E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)04-1118-12

论文引用格式:Wang Y J, Gong J X, Liang Z B, Chong Q P, Cheng X and Xu J D. 2025. Fuzzy diffusion model for seen-through document image restoration. Journal of Image and Graphics, 30(4):1118-1129(王义杰, 龚嘉鑫, 梁宗宝, 崇乾鹏, 程翔, 徐金东. 2025. 面向透射文档图像复原的模糊扩散模型. 中国图象图形学报, 30(4):1118-1129)[DOI:10.11834/jig.240350]

面向透射文档图像复原的模糊扩散模型

王义杰[1],龚嘉鑫[1],梁宗宝[1],崇乾鹏[1],程翔[2],徐金东[1*]

1. 烟台大学计算机与控制工程学院,烟台　264005; 2. 北京大学信息科学与技术学院,北京　100871

摘　要:**目的**　在对文档进行数码成像时,墨水浓度和材质透明度等因素可能会导致文档背面内容透过纸张变得可见,从而导致数字图像中出现透射现象,影响文档图像的实际使用。针对这一现象,提出了一种模糊扩散模型,基于模糊逻辑的均值回归思想,不需要任何先验知识,增强扩散模型处理文档图像中随机因素的能力,不仅解决了文档图像的透射现象,而且增强了图像的视觉效果。**方法**　所提方法通过均值回归随机微分方程连续添加随机噪声降低原始图像质量,将其转换为带有固定高斯噪声的透射均值状态,随后在噪声网络中引入模糊逻辑操作来推理每个像素点的隶属度关系,使模型更好地学习噪声和数据分布,在逆向过程中,通过模拟相应的逆时间随机微分方程来逐渐恢复低质量图像,获得干净的无透射图像。**结果**　将所提方法分别在合成灰度数据集和合成彩色数据集上进行训练,并在3个合成数据集和2个真实数据集上进行测试,与代表性的5种方法进行了比较,所提方法取得了最好的视觉效果,且在一定程度上消除了原始图像中的噪声。在峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity index, SSIM)、学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)和费雷歇初始距离(Fréchet inception distance, FID)4个评价指标上均取得了最好的结果。**结论**　本文方法能够有效地解决不同类型文档图像中的透射现象,提高了文档图像去透射任务的准确性和效率,有望集成到各种摄像头、扫描仪等实际硬件设备。

关键词:扩散模型;模糊逻辑;图像复原;透射去除;随机微分方程(SDE)

Fuzzy diffusion model for seen-through document image restoration

Wang Yijie[1], Gong Jiaxin[1], Liang Zongbao[1], Chong Qianpeng[1], Cheng Xiang[2], Xu Jindong[1*]

1. School of Computer and Control Engineering, Yantai University, Yantai　264005, China;

2. School of Information Science and Technology, Peking University, Beijing　100871, China

Abstract: **Objective** Document images have significant applications across various fields, such as optical character recognition (OCR), historical document restoration, and electronic reading. While scanning or shooting a document, factors such as ink density and paper transparency may cause the content from the reverse side to become visible through the paper, resulting in a digital image with a "seen-through" phenomenon that affects practical applications. Image acquisition is often affected by various sources of uncertainty, including differences in camera equipment performance, paper quality, lighting conditions, lens shake, and variations in the physical properties of the documents themselves. All these random factors contribute to the noise in document images and may complicate the seen-through phenomenon, thereby influencing subsequent tasks such as text recognition, word identification, and layout analysis. Although restoring the content of docu-

收稿日期:2024-07-04;**修回日期**:2024-09-09;**预印本日期**:2024-09-16

*通信作者:徐金东　xujindong@ytu.edu.cn

基金项目:烟台市科技创新发展计划基础研究重点项目(2024JCYT037);国家自然科学基金项目(62072391,62066013)

Supported by: National Natural Science Foundation of China(62072391,62066013)

ment images is important, the backgrounds of many color document images also provide valuable information. Recovering color images with complex backgrounds affected by the seen-through phenomenon presents its own challenges. Despite the improvement in image quality achieved by existing methods for removing seen-through effects from document images, algorithms specifically tailored to handle variations in seen-through effects, complex background colors, and influence of uncertainty factors have not yet been developed. This work aims to develop a comprehensive algorithm for addressing the diverse seen-through problems in regular document images, handwritten document images, and color document images. We propose the fuzzy diffusion model (FDM) that integrates fuzzy logic with conditional diffusion models, introducing a novel approach to document image enhancement and restoration. The objective of this algorithm is to restore document images affected by various types and degrees of seen-through phenomenon. **Method** The overall process of this algorithm can be divided into forward diffusion and corresponding reverse denoising. First, we gradually add continuous Gaussian noise to the input image using mean-reverting stochastic differential equations, resulting in a seen-through mean state with fixed Gaussian noise. We then train a neural network to progressively predict the noise at the current time step from the image with added noise and estimate the score function based on the predicted noise. In the reverse process, we gradually restore the low-quality image by simulating the corresponding reverse-time stochastic differential equation until a clean image without seen-through effects is generated. To address the uncertainty factors in document images, we specifically design a fuzzy block in the skip connection part of the noise network to compute the affiliation of each pixel point in the image. The fuzzy operation uses nine surrounding pixels including the pixel itself, and the final affiliation of the pixel is obtained after fuzzy inference. We draw inspiration from the U-Net structure in denoising diffusion probabilistic model, except that we remove all group normalization layers and self-attention layers to improve inference efficiency. In the middle part, we introduce atrous spatial pyramid pooling (ASPP) to maximize the expansion of the receptive field and extract richer features. Finding matching pairs of seen-through images in the real world is challenging, so we propose a new protocol for synthetic seen-through images. During training, we input seen-through images as conditional information, along with the noise-added images, into the noise network to allow the model to directionally learn the target distribution. After the model is trained, seen-through images are used as conditional input to progressively predict the noise in the noise image, generating clear document images. **Result** We trained our model separately on the synthetic grayscale dataset and the synthetic color dataset and tested it on three synthetic datasets and two real datasets. The test sets include synthetic grayscale document images, synthetic color document images, synthetic handwritten document images, the Media Team Oulu document dataset, and real CET-6 seen-through document images. Compared with five representative existing methods, our proposed method achieved the best visual effects, effectively eliminating the noise present in the original images to some extent. It also achieved the best results on the following four evaluation metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), learned perceptual image patch similarity (LPIPS), and Fréchet inception distance (FID). Our method achieved the best PSNR and FID on the grayscale dataset with values of 35.05 dB and 30.69, respectively. On the synthetic color dataset, our method obtained the highest SSIM, LPIPS, and FID values of 0.986, 0.005 3, and 20.03, respectively. To validate the stability of the proposed method, we also provided the variance values when evaluating the SSIM metric. Our method achieved the best result of 0.0053. **Conclusion** The proposed FDM effectively addresses various challenges in the task of removing seen-through effects from document images, including the lack of paired seen-through document images, residual seen-through effects, difficulty in handling complex backgrounds, and addressing uncertainty factors in images uniformly. It can effectively and accurately remove the seen-through phenomenon in different types of document images and is expected to be integrated into various practical hardware devices such as cameras and scanners.

