Ink Dot-Oriented Differentiable Optimization for Neural Image Halftoning

Hao Jiang
Peking University
jianghao@stu.pku.edu.cn

Bingfeng Zhou
Peking University
zbf@pku.edu.cn

Yadong Mu*
Peking University
myd@pku.edu.cn

Figure 1. Illustration of halftones generated by our proposed ink dot-oriented differentiable neural halftoning method and several previous methods [12, 26, 35]. Images are from the DIV2K dataset [2].

Abstract

Halftoning is a time-honored printing technique that simulates continuous tones using ink dots (halftone dots). The resurgence of deep learning has catalyzed the emergence of innovative technologies in the printing industry, fostering the advancement of data-driven halftoning methods. Nevertheless, current deep learning-based approaches produce halftones through image-to-image black box transformations, lacking direct control over the movement of individual halftone dots. In this paper, we propose an innovative halftoning method termed "neural dot-controllable halftoning". This method allows dot-level image dithering by providing direct control over the motion of each ink dot. We conceptualize halftoning as the process of sprinkling dots on a canvas. Initially, a specific quantity of dots are randomly dispersed on the canvas and subsequently adjusted based on the surrounding grayscale and gradient. To establish differentiable transformations between discrete ink dot positions and halftone matrices, we devise a lightweight dot encoding network to spread dense gradients to sparse dots. Dot control offers several advantages to our approach, including the capability to regulate the quantity of halftone dots and enhance specific areas with artifacts in the generated halftones by adjusting the placement of the dots. Our proposed method exhibits superior performance than previous approaches in extensive quantitative and qualitative experiments.

1. Introduction

In the late 1800s, pioneers such as William Henry Fox Talbot, Georg Meisenbach, Frederic Ives, and Max Levy [32] made significant strides in halftone printing techniques, enabling photographs to be printed on paper using ink. Halftoning is a technique of replicating the color and shading of continuous tone images by employing ink dots (halftone dots) of different sizes and frequencies. It remains a crucial and enduring component in the modern printing

*Corresponding Author.

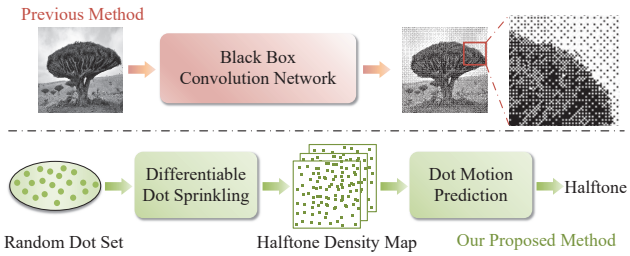


Figure 2. Prior deep networks-based halftoning methods (*top*) map an image to a halftone-like continuous-tone image before applying a binarization operation. In contrast, our proposed method (*bottom*) differs from previous halftoning techniques in that it achieves dot-level halftones through direct control over the movement of each dot.

industry. A successful halftone can closely reproduce the original image with high quality, leveraging the low-pass filtering characteristics of the human visual system to attain near-optimal visual effects.

Halftone technology can be broadly divided into two categories: traditional or deep learning-based methods. The former can be further cast into three groups based on implementations of dithering process: amplitude modulation (AM) [3, 5, 11], frequency modulation (FM) [6, 20, 22, 34], and hybrid modulation [14, 39]. Recently, the renaissance of deep learning as a black-box approximation for high-dimensional functions has catalyzed the emergence of several data-driven methods in the field of halftoning [15, 16, 18, 23, 35], demonstrating superior performance.

In this paper, we argue that current deep learning methods for halftoning, face some weaknesses that impede their deployment in the printing industry. Notably, these methods treat halftones as binary images and heavily rely on a black box encoder-decoder to convert between the original continuous-tone image and the halftone-induced binary image. However, they operate as image-level transforms rather than directly manipulating the halftone dots, leading to a loss of direct control over the desired properties of the resultant halftones. As observed in Figure 1, purely image-level transformations [12, 26, 35] tend to produce undesired artifacts, and they struggle to directly manipulate local dot arrangements to eliminate these artifacts. In addition, traditional methods do not allow for control over the number of halftone dots. In practical applications, opting for halftone techniques that utilize fewer ink dots can be advantageous for consumable savings. For prints with low detail requirements, like halftone cartoon stickers, a modest quantity of dots suffice to faithfully convey the information present in the continuous tone image.

To address the above issues, we propose ink dot-oriented neural image halftoning for the first time in the literature, which directly controls the motion of each ink dot, enabling dot-level differentiable optimization. Nonetheless, a pivotal challenge in achieving this lies in the discrete nature of ink dot positions, while the objective functions piloting the halftone generation are continuous. Directly propagating the gradient flow from halftone pixels to ink dot positions is unfeasible, rendering gradient back-propagation a challenging endeavor. A key contribution of this paper is the presentation of a novel approach to integrate discrete ink dots into continuous halftone optimization in a fully differentiable manner.

Our approach considers halftoning as the process of optimizing dot placements on a blank canvas using halftone quality-related guidance. We commence by randomly spreading a flexible number of dots on the canvas and then progressively move them. To facilitate smooth gradient backpropagation from the objective function to each discrete ink dot, we introduce a dot encoding module that maps dot coordinates to the product of sparse vectors, creating a structure conducive to differentiable optimization. We also develop a motion-estimating module to predict the motion of each ink dot. Our method offers direct control over the quantity of ink dots in the generated halftones, a feat challenging to attain in prior work, owing to the flexibility and scalability of the proposed dot manipulations.

To sum up, this paper introduces ink dot-oriented neural image halftoning, the first approach of its kind to achieve dot-level image dithering by controlling the movement of ink dots. Our proposed model showcases superior performance over previous methods in extensive qualitative and quantitative experiments. Furthermore, the method can be easily extended to arbitrary scale and color images.

