualitative experimental results, it can be obversed that the continuous-tone images generated by the baseline method still have some redundant artifacts, which are dot dithering patterns that are not removed during the inverse halftoning process. In contrast, our model removes the artifacts well without introducing redundant patterns, validating the effectiveness of the proposed method.

5 Conclusion

In this work, we propose a generative halftoning method, which regards the black pixels in halftones as physically moving particles, and makes the randomly distributed particles move under some certain guidance through the reverse diffusion process, so as to obtain the desired halftone dithering patterns. A meta-halftone guided network is introduced to avoid redundant patterns and undesired artifacts in generated halftones. By adopting more diverse halftones from halftone dithering diffusion, we further improve the generalization ability of inverse halftoning model. We conduct experiments on the dataset consisting of 9 halftoning algorithms, and quantitative and qualitative experiments demonstrate the effectiveness of the proposed method.

Broader Impact

A potential negative side effect of this work is that the sample generation model may be used to generate fake images in the halftoning or inverse halftoning process for some certain purpose. In addition, most of the images we used in the dataset are collected from the Internet and may contain some bias. The biases are preserved in the learning of the model and may be reflected in the generated image samples.

Limitation Analysis

A limitation of our model is that the diffusion process is relatively slow and thus requires more inference time than traditional inverse halftoning models, which is limited when the model is applied to mobile devices. One way to address this limitation is to design parallel inverse halftoning methods, such as restoring low-frequency and high-frequency components of the image in parallel, thereby increasing the sampling speed.

References

David Blatner, Stephen F Roth, et al. *Real world scanning and halftones*. Peachpit Press, 1993.

Fergus W Campbell, Janus J Kulikowski, and J Levinson. The effect of orientation on the visual resolution of gratings. *The Journal of physiology*, 187(2):427–436, 1966.

Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14245–14254, 2021.

Li-Ming Chen and Hsueh-Ming Hang. An adaptive inverse halftoning algorithm. *IEEE Transactions on Image Processing*, 6(8):1202–1209, 1997.

Kuo-Liang Chung and Shih-Tung Wu. Inverse halftoning algorithm using edge-based lookup table approach. *IEEE Transactions on Image Processing*, 14(10):1583–1589, 2005.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 2021.

Reiner Eschbach. Error diffusion algorithm with homogenous response in highlight and shadow areas. *Journal of Electronic Imaging*, 6(3):348–356, 1997.

Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

Robert W Floyd. An adaptive algorithm for spatial gray-scale. In *Proc. Soc. Inf. Disp.*, volume 17, pages 75–77, 1976.

Qifan Gao, Xiao Shu, and Xiaolin Wu. Deep restoration of vintage photographs from scanned halftone prints. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4120–4129, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Xianxu Hou and Guoping Qiu. Image companding and inverse halftoning using deep convolutional neural networks. *arXiv preprint arXiv:1707.00116*, 2017.

Tae-Hoon Kim and Sang Il Park. Deep context-aware descreening and rescreening of halftone images. *ACM Transactions on Graphics (TOG)*, 37(4):1–12, 2018.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Thomas D Kite, Niranjan Damera-Venkata, Brian L Evans, and Alan C Bovik. A fast, high-quality inverse halftoning algorithm for error diffused halftones. *IEEE Transactions on Image Processing*, 9(9):1583–1592, 2000.

Daniel L Lau and Gonzalo R Arce. *Modern digital halftoning*. CRC Press, 2018.

Daniel L Lau, Gonzalo R Arce, and Neal C Gallagher. Digital halftoning by means of green-noise masks. *JOSA A*, 16(7):1575–1586, 1999.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018.

Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021.

Murat Mese and Palghat P Vaidyanathan. Look-up table (lut) method for inverse halftoning. *IEEE Transactions on Image Processing*, 10(10):1566–1578, 2001.

Murat Meşe and PP Vaidyanathan. Improved dot diffusion for image halftoning. In *NIP & Digital Fabrication Conference*, volume 1999, pages 350–353. Society for Imaging Science and Technology, 1999.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Hasib Siddiqui and Charles A Bouman. Training-based descreening. *IEEE Transactions on Image Processing*, 16(3):789–802, 2007.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. 2015.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020b.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34, 2021.

Robert L Stevenson. Inverse halftoning via map estimation. *IEEE Transactions on Image Processing*, 6(4): 574–583, 1997.

James R Sullivan, Lawrence A Ray, and Rodney Miller. Design of minimum visual modulation halftone patterns. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(1):33–38, 1991.

Bin Sun, Shutao Li, and Jun Sun. Scanned image descreening with image redundancy and adaptive filtering. *IEEE Transactions on Image Processing*, 23(8):3698–3710, 2014.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

Robert A Ulichney. Void-and-cluster method for dither array generation. In *Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 332–343. International Society for Optics and Photonics, 1993.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*, 2021.

Menghan Xia and Tien-Tsin Wong. Deep inverse halftoning via progressively residual learning. In *Asian Conference on Computer Vision*, pages 523–539. Springer, 2018.

Menghan Xia, Wenbo Hu, Xueting Liu, and Tien-Tsin Wong. Deep halftoning with reversible binary pattern. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14000–14009, 2021.

Yi Xiao, Chao Pan, Xianyi Zhu, Hai Jiang, and Yan Zheng. Deep neural inverse halftoning. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 213–218. IEEE, 2017.

Zixiang Xiong, Michael T Orchard, and Kannan Ramchandran. Inverse halftoning using wavelets. *IEEE Transactions on Image Processing*, 8(10):1479–1483, 1999.

Meng Yao and Kevin J Parker. Modified approach to the construction of a blue noise mask. *Journal of Electronic Imaging*, 3(1):92–97, 1994.

Yan Zhang, Erhu Zhang, Wanjun Chen, Yajun Chen, and Jinghong Duan. Sparsity-based inverse halftoning via semi-coupled multi-dictionary learning and structural clustering. *Engineering Applications of Artificial Intelligence*, 72:43–53, 2018.

Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818, 2017.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] The main claims accurately reflect the paper's contributions and scope.
   (b) Did you describe the limitations of your work? [Yes] See Section 5.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 5.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See Section 1: `https://code-website.wixsite.com/anonymous`.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4.1.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See Section 4.2 and Table 1.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 4.1.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We cite the creators when using existing assets.
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]