Key words: diffusion model; fuzzy logic; image restoration; seen-through removal; stochastic differential equation(SDE)

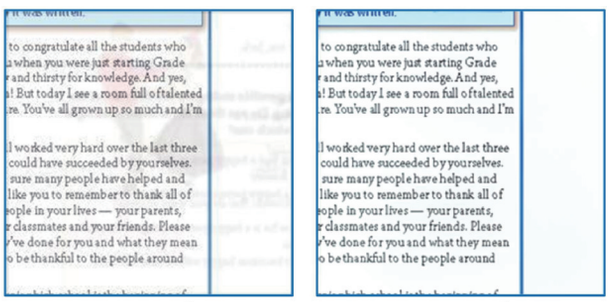
0 引　言

透射现象发生在扫描仪、摄像机或数码相机成像时，文档背面的内容可能会透过纸张变得可见，是文档图像处理和分析中一个重要而具有挑战性的问题。为了获得清晰的文档图像，以便进行后续的文本识别、单词识别和人工阅读等任务，研究可靠的数

字文档图像恢复方法至关重要。如图1所示，透射去除的过程，即透射恢复，从混合图像（图1(a)）中恢复正面内容（图1(b)）。其数学描述为

$$\boldsymbol{I}^{\mathrm{LQ}}=(1-\alpha)\times\boldsymbol{I}^{\mathrm{GT}}+\alpha\times\varphi(\boldsymbol{B}) \tag{1}$$

式中，$\boldsymbol{I}^{\mathrm{LQ}}$是带有透射信息的混合图像，$\boldsymbol{I}^{\mathrm{GT}}$是正面图像，$\boldsymbol{B}$是背面图像，$\alpha$表示透射混合参数，$\varphi(\cdot)$表示透射衰减函数。由此得到的混合图像$\boldsymbol{I}^{\mathrm{LQ}}$可以表示为正面图像$\boldsymbol{I}^{\mathrm{GT}}$和背面图像$\boldsymbol{B}$的线性组合。然而，由于$\boldsymbol{I}^{\mathrm{GT}}$、$\boldsymbol{B}$、$\varphi(\cdot)$和$\alpha$都是未知的，这可能导致从$\boldsymbol{I}^{\mathrm{LQ}}$恢复$\boldsymbol{I}^{\mathrm{GT}}$存在无穷多个可能的解，使其成为一个NP难问题。此外，由于图像属性和内容结构的相似性，在保留正面细节和内容的同时去除背面信息存在极大的困难。尽管现有的传统方法（Rowley-Brooke等，2013；Moghaddam和Cheriet，2010；Sun等，2016；Wolf，2010；Tonazzini等，2007；Tan等，2002；Silva等，2009）在去透射方面取得了一定的效果，但存在一些缺点。例如，Sun等人（2016）和Wolf（2010）的方法对参数设置非常敏感，需要进行精细的参数调整，这可能会导致初始状态下的泛化能力较差。Rowley-Brooke等人（2013）依赖于联合直方图分割和图像分割规则，对于复杂的文档，直方图分割可能无法准确地捕捉图像中的细节或区分不同的区域。Moghaddam和Cheriet（2010）使用变分方法来矫正退化文档中的透射现象。该变分模型需要根据文档图像背面的可用性来估计背景，这意味着在背景估计准确性不足的情况下，性能可能会受到影响。Tonazzini等人（2007）利用盲源分离的思想创建了一个近似但物理上合理的线性模型，但文档图像的透射现象并非是简单的线性重叠，特别是在墨水浓度变化和纸张透明度不均匀的情况下，因此该方法可能在复杂的



(a) 透射图像　(b) 复原图像

图1　彩色文档图像透射去除示例

Fig. 1　An example of color document image seen-through removal ((a) image $\boldsymbol{I}^{\mathrm{LQ}}$ is the seen-through image; (b) the image after seen-through removal)

文档图像透射现象下恢复效果不理想。Tan等人（2002）需要手动进行图像配准，精确度难以保证。此外，Silva等人（2009）在具有复杂背景和不同透射程度的彩色文档图像上的有效性尚未得到很好的验证。对于基于深度学习的文档图像恢复方法，如He和Schomaker（2019）、Souibgui和Kessentini（2022）、De等人（2020）、Feng等人（2021）、Quan等人（2020）和Zheng等人（2021）已在各自的领域取得了令人满意的结果。然而，针对不同类型透射图像恢复的深度学习解决方案尚未得到深入研究。

目前，扩散模型（Ho等，2020；Song和Ermon，2019；Song等，2021b）代表了生成模型的新趋势，并已应用于图像处理的各个领域（Yang等，2023；Zhang等，2023a；Özdenizci和Legenstein，2023；Whang等，2022；Luo等，2023；王珍等，2024）。与其他生成模型（Kingma和Welling，2014；Goodfellow等，2014）不同，扩散模型不需要复杂的网络架构，不基于对抗性原则，训练过程更加稳定可靠。扩散模型的基本思想是通过正向扩散过程系统地扰动数据分布中的结构，然后通过学习逆扩散过程来恢复结构。因此，这种方法产生了高度灵活且易于处理的生成模型，能够捕捉复杂的模式，并生成类似于原始分布的新数据点。由于文档可能受到透射、噪声、模糊和其他随机因素的影响，这些不确定性可以通过具有概率属性的扩散模型更好地建模和处理。此外，涉及从文档中去除透射现象的任务通常需要在缺乏足够先验信息的情况下恢复受损图像。扩散模型可以基于观察到的数据估计图像分布，因此不需要先验知识。

图2说明了逐渐在高质量图像中引入噪声，最终将其退化为具有固定高斯噪声的透射均值状态的过程。初始图像$x(0)$通过随机微分方程（stochastic differential equations，SDE）在相应的时间步下降质为$\boldsymbol{\mu}$（低质量图像）$+\varepsilon$，随后通过模拟逆时间SDE恢复高质量图像。逐渐引入噪声使扩散模型能够更有效地捕获数据的分布和结构。这个过程有助于模拟真实世界的场景，从而帮助模型处理更复杂的透射现象。随后，通过模拟逆时间随机微分方程对数据进行采样，以恢复降质图像的原始状态。通过在噪声网络内部引入模糊操作，模型可以有效捕获和建模文档图像中带有的不确定因素。这有助于模型准确地学习和逼近原始数据的

分布。

本文的贡献总结如下：1）提出了一种名为模糊扩散模型（fuzzy diffusion model，FDM）的新型框架。这是首个专门设计用于去除文档图像中透射现象的扩散方法。它引入了一种新的文档恢复和增强方法。同时展示了扩散模型在解决复杂图像恢复挑战方面的显著潜力。2）提出了一种模糊处理模块并引入跳跃连接中，以帮助模型更有效地建模文档图像中存在的不确定因素，提高模型对数据和噪声分布的学习能力。3）创建了 4 个用于透射图像恢复的数据集，包括一个合成灰度数据集、一个合成手写数据集、一个合成彩色数据集和一个真实数据集。这些源数据可以在 https://github.com/Wangyijie1999/Seen-through-Datasets 上获取。