2. Related Work

2.1. Traditional Halftoning

Traditional halftone methods can be categorized into amplitude modulation (AM) [3, 5, 11], frequency modulation (FM) [7, 8, 25, 38], and hybrid approaches [14, 39]. AM methods use dot clusters of varying sizes and frequencies to represent different shades of gray and reproduce the original continuous tone image. FM is represented by the error diffusion algorithm, which distributes the quantization error generated by pixel thresholding to adjacent pixels, achieving superior results. For example, Ostromoukhov [26] proposed an improved error-diffusion algorithm based on the deliberately restricted choice of the distribution coefficients. Unlike these traditional methods, which rely on some prior or handcrafted designs, our approach is data-driven, enabling halftone optimization with greater fitting ability and model capacity.

2.2. Deep Learning Based Halftoning

With the emergence of deep learning techniques, researchers have explored the use of neural networks to gen-

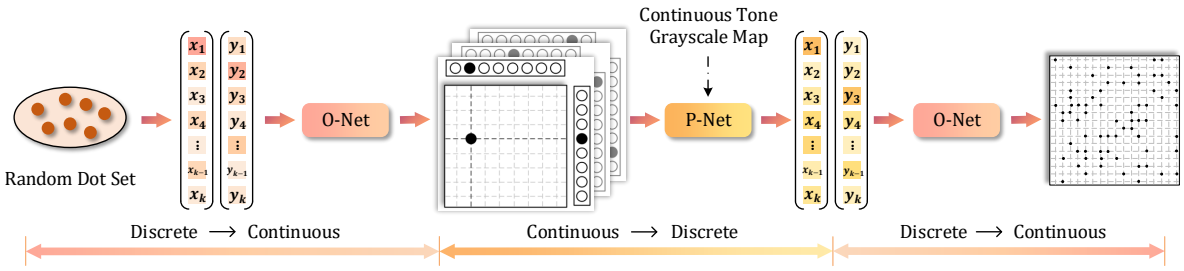


Figure 3. Schematic illustration of the proposed method. The proposed method takes a set of random dots as input. O-Net differentiably transforms discrete ink dot positions into a halftone matrix, effectively spreading dense gradients over sparse dots. P-Net predicts the motion of each halftone dot.

erate halftones [16, 17, 23, 35]. For example, Xia et al. [35] utilized stacked convolutional networks to learn reversible halftones capable of concealing color information. However, most of prior methods treat the entire image as a whole and cannot achieve fine control of halftone dots. In contrast, our approach facilitates dot-level controllable halftones, offering greater interpretability and maneuverability.

3. Method

3.1. O-Net: Spreading Dense Gradients onto Sparse Ink Dots

Halftone images consist solely of black and white pixels, with the arrangement of black pixels determining the brightness and visual structure of each area in the image. A straightforward approach to generate halftone images is to first learn a mapping from a continuous-tone image to an almost binary image, which is then rounded to produce binary halftone dots, as demonstrated in prior work [15–18, 23, 35]. Nevertheless, this approach is subject to several limitations. Primarily, it lacks direct control over individual halftone dots. In the event of defects in some generated halftone areas, the entire halftone have to be regenerated, posing escalating challenges, particularly for larger images. Additionally, previous image-level halftoning methods lack the capability to regulate the quantity of ink dots in halftones. In scenarios where printing equipment features limited or costly consumables, this lack of flexibility hinders the generation of halftones with varying consumable levels.

In Figure 3, we present a novel approach for dot-centric halftoning. The process begins by initializing a set of dots at random. Denote the dot set as $\mathcal{Q} = \{\boldsymbol{q}_0, \boldsymbol{q}_1, ..., \boldsymbol{q}_{k-1}\}$, where k represents the number of dots. Each dot $\boldsymbol{q}_i$ is represented by its two-dimensional coordinates within the image, *i.e.*, $\boldsymbol{q}_i = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, where $i \in [0, k-1], \boldsymbol{x}_i \in [0, L_1 - 1], \boldsymbol{y}_i \in [0, L_2 - 1]$, with L_1 and L_2 being the height and width of the image, respectively. The coordinates of these dots can be used to index their corresponding positions on a "white canvas" $\mathbf{H}$, which serves as a layout for the ink dots. $\mathbf{H}$ consists solely of elements that are either 0 or 1 (after normalization). A value of 1 in $\mathbf{H}$ indicates the presence of an ink dot at the corresponding position, while an element 0 indicates the absence of a dot. To pin the dots in set $\mathcal{Q}$ onto the white canvas $\mathbf{H}$, a straightforward approach is to directly fill these halftone dots into the matrix $\mathbf{H}$:

$$\mathbf{H}\big[\lfloor \boldsymbol{x}_i \rfloor, \lfloor \boldsymbol{y}_i \rfloor\big] = 1, \quad \forall (\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{Q}. \tag{1}$$

Performing such an operation is challenging to optimize in an end-to-end neural network due to the discreteness of ink dot coordinates $\boldsymbol{q}_i$, and the continuity of $\mathbf{H}$. Directly passing the gradient flow between matrix coordinates $\boldsymbol{x}_i, \boldsymbol{y}_i$ and $\mathbf{H}$ is not feasible, and backpropagation is difficult to conduct. While reinforcement learning ideas [10] suggest that decision-making agents can help networks execute decisions under reward functions for discrete action spaces, these approaches are relatively complicated. Instead, we aim to design a simple, differentiable transformation between $\mathcal{Q}$ and $\mathbf{H}$ to enable end-to-end training of the network.