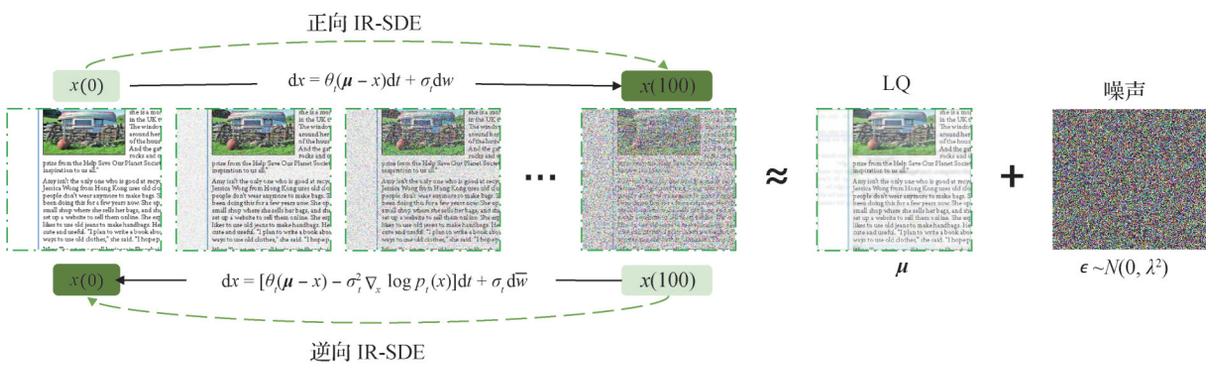


图 2　通过均值回归随机微分方程实现图像透射去除过程

Fig. 2　The process of image seen-through removal is carried out by mean-reverting stochastic differential equation

1　相关工作

1.1　文档增强

文档图像的内容结构与自然图像显著不同，文档图像通常呈现信息密度高的特征，包含大量文本和数据，并可能涵盖不同的字体、背景和样式。因此，为了有效且一致地处理不同类型文档中的透射现象，开发专门的文档图像恢复模型至关重要。目前，大多数文档增强方法（Souibgui 和 Kessentini，2022；Jemni 等，2022；Souibgui 等，2022；Zhao 等，2019；刘雨轩 等，2023）主要集中在提高文档内容的可读性，而忽视了文档背景的重要性。

1.2　基于扩散模型的图像到图像任务

扩散模型在图像生成领域中广泛应用（Yang 等，2023；Luo 等，2023；Fei 等，2023；Li 等，2022；Saharia 等，2023；Zhang 等，2023b；Zhang 等，2024；Shen 等，2024；Liu 等，2024；Zheng 等，2024）。Saharia 等人（2023）提出了 SR3（image super-resolution via iterative refinement），将扩散模型应用于图像超分辨率。Fei 等人（2023）提出了生成式扩散先验（generative diffusion prior，GDP），它利用预训练的 DDPM（denoising diffusion probabilistic models）来解决线性逆问题、非线性问题或盲问题。Shen 等人（2024）提出了 RainDiff（deraining via an unpair degradation-conditioned diffusion model），通过引入稳定且非对抗性的不成对循环一致架构来解决收集现实世界成对数据集困难问题；提出退化条件扩散模型，通过一个由多种雨退化的学习先验条件的扩散生成过程来优化所需的输出。Liu 等人（2024）提出了一个新的 Diff-Plugin 框架，使单个预训练扩散模型能够在各种低层任务中生成保真效果。Zheng 等人（2024）提出了 DiffUIR（efficient diffusion model for image restoration），是一种基于扩散模型的选择性沙漏映射策略，通过整合共享分布项和强条件引导，在多个图像复原任务中取得了优良的性能。Yang 等人（2023）提出了 DocDiff（document enhancement via residual diffusion models），通过将扩散模型引入文档图像增强任务中，避免了图像中高频信息的丢失，但其不适用于彩色文档图像。Zhang 等人（2024）提出了 DECDM（document enhancement using cycle-consistent diffusion models），是一种端到端文档图像翻译方法，通过独立训练源模型和目标模型，克服了成对训练的局限性，但未涉及透射图像复原。Luo 等人（2023）提出了 IR-SDE（image restoration with mean-reverting stochastic differential equations），在自然图像的恢复

方面表现出了良好的效果，但尚未应用于文档图像处理领域。

1.3 随机微分方程

随机微分方程对于扩散建模的整个发展历程至关重要，它提供了描述随机过程中随机变化的数学框架。Song等人(2021b)利用SDE在理论上将DDPM(Ho等，2020)与基于分数的模型(Song和Ermon，2019)统一起来。通过构建扩散模型，复杂的数据分布$x(0)\sim p_{\text{data}}$可以逐渐转化为简单的噪声分布$x(T)\sim p_T = N(0, \boldsymbol{I})$，实现从噪声中对数据进行采样。

扩散过程可以建模为一个随机微分方程，通常是Itô型SDE的解。该SDE描述了数据点在时间段T内的演变过程。在这个过程中，得分函数模拟了数据点相对于时间的位置。具体来说，正向过程为

$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w \tag{2}$$

式中，$f(\cdot,t)\colon \mathbf{R}^d \to \mathbf{R}^d$是表示随机过程$x(t)$的漂移系数的矢量值函数，$g(\cdot)\colon \mathbf{R} \to \mathbf{R}$表示随机过程$x(t)$的扩散系数的标量函数，$w$是标准维纳过程。该方程描述了在随机高斯噪声的扰动下，数据x随时间t的演化过程。根据式(2)，可以在逆时间下模拟该过程从噪声中采样数据(Song等，2021b；Anderson，1982)，具体为

$$\mathrm{d}x = \left[f(x,t) - g(t)^2\nabla_x \log p(x_t)\right]\mathrm{d}t + g(t)\mathrm{d}\overline{w} \tag{3}$$

式中，$\overline{w}$表示逆时间的维纳过程，$p_t(x)$表示在时间t时$x(t)$的边缘概率密度函数。获得得分函数$\nabla_x \log p_t(x)$是困难的，通常需要训练一个依赖于时间的神经网络(Song等，2021a；Hyvärinen，2005)来近似计算。一旦在所有t下得到了得分函数就可以从式(3)推导出逆扩散过程，并模拟它从p_{data}中进行采样。

2 本文方法

2.1 模糊扩散模型

本文所提出的方法使用FDM作为去除文档图像透射现象的基础扩散框架。如图3所示，使用U-Net(Ho等，2020)作为噪声网络。为了提高推理效率，移除了所有的组归一化层和自注意力层。在中间部分，引入空洞空间金字塔池化(atrous spatial pyramid pooling，ASPP)，以最大化感受野，从而提取更丰富的特征。此外，在跳跃连接处提出了一种模糊处理模块，有助于模型更好地学习数据和噪声分布(详见第2.3节)。通过将状态x、时间t和均值$\boldsymbol{\mu}$作为输入，噪声网络输出一个估计的噪声，并且通过L_1损失不断更新网络权重。网络的训练过程涉及在x的不同时间步下预测噪声，并根据噪声和得分函数之间的关系估计得分函数。最终，这个过程使用逆时间SDE公式完成最终的采样。图像退化过程为

$$\mathrm{d}x = \theta_t(\boldsymbol{\mu} - x)\mathrm{d}t + \sigma_t\mathrm{d}w \tag{4}$$