Our key observation is that the halftone $\mathbf{H}$ can be generated through the Hadamard product of multiple sparse binary matrices, and these sparse binary matrices can be expressed as the product of one-hot encodings of dot coordinates. For a halftone $\mathbf{H}$ with k dots, we decompose it into the Hadamard product of k independent matrices $\mathbf{H}_0, \mathbf{H}_1, ..., \mathbf{H}_{k-1}$:

$$\mathbf{H} = \mathbf{I} - (\mathbf{I} - \mathbf{H}_0) \odot (\mathbf{I} - \mathbf{H}_1) \odot ... \odot (\mathbf{I} - \mathbf{H}_{k-1}), \tag{2}$$

where $\odot$ is the Hadamard product, and $\mathbf{I}$ stands for an all-one matrix. $\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{k-1}$ have the property that each matrix contains only one element equal to 1, while the rest are 0. Intuitively, $\mathbf{H}_0, \mathbf{H}_1, ..., \mathbf{H}_{k-1}$ represent halftones formed by printing dots $\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k-1}$, respectively. For a dot $\boldsymbol{q}_i = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, we can get its one-hot encoding from $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$:

$$\begin{aligned} \boldsymbol{o}(\boldsymbol{x}_i) &= \mathrm{OneHotEncoding}(\boldsymbol{x}_i), \\ \boldsymbol{o}(\boldsymbol{y}_i) &= \mathrm{OneHotEncoding}(\boldsymbol{y}_i). \end{aligned} \tag{3}$$

Assuming $\boldsymbol{o}(\boldsymbol{x}_i)$ and $\boldsymbol{o}(\boldsymbol{y}_i)$ have been obtained, $\mathbf{H}_i$ can be calculated according to $\boldsymbol{o}(\boldsymbol{x}_i)$ and $\boldsymbol{o}(\boldsymbol{y}_i)$ via $\mathbf{H}_i = \boldsymbol{o}(\boldsymbol{x}_i)^\top \cdot \boldsymbol{o}(\boldsymbol{y}_i)$.

Substituting the calculated $\mathbf{H}_i$ into Equation 2, the halftone $\mathbf{H}$ can be obtained. The only non-differentiable step occurs in Equation 3 (*i.e.*, OneHotEncoding). The non-differentiable problem of converting discrete ink dots to continuous halftone matrics is transformed into the non-differentiable problem of one-hot encodings, resulting in a simplified problem. We learn a model O-Net ('O' stands for "one-hot encoding") to perform the mapping from discrete coordinates $\boldsymbol{x}_i$, $\boldsymbol{y}_i$ to one-hot encoding $\boldsymbol{o}(\boldsymbol{x}_i)$, $\boldsymbol{o}(\boldsymbol{y}_i)$. Denote O-Net as $\mathbf{O}$ for brevity, which consists of fully connected layers in our implementation. ReLU activation function is used between each layer. During the optimization of O-Net, the potential coordinate values $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ are fed into $\mathbf{O}$, and the gradient is computed utilizing the following loss function $\mathcal{L}_{\mathrm{O-Net}}$:

$$\begin{aligned}\mathcal{L}_{\mathrm{O-Net}} = &\\ \frac{1}{\max(L_1, L_2)/s} &\sum_{l=0}^{\max(L_1,L_2)-1} |\mathbf{O}(\tilde{\boldsymbol{x}}_l) - \mathrm{OneHotEnc}(\tilde{\boldsymbol{x}}_l)| + \\ \frac{1}{\max(L_1, L_2)/s} &\sum_{l=0}^{\max(L_1,L_2)-1} |\mathbf{O}(\tilde{\boldsymbol{y}}_l) - \mathrm{OneHotEnc}(\tilde{\boldsymbol{y}}_l)|,\end{aligned} \tag{4}$$

where $\mathrm{OneHotEncoding}(\tilde{\boldsymbol{x}}_l)$ and $\mathrm{OneHotEncoding}(\tilde{\boldsymbol{y}}_l)$ denote the one-hot vector determined by $\tilde{\boldsymbol{x}}_l$ and $\tilde{\boldsymbol{y}}_l$, respectively. s represents the sampling step. It is worth noting that the dot coordinates $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are decimal numbers. Therefore, to ensure that the network $\mathbf{O}$ can accurately handle different coordinate values, it is necessary to densely sample the interval $[0, \max(L_1, L_2) - 1]$. In our experiments, we use a sampling step of 0.01 for this interval. The architecture of O-Net is illustrated in Figure 4.

The precision $\mathcal{P}(\mathbf{O})$ is calculated as follows:

$$\begin{aligned}\mathcal{P}(\mathbf{O}) = &\frac{1}{\max(L_1, L_2)/s} \sum_{l=0}^{\max(L_1,L_2)-1} \mathbb{1}\{\arg\max(\mathbf{O}(\tilde{\boldsymbol{x}}_l)), \\ &\arg\max(\mathrm{OneHotEncoding}(\tilde{\boldsymbol{x}}_l))\} + \\ &\frac{1}{\max(L_1, L_2)/s} \sum_{l=0}^{\max(L_1,L_2)-1} \mathbb{1}\{\arg\max(\mathbf{O}(\tilde{\boldsymbol{y}}_l)), \\ &\arg\max(\mathrm{OneHotEncoding}(\tilde{\boldsymbol{y}}_l))\},\end{aligned} \tag{5}$$

where $\mathbb{1}\{a, b\}$ is 1 when a and b are equal, and 0 otherwise. In the experiments, O-Net is first trained to convergence and its parameters are fixed during halftone generation.

Discussion. We eschew the use of embedding matrix lookups to establish the mapping between dot coordinates and one-hot encodings due to their non-differentiability.

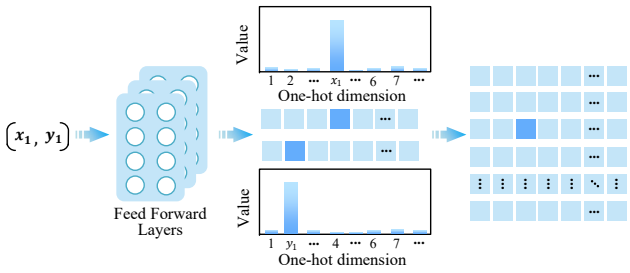


Figure 4. Illustration of O-Net for differentiable transformation from discrete dots to continuous vector-based halftone representations.

This characteristic disrupts gradients and impedes the process of backpropagation.

3.2. P-Net: Ink Dot Motion Prediction

A primary contribution of this work is the achievement of dot controllable halftones, with the prediction of the ink dot motion playing a crucial role in this realization. While some deep learning methods have introduced techniques for point motion control and prediction, exemplified in point cloud generation [1, 4, 24, 28, 33, 36], applying these methods directly to halftoning is not feasible for the following reasons: (1) Point clouds typically consist of a small number of points, for instance, the $2,048$ points utilized in [24] to generate point clouds of various shapes. In contrast, halftones comprise a substantial number of ink dots, with $20,000 \sim 30,000$ ink dots in a 256×256 resolution image. As the dot count escalates, optimizing the model becomes progressively more challenging. (2) Point cloud generation methods are primarily employed to create point cloud shapes with a coarse level of detail. In contrast, halftones incorporate intricate details and textures, demanding precise and meticulous control over the movement of each ink dot to achieve a fine-grained result.

To tackle the aforementioned challenges, we propose P-Net, a network that employs image gray levels as guidance to achieve ink dot motion prediction, as shown in Figure 5. Inspired by recent diffusion generative models [9, 29–31], we explore the efficacy of simple and lightweight convolutional blocks for dot motion prediction. P-Net takes the halftone $\mathbf{H}$ formed by ink dots (Equation 2) and the continuous tone image $\mathbf{C}$ as input. Each element in the matrix $\mathbf{C}$ represents the gray level of the image at the corresponding position. The approach employed by P-Net to integrate information from $\mathbf{H}$ and $\mathbf{C}$ is relatively straightforward, involving concatenation of the two matrices along the channel dimension:

$$\mathbf{P}_{\mathrm{in}} = \mathrm{Conv2D}(\mathbf{H} \,||\, \mathbf{C}), \tag{6}$$

where $\mathrm{Conv2D}$ means two-dimensional convolution, $||$ means the concatenation operation. While alternative approaches, including post-smooth concatenation and density-

Algorithm 1 Algorithm for Dot Controllable Halftoning

Input:
The set of random dots, $\mathcal{Q} = \{\boldsymbol{q}_0, \boldsymbol{q}_1, ..., \boldsymbol{q}_{k-1}\}$;
The number of dots in the set $\mathcal{Q}$, k;
Continuous tone image, $\mathbf{C}$;
The number of iterations, τ;
Output:
Generated halftone, $\mathbf{H}_{\text{out}}$;

Transform the dots in the set $\mathcal{Q}$ into halftone $\mathbf{H}^{(0)}$ using `O-Net`; /* Equation 2 */
for $i = 0$ to $\tau - 1$ **do**
  Feed $\mathbf{H}^{(i)}$ and $\mathbf{C}$ into `P-Net` as input; /* Equation 6 */
  Obtain motion predictions of each dot in $\mathcal{Q}$: $\mathbf{P}_{\text{out}}^{(i)} \leftarrow$ `P-Net`$(\mathbf{H}^{(i)}, \mathbf{C})$; /* Equation 8 */
  Update the position of ink dot $\forall \boldsymbol{q}_i \in \mathcal{Q}$, and update the set $\mathcal{Q}$; /* Equation 9 */
  Transform the dots in the set $\mathcal{Q}$ into halftone $\mathbf{H}^{(i+1)}$ using `O-Net`; /* Equation 2 */
  Calculate $\mathcal{L}_{\text{P}}$ and update the model parameters of `P-Net`; /* Equation 13 */
end for
return $\mathbf{H}_{\text{out}} \leftarrow \mathbf{H}^{\tau}$;

integral map-based methods, have been explored in the experiments, they demand greater computation and do not yield substantial performance enhancements. P-Net utilizes a sequence of convolution modules to acquire the fusion representation of matrices $\mathbf{H}$ and $\mathbf{C}$, followed by several downsampling modules to reduce the image resolution:

$$\widetilde{\mathbf{P}} = \text{DownSample}(\text{StackedConv2D}(\mathbf{P}_{\text{in}})), \tag{7}$$

where $\widetilde{\mathbf{P}}$ represents the feature matrix of high-level gray scale information, which coarsely expresses the overall movement trend of the ink dots within adjacent image blocks. Next, a prediction head composed of convolutions is used on the upsampled $\widetilde{\mathbf{P}}$ matrix to make the output dot motion $\mathbf{P}_{\text{out}}$ consistent with the resolutions of the halftone matrix $\mathbf{H}$:

$$\mathbf{P}_{\text{out}} = \text{StackedConv2D}(\text{UpSample}(\widetilde{\mathbf{P}})). \tag{8}$$

Each element in $\mathbf{P}_{\text{out}}$ represents the displacement vector of the ink dot at the corresponding position, and it has two channels representing the motion prediction along x and y directions, respectively. For $\forall \boldsymbol{q}_i \in \mathcal{Q}, \boldsymbol{q}_i = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, its new position can be derived as:

$$(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{y}}_i) = (\boldsymbol{x}_i, \boldsymbol{y}_i) + \mathbf{P}_{\text{out}}\left[\lfloor \boldsymbol{x}_i \rfloor, \lfloor \boldsymbol{y}_i \rfloor\right]. \tag{9}$$

The ink dot coordinates $\hat{\boldsymbol{q}}_i = \{\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{y}}_i\}$ after the motion can be differentiated to form a new halftone matrix $\widehat{\mathbf{H}}$ through O-Net, and the iteration is completed.

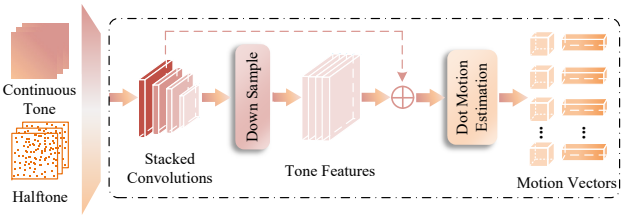


Figure 5. Schematic diagram of the proposed P-Net, which is used to predict the motion of each ink dot.