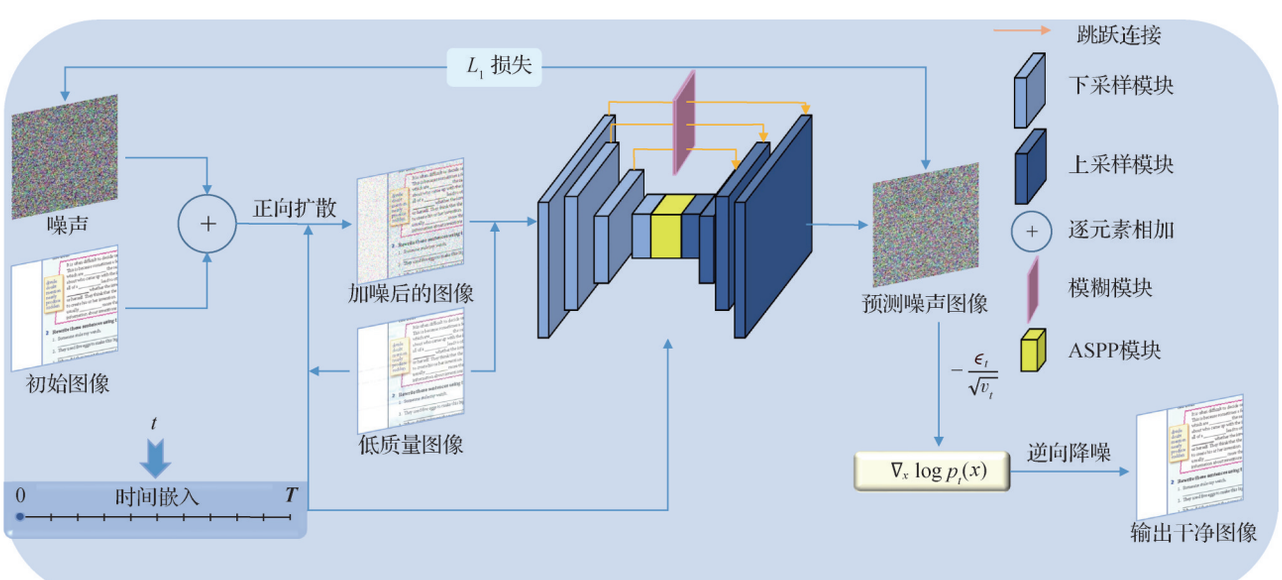


图3 使用模糊扩散模型进行透射去除的简化框架图

Fig. 3 A simplified framework diagram for seen-through removal using the fuzzy diffusion models

式中，θ_t和σ_t表示与时间相关的依赖参数，分别表示均值回归的速率和随机波动的规模。为了更好地捕获图像的退化过程，可以将$\boldsymbol{\mu}$和$x(0)$构造成图像对，$x(0)$表示初始图像，$\boldsymbol{\mu}$表示其退化的透射图像。为了获得方程的闭合形式解，可以建立关系式$\sigma_t^2/\theta_t = 2\lambda^2$，其中$\lambda^2$是稳态方差。如果式(4)中的系数满足$\sigma_t^2/\theta_t = 2\lambda^2$在所有时间$t$下成立，那么边缘分布$p_t(x)$的结果为

$$\begin{cases} p_t(x) \quad = N\left(x(t)\middle|m_t(x), v_t\right) \\ m_t(x) = \boldsymbol{\mu} + (x(0) - \boldsymbol{\mu})\mathrm{e}^{-\bar{\theta}_t} \\ v_t = \lambda^2\left(1 - \mathrm{e}^{-2\bar{\theta}_t}\right) \end{cases} \tag{5}$$

对于任意给定的时间$s \in [0, T]$且$t > s$，SDE的解为

$$x(t) = \boldsymbol{\mu} + (x(s) - \boldsymbol{\mu})\mathrm{e}^{-\bar{\theta}_{s:t}} + \int_s^t \sigma_z \mathrm{e}^{-\bar{\theta}_{z:t}} \mathrm{d}w(z) \tag{6}$$

式中，$x(s)$表示x在先前时间s的值，$\bar{\theta}_{s:t} = \int_s^t \theta_z \mathrm{d}z$表示从$s$到$t$随时间变化的已知参数($\theta_t$预先设置)，转移核$p(x(t)\big|x(s)) = N\left(x(t)\middle|m_{s:t}(x(s)), v_{s:t}\right)$是以$m_{s:t}$为均值、$v_{s:t}$为方差的正态分布，具体为

$$\begin{cases} m_{s:t}(x(s)) = \boldsymbol{\mu} + (x(s) - \boldsymbol{\mu})\mathrm{e}^{-\bar{\theta}_{s:t}} \\ v_{s:t} = \int_s^t \sigma_z^2 \mathrm{e}^{-2\bar{\theta}_{z:t}} \mathrm{d}z = \lambda^2\left(1 - \mathrm{e}^{-2\bar{\theta}_{s:t}}\right) \end{cases} \tag{7}$$

在式(5)中，随着时间t的增加，均值$m_{s:t}$不断接近$\boldsymbol{\mu}$，方差稳步接近λ^2。这意味着给定的$x(0)$逐渐退化并收敛于均值$\boldsymbol{\mu}$，同时保持固定的噪声水平λ^2。这个过程称为均值回归。为了将低质量图像逆转回其初始状态，可以将逆时间SDE(Song等，2021b)表示为

$$\mathrm{d}x = \left[\theta_t(\boldsymbol{\mu} - x) - \sigma_t^2 \nabla_x \log p_t(x)\right]\mathrm{d}t + \sigma_t \mathrm{d}\overline{w} \tag{8}$$

式中，$\theta_t(\boldsymbol{\mu} - x)$为均值回归项，表示随着时间$t$的推移，$x$逐渐收敛到均值$\boldsymbol{\mu}$。$\nabla_x \log p_t(x)$是当前时间$t$下的得分函数。根据式(5)真实的得分函数计算为

$$\nabla_x \log p_t(x) = -\frac{x(t) - m_t(x)}{v_t} \tag{9}$$

通过计算真实的得分函数$\nabla_x \log p_t(x)$可以用来更新模型参数以最大化或最小化特定的目标函数。此外，使用重参数化技术使得$x(t)$的采样可以表示为$x(t) = m_t(x) + \sqrt{v_t}\,\epsilon_t$，其中$\epsilon_t$是标准高斯噪声$\epsilon_t \sim N(0, \boldsymbol{I})$。这样便可以使用式(9)重新定义$\nabla_x \log p_t(x)$为$-\frac{\epsilon_t}{\sqrt{v_t}}$。为了估计低质量图像的噪声，一个常见的方法是训练一个卷积神经网络(Ho等，2020；Song等，2021a)，然后在测试阶段使用逆时间SDE将低质量图像恢复为高质量版本。

2.2 模糊运算

鉴于噪声网络中的噪声预测涉及对数据分布和噪声分布的学习，引入了模糊操作来增强网络对不同分布数据的识别和理解能力。在具有不同背景的彩色图像中，当像素与其相邻像素显著不同时，通常将其分类为背景边界、噪声像素或受其他随机因素影响的像素。为了解决这个问题，本文提出了一种模糊处理模块，它利用二维高斯模糊方法来检测具有不确定性的像素，并整合到U-Net架构的跳跃连接中。通过模糊块操作获得的特征有助于区分噪声和文档内容，从而产生更平滑的图像，图像的内容和背景变得更加统一。此外，图像中原有的噪声得到了一定的抑制。通过建立模糊关系来调节关联程度，具体为

$$\boldsymbol{A}_i = \left\{\left(x, \xi_{A_i}(x, y)\right)\middle|\forall x \in \boldsymbol{F}\right\} \tag{10}$$