3.3. Training Scheduling and Optimization

To optimize our model, following [27, 35], we use the tone loss $\mathcal{L}_T$, which computes the dissimilarity between halftone predictions $\mathbf{H}_{\text{out}}$ and continuous tone images $\mathbf{C}$:

$$\mathcal{L}_T = \mathbb{E}_{\mathbf{H}} \left\| \text{GauSmooth}\left(\mathbf{H}_{\text{out}}\right) - \text{GauSmooth}\left(\mathbf{C}\right) \right\|_2, \tag{10}$$

where $\|\cdot\|_2$ represents the L2 norm. To improve the structural similarity of the generated halftones, the structural loss $\mathcal{L}_S$ is employed:

$$\mathcal{L}_S = \mathbb{E}_{\mathbf{H}} \left\| \text{Struc\_Measure}\left(\mathbf{H}_{\text{out}}, \mathbf{C}\right) \right\|_1. \tag{11}$$

Additionally, the pixel loss $\mathcal{L}_R$ is adopted to measure the difference between the generated halftone after inverse halftoning transformation and the continuous tone image:

$$\mathcal{L}_R = \mathbb{E}_{\mathbf{H}} \left\| \mathcal{R}\left(\mathbf{H}_{\text{out}}\right) - \mathbf{C} \right\|_2, \tag{12}$$

where $\mathcal{R}$ represents the frozen inverse halftoning network [35]. The total loss $\mathcal{L}_P$ is as follows:

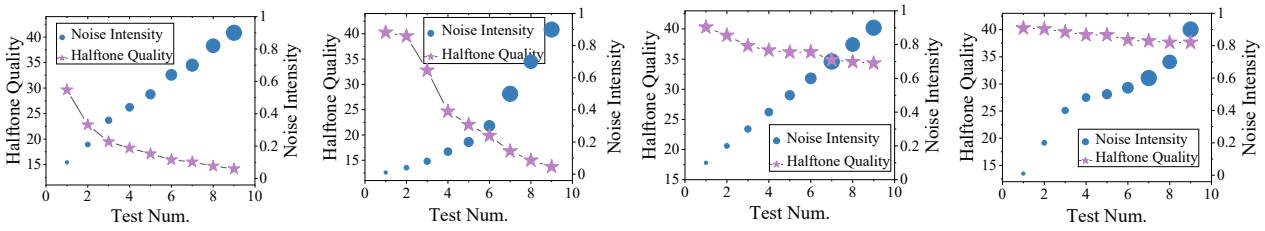
$$\mathcal{L}_P = \mathcal{L}_T + \gamma_S \mathcal{L}_S + \gamma_R \mathcal{L}_R, \tag{13}$$

where γ_S and γ_R are hyperparameters to trade-off each loss term. We detail the training process in Algorithm 1.

4. Experiments

4.1. Datasets, Baselines, and Evaluation Protocols

Datasets and Baselines. Following previous research [13, 35], we curate a halftoning dataset by selecting images from UTKFace [37] and DIV2K [2]. The dataset comprises 4000, 1000, and 400 images in the training, validation, and test sets, respectively. We select baseline models from two categories: traditional halftoning methods [12, 21, 26] and deep learning-based halftoning methods [35]. The traditional methods include amplitude modulation techniques (such as Ordered Dithering), frequency modulation techniques (such as Floyd-Steinberg Error Diffusion, Simpler Floyd-Steinberg Error Diffusion, Jarvis, Judice, and Ninke Dithering, Stucki Dithering, Atkinson Dithering, Burkes Dithering, Sierra Dithering, Two Row Sierra Dithering, and Sierra Lite Dithering), and improvements of Floyd-Steinberg error diffusion algorithm (such as Ostromoukhov's Method).



(a) training iteration=1000 (b) training iteration=3000 (c) training iteration=6000 (d) training iteration=9000

Figure 6. O-Net robustness test experiment. The figure shows the variation trend of the halftone quality obtained by O-Net with the noise intensity. a, b, c, and d represent 1000, 3000, 6000, and 9000 model iterations ($\times 10^4$), respectively.

Table 1. Performance comparison of different halftoning methods on test images in terms of peak signal to noise ratio.

Method	Min.	Max.	Std. Dev.	Avg.
Ordered Dithering [21]	29.69	36.49	1.14	33.04
Floyd-S. [12]	35.31	44.40	1.72	39.52
Simpler Floyd-S. [21]	32.52	41.61	1.64	36.81
Jarvis, J., and N. [21]	27.70	38.29	1.85	32.67
Stucki [21]	28.57	39.10	1.85	33.40
Atkinson [21]	21.78	31.77	1.82	26.36
Burkes [21]	31.74	41.58	1.75	35.62
Sierra [21]	28.23	38.51	1.87	33.25
Two Row Sierra [21]	30.82	41.21	1.73	34.86
Sierra Lite [21]	35.95	45.58	1.65	40.90
Ostromoukhov [26]	36.04	45.90	1.76	41.18
Xia's Method [35]	29.89	35.94	0.81	33.83
Our Proposed Method	34.74	46.64	2.18	**43.04**

Regarding deep learning methods, we select [35] as the baseline, which is a recently proposed deep learning-based halftoning approach.

Evaluation Protocols and Implementation Details. Following previous work [35], we use Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) to measure the quality of the generated halftones. Before computing PSNR, we first smooth both the halftone and continuous tone image with an 11×11 Gaussian filter, and then use the smoothed images to calculate the metrics. In addition, we also use qualitative experiments to evaluate the performance of different methods. The quantity of initial dots k is determined through a linear transformation based on the grayscale of the continuous tone. In P-Net, the number of channels is 64, the convolution kernel size is set to 3×3. O-Net consists of 4 fully connected layers, and the activation function between layers is ReLU. γ_S and γ_R are set to 0.04 and 0.08 respectively. Adam optimizer [19] is used to optimize the model, and the learning rate is set to 0.0001.