式中，ξ_A: $\boldsymbol{F} \to [0, 1]$，$\boldsymbol{A}_i$表示特征图$\boldsymbol{F}$上类别$i$的模糊集合。$\xi_A$是$\boldsymbol{A}_i$的隶属度函数，$\xi_{A_i}(x, y)$表示像素$(x, y)$属于模糊集合$\boldsymbol{A}_i$的隶属度。本文使用二维高斯模糊函数来获取模糊集合$\boldsymbol{A}_i$，每个像素对模糊集合$\boldsymbol{A}_i$的隶属度计算式为

$$\xi'_{A_i}(x, y) = \frac{\mathrm{e}^{-\frac{\left(x_p - x\right)^2 + \left(y_p - y\right)^2}{2\sigma^2}}}{2\pi\sigma^2} \tag{11}$$

式中，$\left(x_p, y_p\right)$表示邻域像素的坐标，而(x, y)表示受影响像素的坐标。σ代表方差，在不同时间步下是可学习的。在像素点的模糊操作中使用了包括自身在内的9个周围像素，并且模糊推理后的最终隶属度为

$$\xi_{A_i}(x, y) = \sum_{\gamma=1}^{9} V_\gamma BN\left(Conv\left(BN\left(\xi'_{A_i}(x, y)\right)\right)\right) \tag{12}$$

式中，V_γ表示元素γ的当前值。该元素的权重由其在分布中的位置决定，中心像素本身具有最大的权重，其次是周围和对角线像素。

图4展示了从噪声图像恢复到高质量图像的可视化过程，其中虚线框中标识了复杂的透射现象。采样过程逐渐去除了图像中的透射现象，同时保留了原始图像的纹理细节和颜色信息。最终的结果生成了一幅视觉上更加清晰、更优质的图像。

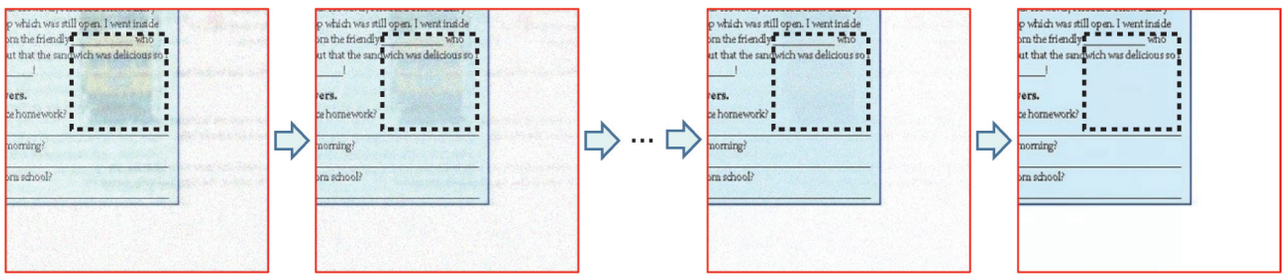


图4　图像去透射的可视化过程

Fig. 4　Visualization process of image seen-through removal

算法1 FDM

干净和退化的文档图像对$(x_0,\boldsymbol{\mu})$，采样步数T，条件扩散模型s_θ。

1）repeat；

2）$(x_0,\boldsymbol{\mu})\sim q(x_0,\boldsymbol{\mu})$；

3）$t\sim \text{Uniform}\{1,\cdots,T\}$；

4）$\epsilon\sim N(0,\boldsymbol{I})$；

5）$\mathrm{d}x=\theta_t(\boldsymbol{\mu}-x)\mathrm{d}t+\sigma_t\mathrm{d}w$；

6）$\xi'_{A_i}(x,y)=\dfrac{\mathrm{e}^{-\frac{(x_p-x)^2+(y_p-y)^2}{2\sigma^2}}}{2\pi\sigma^2}$；

7）采样梯度下降；

$L\left\|\epsilon_t,s_\theta\left(x_t,\boldsymbol{\mu},t\right)\right\|$

8）until converged；

9）$x_T\sim N(0,\boldsymbol{I})$；

10）for $t=T,\cdots,1$ do；

11）$\nabla_x\log p_t(x)\approx-\dfrac{s_\theta\left(x_t,\boldsymbol{\mu},t\right)}{\sqrt{v_t}}$；

12）$\mathrm{d}x=\left[\theta_t(\boldsymbol{\mu}-x)-\sigma_t^2\nabla_x\log p_t(x)\right]\mathrm{d}t+\sigma_t\mathrm{d}\overline{w}$；

13）end for；

14）return $x(0)$。

3　实　验

3.1　实验设置

3.1.1　实施细节

实验硬件环境为：GPU为两张NVIDIA GeForce RTX 3090，显存均为24 GB；CPU为Intel Core i9-12900K 16核处理器，内存为62 GB。本文的实验代码使用Pytorch1.13.0框架，并采用Adam优化器实现。在训练阶段，输入图像调整为256×256像素，训练轮数、批量大小和学习率分别设置为112、4和0.000 1。所提出的方法在5个数据集上进行实验，包括3个合成数据集和两个真实数据集，训练集总共包含2 900幅图像。此外，固定方差λ^2设为10，扩散步骤总数T设为100。

3.1.2　数据集

实际场景中难以找到配对图像，为了验证所提出方法的有效性，本文使用了两个合成数据集：G-1代表灰度图像，而C-1代表彩色图像。对于G-1数据集，本文从网络（https://www.planetebook.com）中收集数据，其中900幅用于训练，160幅用于测试。对于C-1数据集，从初中英语教科书（https://www.pep.com.cn/）中收集数据，其中2 000幅用于训练，316幅用于测试。同时还在一个单独生成的手写合成数据集（https://storage.googleapis.com/openimages/web/index.html）上进行实验。在这些合成数据集中，采用了与$\varphi(\cdot)$相结合的混合线性透射模型，并应用了标准差为2的5×5高斯模糊滤波器，以模拟真实世界的透射现象。$\varphi(\cdot)$的参数在区间0.15～0.25随机选择。此外，为了更准确地模拟透射现象，所有合成数据集都在水平镜像透射层后进行合成。对于真实图像数据集，采用Media Team Oulu Document数据集（Sauvola和Kauniskangas，1999），这是一个包含文档扫描图像和相关文本内容的集合，其中包含了1975年之前发表的500份文档的信息。为了测试，选择了68幅具有透射效果的图像。此外，使用Kyocera FS-1020MFP扫描了CET-6（大学英语六级）试卷，并获取了105幅具有透射现象的真实图像。为了方便训练，数据集中的图像尺寸均为256×256像素，除了手写合成数据集，其图像尺寸为512×512像素。对于测试阶段，本文使用在G-1和C-1数据集上训练的模型。

3.2　评估指标

本文采用传统的像素级度量峰值信噪比（peak signal-to-noise ratio，PSNR）来衡量重建图像与原始

图像之间噪声影响程度，并通过比较图像像素值之间的均方误差（mean square error，MSE）来评估图像的质量。除此之外，为了更准确地评估图像之间的相似度，选择结构相似性指数（structure similarity index measure，SSIM）（Wang等，2004）。SSIM综合考虑了图像的亮度、对比度和结构等方面，为图像相似度提供了介于0和1之间的分数，越接近1表示两个图像之间的相似程度越高。作为图像重建质量的客观指标，学习感知图像块相似性（learned perceptual image patch similarity，LPIPS）（Zhang等，2018）是一种基于深度学习的视觉感知度量方法，用于建模人眼对图像感知的差异。通过深度学习模型学习图像感知的特征表示，并计算图像之间的相似度，LPIPS更加关注图像的语义内容，因此能够更准确地衡量图像之间的视觉差异。另外，费雷歇初始距离（Fréchet inception distance，FID）（Heusel等，2017）利用神经网络提取的特征来衡量生成图像与真实图像分布之间的距离。较低的FID值意味着生成图像分布与真实图像分布更接近，反映了生成模型性能的优越性。

3.3 实验结果与分析

本文在5个数据集上对所提方法与目前流行的方法进行了对比实验。对比方法包括DeepOtsu（document enhancement and binarization using iterative deep learning）（He和Schomaker，2019）、Zou等人（2020）方法、AGAN（attentional generative adversarial network）（Sun等，2022）、S-CycleGAN（self-supervised cycle-consistent generative adversarial networks）（徐金东等，2023）、MPRNet（multi-stage progressive image restoration architecture）（Mehri等，2021）和RAGNet（reflection-aware guidance network）（Li等，2023）。利用PSNR来衡量数据集G-1和C-1中图像的失真程度，通过直接计算整个图像的均方误差来获得PSNR值。采用SSIM指标来评估图像的结构相似性。对于彩色图像，分别计算每个通道的SSIM值，然后对这些值取平均以获得最终的SSIM值。此外，在表1中提供了SSIM值的标准差，以体现所提方法的稳定性。为了评估生成图像的质量，本文引入了FID指标，并取得了最佳结果。尤其对于数据集C-1，本文引入了专门设计用于模拟人眼视觉系统的LPIPS指标以更准确地测量图像之间的差异。提出的方法在G-1数据集上表现出最高的PSNR值，在C-1数据集上表现出最高的SSIM值。在LPIPS指标方面表现出了卓越的性能并接近最佳指标水平。这归因于所提出的方法有效地去除了透射，同时适度降低了原始图像中的噪声，从而提高了整体的视觉效果。

图5展示了在合成和真实数据集上的定性比较实验结果，其中蓝色和红色框中突出显示了与其他方法相比所提出的方法在颜色和细节恢复方面取得的卓越结果。具体而言，DeepOtsu在透射去除方面表现较好，但无法恢复原始图像的颜色信息。Zou等人（2020）和AGAN去除了部分透射现象并保留了部分颜色信息，但总体视觉效果并不理想。S-CycleGAN和MPRNet难以去除复杂的透射现象，并且难以恢复彩色图像的高频信息，恢复效果图不够平滑。RAGNet在灰度数据集上表现优异，能够很

表1 不同算法在G-1和C-1数据集下的指标评估结果

Table 1 Quantitative comparison of different algorithms on G-1 and C-1 datasets

方法	G-1数据集			C-1数据集				
	PSNR/dB ↑	SSIM ↑	FID ↓	PSNR/dB ↑	SSIM ↑	σ ↓	LPIPS ↓	FID ↓
DeepOtsu	22.37	0.902	35.81	20.61	0.824	0.014 7	0.569	60.98
Zou等人（2020）	29.20	0.967	43.20	29.37	0.973	<u>0.008 5</u>	0.023	29.97
AGAN	34.55	<u>0.984</u>	<u>34.13</u>	29.72	0.965	0.020 4	0.027	45.09
S-CycleGAN	<u>34.59</u>	0.977	38.44	**31.54**	<u>0.982</u>	0.054 0	**0.016**	<u>20.42</u>
MPRNet	34.25	**0.985**	51.34	29.57	0.960	0.012 0	0.024	36.47
RAGNet	34.48	**0.985**	34.53	28.46	0.963	0.010 3	0.025	27.59
本文	**35.05**	<u>0.984</u>	**30.69**	<u>30.00</u>	**0.986**	**0.005 3**	<u>0.017</u>	**20.03**

注：加粗、下划线字体分别表示各列最优、次优结果。“↑”表示值越大越好，“↓”表示值越小越好。

好地去除图像中的透射现象，但在彩色数据集和真实数据集上表现并不理想，图像的背景颜色偏差过大，仍然存在部分透射残余。本文方法不仅很好地去除了图像中的透射现象，而且在彩色图像的颜色信息以及高频细节恢复方面表现较好，整体的视觉效果表现优异。通过对扫描的CET-6试卷等真实场景中常见的透射现象进行实验，更加充分地证明了所提方法的适用性。综合多个数据集的实验结果和平均指标验证了所提出方法在消除透射效果方面的优越性。

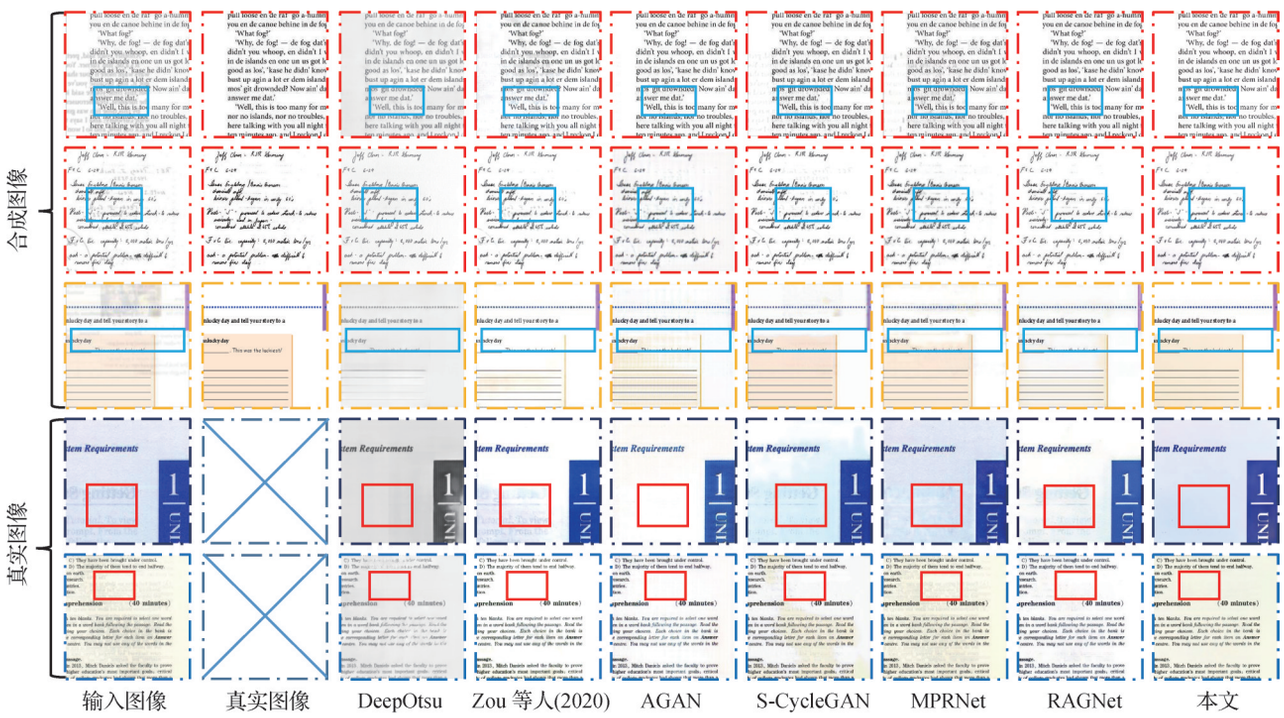


图5　合成数据集和真实数据集文档图像去透射对比结果

Fig. 5　Comparative results of seen-through removal on the synthesis datasets and real datasets