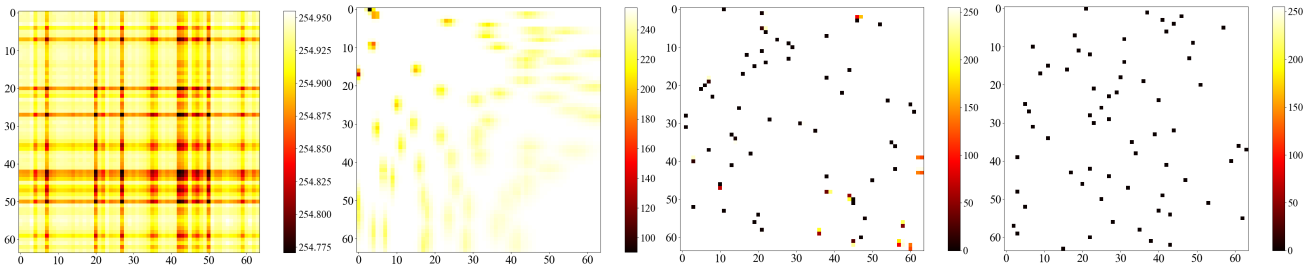
4.2. Performance Comparison

We conduct quantitative experiments on images with 64×64 resolutions, and the experimental results are shown in Table 1. It can be observed that our proposed method achieves superior performance than previous methods. Traditional methods (*e.g.*, Ordered Dithering, Floyd-Steinberg Error Diffusion, Ostromoukhov's Method) usually predefine some dithering matrices or diffusion coefficients, which achieve relatively good results. However, these traditional methods apply a hand-crafted halftoning criterion to all images indiscriminately, and thus have limited flexibility compared to data-driven approaches like [35]. Besides, previous methods perform halftone generation on the global image and cannot adaptively control the motion of each ink dot, which makes it difficult to flexibly remove some undesired artifacts in the generated halftones. In contrast, our method can flexibly control the motion of each ink dot on the halftone, thus achieving better experimental results.

4.3. Study on the Effectiveness of O-Net

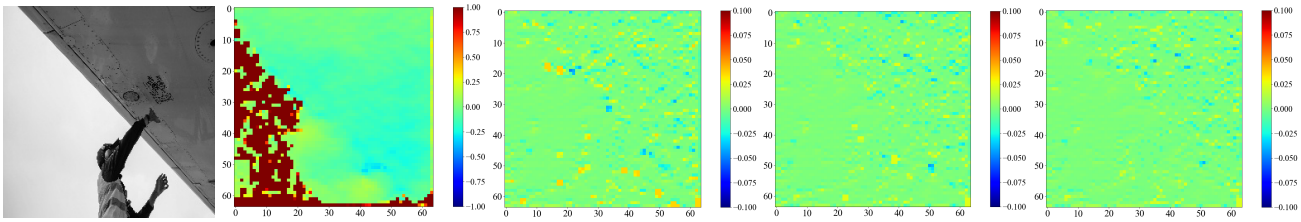
A well-performing O-Net should be able to handle various coordinate values produced by model predictions. Although we sample the $[0, \max(L_1, L_2) - 1]$ interval as densely as possible during training, it is still not enough to cover various float coordinate values that may be encountered during model training. In light of this, we design some experiments to verify the robustness of O-Net to numerical noise of float numbers, as shown in Figure 6. The main process of the experiment is that we randomly add noises of different intensities to the correct coordinates of ink dots, and observe the prediction precision of O-Net after adding noises. The magnitude of the noise is constrained between $[0, 1]$, which is to ensure that the original position of the ink dot will not be changed when the noise is added to the coordinates. The intensity of the noise is defined by its coverage percentage of the coordinate value. As illustrated in Figure 6, during the initial stages of training, noise interference adversely impacts the quality of the halftones generated by the model. However, as training advances towards convergence, the model becomes less sensitive to noise interference, affirming the robustness of O-Net.

In addition, we visualize the heat maps of O-Net prediction results at different stages of training, as shown in Figure 7. We first randomly initialize 64 ink dot coordinates on a 64×64 resolution image, then feed these 64 dot coordinates into O-Net, and visualize the model prediction results. As depicted in Figure 7, prior to the convergence of O-Net, the predicted image exhibits numerous non-binarized gray



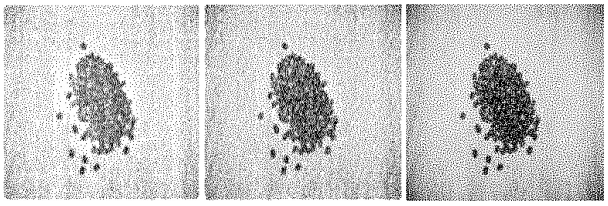
(a) random initialized (b) training iteration=10 (c) training iteration=6000 (d) training iteration=9000

Figure 7. Visualization of model predictions at different stages of O-Net training. a, b, c, and d represent 0, 10, 6000, and 9000 model iterations ($\times 10^4$), respectively. The effectiveness of O-Net increases as the image transitions towards a binary representation.



(a) continuous tone (b) iteration #1 (c) iteration #2 (d) iteration #3 (e) iteration #4

Figure 8. Visualization of ink dot motion predictions over several iterations. Here we visualize the motions on x-axis, normalized to 0 to 1.



(a) quantity level 1 (b) quantity level 2 (c) quantity level 3

Figure 9. Illustration of the resulting halftones with different numbers of ink dots. Number of ink dots: (a): 10K; (b): 14K; (c): 18K. More ink dots enrich the texture details of the image.

pixels, indicating that the one-hot encoding at this stage is not fully optimized. In contrast, upon O-Net convergence, the generated halftone comprises solely black and white pixels, affirming the capability of O-Net in producing binary halftones.

4.4. Comparison of Structural Quality

The structural quality of halftones is also an important indicator for evaluating the performance of an algorithm. We compare the structural similarity with methods that generate halftones with higher PSNR, as shown in Table 2. It can be observed that our proposed method achieves better structural similarity than the baseline methods.