3.4　性能分析

表2列出了所有方法的操作框架、单幅图像的处理时间和模型参数量。尽管本文方法并不复杂，但由于扩散模型的复杂推理过程，导致每幅图像的处理时间较长。现有的一些方法(Song 等，2021a；Rombach 等，2022；Kingma 等，2021；Yang 等，2023)加速了扩散模型的运算效率，从而在处理时间上表现更优。例如，Song等人(2021a)采用非马尔可夫过程推理，逆过程可以跳步采样，避免了多余步骤，因此能够更快速生成高质量样本。Rombach 等人(2022)通过在潜在空间中进行扩散过程，避免了对高维数据进行迭代处理，显著减少了计算量，提高了生成速度。Kingma 等人(2021)对噪声计划和模型参数同时进行优化，同时将变分下限(variational lower bound，VLB)简化为与信噪比相关的短表达，减少计算复杂度，此外，利用VLB对噪声计划的不变性，设计出最小化VLB估计器方差的噪声计划，从而加速优化。Yang等人(2023)通过设计轻量级的网络架构进行辅助任务并与扩散模型进行联合训练，同时将残差图作为扩散模型端的输入，减轻了推理负担，加速了推理过程。本文方法在文档透射去除方面表现出色，但其主要缺点在于扩散模型推理过程的复杂性。这将影响该方法在实际应用中的可

表2　方法性能分析

Table 2　Performance analysis

方法	框架	时间/s	参数量/M
DeepOtsu	TensorFlow	1.803 6	10.9
Zou 等人(2020)	PyTorch	0.040 2	44.6
AGAN	TensorFlow	0.015 7	43.2
S-CycleGAN	PyTorch	0.243 8	14.4
MPRNet	PyTorch	0.164 2	20.1
RAGNet	PyTorch	0.150 0	146.0
本文	PyTorch	15.660 0	13.7

行性，特别是在资源有限的环境中。因此，未来将根据现有的加速策略研究高效的扩散网络模型，通过设计多分支或动态处理策略，减轻扩散模型一端的推理负担，从而实现更快的图像处理速度，同时保持恢复效果。

4 结 论

本文提出了一种模糊扩散模型（FDM）来解决文档图像中的透射现象，是首次利用模糊理论来优化扩散模型的算法，这为由透射现象引起的文档质量问题提供了一种新的解决方案。通过均值回归SDE对复杂的透射信息分布进行建模，并引入模糊处理模块来帮助模型学习噪声和数据分布，随后，对SDE进行反向模拟以从真实数据分布中采样，最终生成高质量的图像。采用模糊扩散模型进行透射去除不仅有效地恢复了受此现象影响的图像，同时去除了原始图像中的部分噪声，这个过程进一步提高了文档图像的可读性。本文方法与多种主流的图像复原方法进行比较，结果表明，本文方法生成的图像质量更好、可读性更高，能够有效解决文档图像的透射问题。

尽管本文方法在多种数据集下表现出卓越的去透射成果，但是扩散模型推理过程中需要多次迭代，从而导致计算时间过长的问题。因此，未来将考虑优化模型以减少不必要的采样步数，在推理过程中采用动态调整采样步数，针对不同透射复杂度的图像，使用不同的步数进行推理，以便在确保去透射性能的情况下加快推理速度。

参考文献（References）

Anderson B D O. 1982. Reverse-time diffusion equation models. Stochastic Processes and their Applications, 12(3): 313-326 [DOI: 10.1016/0304-4149(82)90051-5]

De R, Chakraborty A and Sarkar R. 2020. Document image binarization using dual discriminator generative adversarial networks. IEEE Signal Processing Letters, 27: 1090-1094 [DOI: 10.1109/lsp.2020.3003828]

Fei B, Lyu Z Y, Pan L, Zhang J Z, Yang W D, Luo T Y, Zhang B and Dai B. 2023. Generative diffusion prior for unified image restoration and enhancement//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9935-9946 [DOI: 10.1109/CVPR52729.2023.00958]

Feng X, Pei W J, Jia Z H, Chen F L, Zhang D and Lu G M. 2021. Deep-masking generative network: a unified framework for background restoration from superimposed images. IEEE Transactions on Image Processing, 30: 4867-4882 [DOI: 10.1109/TIP.2021.3076589]

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2672-2680

He S and Schomaker L. 2019. DeepOtsu: document enhancement and binarization using iterative deep learning. Pattern Recognition, 91: 379-390 [DOI: 10.1016/j.patcog.2019.01.025]

Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S. 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6629-6640

Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #574

Hyvärinen A. 2005. Estimation of non-normalized statistical models by score matching. The Journal of Machine Learning Research, 6: 695-709

Jemni S K, Souibgui M A, Kessentini Y and Fornés A. 2022. Enhance to read better: a multi-task adversarial network for handwritten document image enhancement. Pattern Recognition, 123: #108370 [DOI: 10.1016/j.patcog.2021.108370]

Kingma D P, Salimans T, Poole B and Ho J. 2021. Variational diffusion models//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc.: #1660

Kingma D P and Welling M. 2014. Auto-encoding variational Bayes//Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada: ICLR

Li H Y, Yang Y F, Chang M, Chen S Q, Feng H J, Xu Z H, Li Q and Chen Y T. 2022. SRDiff: single image super-resolution with diffusion probabilistic models. Neurocomputing, 479: 47-59 [DOI: 10.1016/j.neucom.2022.01.029]

Li Y, Liu M, Yi Y L, Li Q C, Ren D W and Zuo W M. 2023. Two-stage single image reflection removal with reflection-aware guidance. Applied Intelligence, 53(16): 19433-19448 [DOI: 10.1007/s10489-022-04391-6]

Liu Y H, Ke Z H, Liu F, Zhao N X and Lau R W H. 2024. Diff-plugin: revitalizing details for diffusion-based low-level tasks//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4197-4208 [DOI: 10.1109/

CVPR52733.2024.00402]

Liu Y X, Zhao Q J, Pan F, Gao D G and Pubu D Z. 2023. Structure prior guided text image inpainting model. Journal of Image and Graphics, 28(12): 3699-3712 (刘雨轩, 赵启军, 潘帆, 高定国, 普布旦增. 2023. 结构先验指导的文本图像修复模型. 中国图象图形学报, 28(12): 3699-3712) [DOI: 10.11834/jig.220960]

Luo Z W, Gustafsson F K, Zhao Z, Sjölund J and Schön T B. 2023. Image restoration with mean-reverting stochastic differential equations//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: ACM: #957

Mehri A, Ardakani P B and Sappa A D. 2021. MPRNet: multi-path residual network for lightweight image super resolution//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 2703-2712 [DOI: 10.1109/WACV48630.2021.00275]

Moghaddam R F and Cheriet M. 2010. A variational approach to degraded document enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(8): 1347-1361 [DOI: 10.1109/TPAMI.2009.141]

Özdenizci O and Legenstein R. 2023. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8): 10346-10357 [DOI: 10.1109/TPAMI.2023.3238179]