4.5. Visualization of Motion Predictions

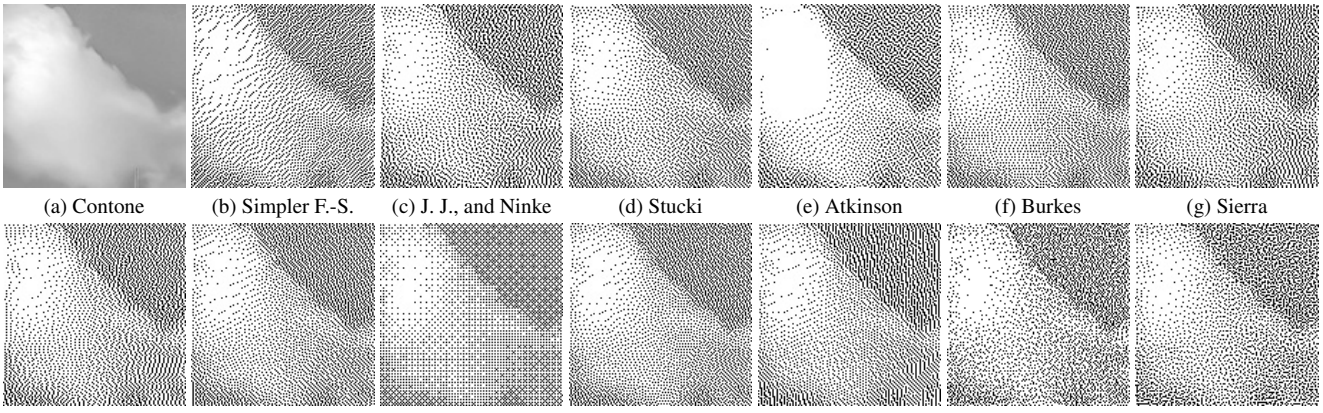
We depict the dot motion predictions generated by P-Net across multiple iterations, as illustrated in Figure 8. Notably, the model rapidly shifts the ink dots from regions with elevated grayscale to those with lower grayscale in the initial stage. This observation substantiates that the motion predictions of P-Net are grounded in a keen awareness of image grayscale. With the progression of iterations, the motions converge towards zero across the majority of the image, leading the model to confine ink dot movements to a limited local area.

Table 2. Performance comparison of different halftoning methods on test images in terms of structural similarity.

Method	Min.	Max.	Std. Dev.	Avg.
Ordered Dithering [21]	0.0254	0.3070	0.0486	0.1384
Floyd-S. [12]	0.0296	0.4335	0.0604	0.1687
Simpler Floyd.-S. [21]	0.0295	0.4610	0.0629	0.1664
Sierra Lite [21]	0.0284	0.4110	0.0596	0.1660
Ostromoukhov [26]	0.0265	0.4185	0.0590	0.1628
Our Proposed Method	0.0305	0.4159	0.0628	**0.1709**

4.6. Qualitative Experiments

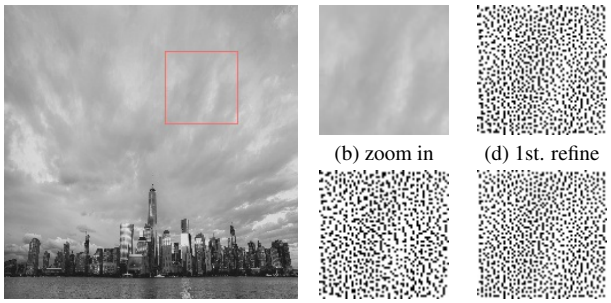
Figure 10 presents qualitative experimental results of different halftone methods, which focus on generating halftones on flat textures with subtle variations in image tone (*e.g.*, light and shadow on clouds). Earlier efforts yield halftones with either undesirable white holes or repetitive streaks of ink dots, compromising the overall quality of the output. Additionally, deep learning methods (Figure 10m) introduce visual artifacts and struggle to faithfully reproduce subtle variations in light and shade. In contrast, our method better preserves tonal details in flat regions and achieves su-



(a) Contone (b) Simpler F.-S. (c) J. J., and Ninke (d) Stucki (e) Atkinson (f) Burkes (g) Sierra

(h) Two Row Sierra (i) Sierra Lite (j) Ordered Dither (k) Floyd-Steinberg (l) Ostromoukhov (m) Xia (n) Ours

Figure 10. Qualitative experimental results. Baseline methods yields halftones with some imperfections, including holes (e, m), striped dot distribution (b, f, g, h, l), or other regular visual artifacts (c, d, i, j, k). Images are from the DIV2K dataset [2].



(a) continuous tone (b) zoom in (c) init. pred. (d) 1st. refine (e) 2nd. refine

Figure 11. Experimental results of refining areas with artifacts in generated halftones by manipulating local ink dots. PSNR: (c): 37.24; (d): 44.10; (e): 44.43. SSIM: (c): 0.00756; (d): 0.00772; (e): 0.00776.

Table 3. Statistical results of collisions during ink dot movement.

Iteration (τ)	4	8	12	16	20
Reduction Ratio	0.0314	0.0249	0.0296	0.0316	0.0287

perior visual results. Additional qualitative experimental results, encompassing high-resolution and color images, are presented in the supplementary material.

4.7. Study on the Collision Rate of Ink Dots

Ink dots may experience collisions during their movement. We compute the reduction ratio in the number of ink dots resulting from these collisions and overlaps, as outlined in Table 3. It can be observed that the reduction in the number of ink dots due to collisions is minimal, constituting approximately 2% to 3% of the total ink dot count. This observation suggests that collisions have a negligible impact on both the overall quantity of ink dots and the dithering process in halftones. We will investigate methods to further minimize ink dot collisions and develop strategies for managing colliding ink dots in our future work.

4.8. Dot Control Helps Remove Local Artifacts

We explore an advantage of controllable ink dots: the capability to enhance specific areas within generated halftones through precise manipulation of the ink dots. As illustrated in Figure 11, the initially generated halftone may exhibit artifacts in local regions (Figure 11c). To address this, we refine the halftone by readjusting the ink dots in these areas, resulting in a further enhancement of the generated halftone's quality and the elimination of artifacts (Figure 11d, 11e). This process serves as validation for the efficacy of our proposed concept of controllable ink dots.