Quan Y H, Yang J T, Chen Y X, Xu Y and Ji H. 2020. Collaborative deep learning for super-resolving blurry text images. IEEE Transactions on Computational Imaging, 6: 778-790 [DOI: 10.1109/TCI.2020.2981758]

Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]

Rowley-Brooke R, Pitié F and Kokaram A. 2013. A non-parametric framework for document bleed-through removal//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 2954-2960 [DOI: 10.1109/CVPR.2013.380]

Saharia C, Ho J, Chan W, Salimans T, Fleet D J and Norouzi M. 2023. Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4713-4726 [DOI: 10.1109/TPAMI.2022.3204461]

Sauvola J and Kauniskangas H. 1999. MediaTeam document database II, a CD-ROM collection of document images. Finland: University of Oulu

Shen Y Y, Wei M Q, Wang Y Z, Fu X Y and Qin J. 2024. Rethinking real-world image deraining via an unpaired degradation-conditioned diffusion model [EB/OL]. [2024-05-01]. https://arxiv.org/pdf/2301.09430.pdf

Silva J M, Lins R D and Silva G P. 2009. Enhancing the quality of color documents with back-to-front interference//Proceedings of the 6th International Conference on Image Analysis and Recognition. Halifax, Canada: Springer: 875-885 [DOI: 10.1007/978-3-642-02611-9_86]

Song J M, Meng C L and Ermon S. 2021a. Denoising diffusion implicit models//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: OpenReview.net

Song Y and Ermon S. 2019. Generative modeling by estimating gradients of the data distribution//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada: NeurIPS: 11918-11930

Song Y, Sohl-Dickstein J, Kingma D P, Kumar A, Ermon S and Poole B. 2021b. Score-based generative modeling through stochastic differential equations//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: OpenReview.net

Souibgui M A, Biswas S, Jemni S K, Kessentini Y, Fornés A, Lladós J and Pal U. 2022. DocEnTr: an end-to-end document image enhancement Transformer//Proceedings of the 26th International Conference on Pattern Recognition. Montreal, Canada: IEEE: 1699-1705 [DOI: 10.1109/ICPR56361.2022.9956101]

Souibgui M A and Kessentini Y. 2022. DE-GAN: a conditional generative adversarial network for document enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(3): 1180-1191 [DOI: 10.1109/TPAMI.2020.3022406]

Sun B, Li S T, Zhang X P and Sun J. 2016. Blind bleed-through removal for scanned historical document image with conditional random fields. IEEE Transactions on Image Processing, 25(12): 5702-5712 [DOI: 10.1109/TIP.2016.2614133]

Sun X, Xu J D, Ma Y L, Zhao T Y, Ou S F and Peng L Z. 2022. Blind image separation based on attentional generative adversarial network. Journal of Ambient Intelligence and Humanized Computing, 13(3): 1397-1404 [DOI: 10.1007/s12652-020-02637-0]

Tan C L, Cao R and Shen P Y. 2002. Restoration of archival documents using a wavelet technique. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(10): 1399-1404 [DOI: 10.1109/TPAMI.2002.1039211]

Tonazzini A, Salerno E and Bedini L. 2007. Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique. International Journal of Document Analysis and Recognition (IJDAR), 10(1): 17-25 [DOI: 10.1007/s10032-006-0015-z]

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Wang Z, Huo G L, Lan H, Hu J M and Wei X. 2024. Fundus image enhancement algorithm based on convolutional dictionary diffusion model. Journal of Image and Graphics, 29(8): 2426-2438 (王珍,

霍光磊，兰海，胡建民，魏宪. 2024. 基于卷积字典扩散模型的眼底图像增强算法. 中国图象图形学报，29(8)：2426-2438)［DOI：10.11834/jig.230595］

Whang J，Delbracio M，Talebi H，Saharia C，Dimakis A G and Milanfar P. 2022. Deblurring via stochastic refinement//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA：IEEE：16272-16282［DOI：10.1109/CVPR52688.2022.01581］

Wolf C. 2010. Document ink bleed-through removal with two hidden Markov random fields and a single observation field. IEEE Transactions on Pattern Analysis and Machine Intelligence，32(3)：431-447［DOI：10.1109/TPAMI.2009.33］

Xu J D，Ma Y L，Liang Z B and Ni M Y. 2023. Single bleed-through image restoration with self-supervised learning. Acta Automatica Sinica，49(1)：219-228（徐金东，马咏莉，梁宗宝，倪梦莹. 2023. 自监督学习的单幅透射图像恢复. 自动化学报，49(1)：219-228）［DOI：10.16383/j.aas.c220165］

Yang Z Y，Liu B L，Xiong Y P，Yi L，Wu G B，Tang X J，Liu Z Q，Zhou J J and Zhang X. 2023. DocDiff：document enhancement via residual diffusion models//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa，Canada：ACM：2795-2806［DOI：10.1145/3581783.3611730］

Zhang G H，Ji J B，Zhang Y，Yu M，Jaakkola T and Chang S Y. 2023a. Towards coherent image inpainting using denoising diffusion implicit models//Proceedings of the 40th International Conference on Machine Learning. Honolulu，USA：ACM：#1726

Zhang J X，Rimchala J，Mouatadid L，Das K and Kumar S. 2024. DECDM：document enhancement using cycle-consistent diffusion models//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa，USA：IEEE：8021-8030［DOI：10.1109/WACV57701.2024.00785］

Zhang L M，Rao A Y and Agrawala M. 2023b. Adding conditional control to text-to-image diffusion models//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris，France：IEEE：3813-3824［DOI：10.1109/ICCV51070.2023.00355］

Zhang R，Isola P，Efros A A，Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA：IEEE：586-595［DOI：10.1109/CVPR.2018.00068］

Zhao J Y，Shi C Z，Jia F X，Wang Y N and Xiao B H. 2019. Document image binarization with cascaded generators of conditional generative adversarial networks. Pattern Recognition，96：#106968［DOI：10.1016/j.patcog.2019.106968］

Zheng D，Wu X M，Yang S Z，Zhang J，Hu J F and Zheng W S. 2024. Selective hourglass mapping for universal image restoration based on diffusion model//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：25445-25455［DOI：10.1109/CVPR52733.2024.02404］

Zheng Q，Shi B X，Chen J N，Jiang X D，Duan L Y and Kot A C. 2021. Single image reflection removal with absorption effect//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，USA：IEEE：13390-13399［DOI：10.1109/CVPR46437.2021.01319］

Zou Z X，Lei S，Shi T Y，Shi Z W and Ye J P. 2020. Deep adversarial decomposition：a unified framework for separating superimposed images//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：12803-12813［DOI：10.1109/CVPR42600.2020.01282］

作者简介

王义杰，男，硕士研究生，主要研究方向为计算机视觉与图像复原。E-mail：15610121683@163.com

徐金东，通信作者，男，教授，主要研究方向为图像处理、模式识别和盲源分离。E-mail：xujindong@ytu.edu.cn

龚嘉鑫，男，硕士研究生，主要研究方向为计算机视觉与图像复原。E-mail：202300358064@s.ytu.edu.cn

梁宗宝，男，硕士研究生，主要研究方向为计算机觉与图像复原。E-mail：liangzongbao_yut@163.com

崇乾鹏，男，硕士研究生，主要研究方向为计算机觉与图像分割。E-mail：chongqp1999@s.ytu.edu.cn

程翔，男，教授，主要研究方向为无线通信和网络与数据分析处理。E-mail：cheng@pku.edu.cn