4.9. Study on Ink Dot Quantity Control

We investigate another advantage of dot control: the capacity to regulate the quantity of ink dots. In Figure 9, we present some examples of halftones with a gradual increase in the quantity of ink dots, from level 1 (Figure 9a) to level 3 (Figure 9c). Notably, even with a smaller number of ink dots, the halftones still maintain visually pleasing levels of detail. The experimental results verify that the proposed method can control the number of ink dots and generate visually pleasing halftones, making it more flexibly applicable to halftone prints of different consumable levels.

5. Conclusion

We introduce ink dot-oriented neural halftoning in this work. O-Net propagates dense gradients to sparse halftone dots, and completes the differentiable transformation between dot coordinates and halftone matrices. P-Net achieves direct control of halftone dots through dot motion predictions. Extensive quantitative and qualitative experiments verify the effectiveness of our proposed method.

Acknowledgement: This work is supported by National Key R&D Program of China (2022ZD0160300).

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018. 4

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 1, 5, 8

[3] David Blatner, Stephen F Roth, et al. *Real world scanning and halftones*. Peachpit Press, 1993. 2

[4] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snavely, and Bharath Hariharan. Learning gradient fields for shape generation. In *European Conference on Computer Vision*, pages 364–381. Springer, 2020. 4

[5] 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. 2

[6] Jianghao Chang, Benoît Alain, and Victor Ostromoukhov. Structure-aware error diffusion. In *ACM SIGGRAPH Asia*, pages 1–8. 2009. 2

[7] Niranjan Damera-Venkata and Brian L Evans. Adaptive threshold modulation for error diffusion halftoning. *IEEE Transactions on Image Processing*, 10(1):104–116, 2001. 2

[8] Niranjan Damera-Venkata and Brian L Evans. Design and analysis of vector color error diffusion halftoning systems. *IEEE transactions on image processing*, 10(10):1552–1565, 2001. 2

[9] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 4

[10] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces. *arXiv preprint arXiv:1512.07679*, 2015. 3

[11] Žitinski Elías, Daniel Nyström Paula, and Sasan Gooran. Multi-channel printing by orthogonal and non-orthogonal am halftoning. In *Proceedings of 12th International AIC Colour Congress: Bringing Colour to Life*, 2013. 2

[12] R.W. Floyd and L. Steinberg. An adaptive algorithm for spatial grey scale. In *Proceedings of the Society of Information Display*, pages 75–77, 1976. 1, 2, 5, 6, 7

[13] 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. 5

[14] Sasan Gooran. A novel hybrid amplitude modulated/frequency modulated halftoning based on multilevel halftoning. *Journal of Imaging Science and Technology*, 50 (2):157–167, 2006. 2

[15] Jianjin Gu and Li Li. A new method of halftoning and inverse halftoning based on gan network. In *Advances in Artificial Intelligence and Security: 7th International Conference, ICAIS 2021, Dublin, Ireland, July 19-23, 2021, Proceedings, Part I 7*, pages 119–131. Springer, 2021. 2, 3

[16] Jing-Ming Guo and S Sankarasrinivasan. H-gan: Deep learning model for halftoning and its reconstruction. In *2020 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–2. IEEE, 2020. 2, 3

[17] Hao Jiang and Yadong Mu. Conditional diffusion process for inverse halftoning. *Advances in Neural Information Processing Systems*, 35:5498–5509, 2022. 3

[18] 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. 2, 3

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[20] Thomas D Kite, Brian L Evans, and Alan C Bovik. Modeling and quality assessment of halftoning by error diffusion. *IEEE Transactions on image processing*, 9(5):909–922, 2000. 2

[21] Daniel L Lau and Gonzalo R Arce. *Modern digital halftoning*. CRC Press, 2018. 5, 6, 7

[22] Pingshan Li and Jan P Allebach. Tone-dependent error diffusion. *IEEE Transactions on Image Processing*, 13(2):201–215, 2004. 2

[23] Jianfeng Lu, Zhiwen Wang, Li Li, Ching-Chun Chang, Ting Luo, and Wei Gu. Towards high-resolution copy-evident ceramic tiles: a deep learning framework for halftoning and watermarking. *Electronics*, 10(15):1833, 2021. 2, 3

[24] 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. 4

[25] Vishal Monga, Niranjan Damera-Venkata, and Brian L Evans. Design of tone-dependent color-error diffusion halftoning systems. *IEEE Transactions on Image Processing*, 16(1):198–211, 2006. 2

[26] Victor Ostromoukhov. A simple and efficient error-diffusion algorithm. *ACM SIGGRAPH*, 9:567–572, 2001. 1, 2, 5, 6, 7

[27] Wai-Man Pang, Yingge Qu, Tien-Tsin Wong, Daniel Cohen-Or, and Pheng-Ann Heng. Structure-aware halftoning. In *ACM SIGGRAPH 2008 papers*, pages 1–8. 2008. 5

[28] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3859–3868, 2019. 4

[29] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020. 4

[30] 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*, 2021.

[31] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, 2023. 4

[32] Dusan C Stulik and Art Kaplan. Halftone. *The Atlas of Analytical Signatures of Photographic Processes*, 2013. 1

[33] Lyne P Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. 4

[34] Ping Wah Wong. Adaptive error diffusion and its application in multiresolution rendering. *IEEE Transactions on image Processing*, 5(7):1184–1196, 1996. 2

[35] 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. 1, 2, 3, 5, 6

[36] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4541–4550, 2019. 4

[37] 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. 5

[38] Bingfeng Zhou and Xifeng Fang. Improving mid-tone quality of variable-coefficient error diffusion using threshold modulation. In *ACM SIGGRAPH*, pages 437–444. 2003. 2

[39] Xia Zhuge, Yuki Hirano, and Koji Nakano. A new hybrid multitoning based on the direct binary search. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2008. 2