基于深度学习的图像补全(Inpainting)方法综述

中国海洋大学 21220211109 张新宇

Email： 21220211109@stu.ouc.edu.cn

abstract

图像修复(Image Inpainting)，作为图像处理领域前沿的重要方向，各种前沿问题都受到研究者们广泛的关注，图像补全的目标是将图像中损坏的部分修复起来。该技术可以应用在图像编辑上，例如包括不规则形状遮挡、文本遮挡、目标遮挡、相片划痕等。传统的图像修复方法有 diffusion-based 和 patch-based 等方法，这些方法在某些情况下取得了不错的效果，而近些年来，随着深度学习的兴起，越来越多的基于深度学习的图像补全方法取得了极佳的效果，本文我将介绍深度学习在图像补全领域的发展脉络，并介绍几种具有代表性的基于深度学习的图像补全模型架构和算法。

1 Introduction

一直以来，计算机图形学和计算机视觉(Computer Vision) 是计算机科学领域中两个非常重要的研究方向。图形学领域的研究的可以总结为如何生成和处理图像，计算机视觉领域的研究可以总结为如何认识和深入理解图像。虽然这两个领域研究的问题有一些差距，但是由于都是对图像这个对象的领域进行研究，所以研究的相关性也同样紧密。本文介绍的问题，图像补全(image inpainting)，同样涉及了图形学和视觉的相关领域。

图像补全，就是将图片中缺失的像素区域补全，目的是使得陌生的观察者无法辨认出这其实是补全的图像。传统的方法可以简单的处理修补图像的局部失真等问题，但是无法有效的处理对某些全局信息失真的问题。而基于深度学习的图像补全方法则可以非常有效的解决这些问题。

本文将按照时间顺序，重点介绍具有代表性的基于深度学习的图像补全的工作。了解相关领域的研究者们，如何使用深度学习方法来解决这些棘手的问题。期待着我们可以通过对这一领域的发展历程来启发我们自己的思考，来启发自己对这一系列问题的研究。

2 Tranditional methods

2.1 diffusion-based method

这是一种将信息从缺失区域外逐步扩散到缺失区域的方法，采用迭代补全的方法，从缺失区域的边界一步步向里面补全，计算如下：

$$I^{n+1}(i,j) = I^{n}(i,j) + \Delta t I_t^n(i,j),\ \forall (i,j) \in \Omega$$

每一步迭代就将图像的 smooth 信息沿着等照度线向缺失区域内传播。这种将信息从缺失区域外逐步扩散到缺失区域的方法，对于较小的缺失区域，补全效果较好，但是对于较大面积的区域补全效果不理想。

2.2 PatchMatch method

如图所示，具体来说，就是从 **B** 图中寻找与 **A** 图 **Patch** 匹配的结果：

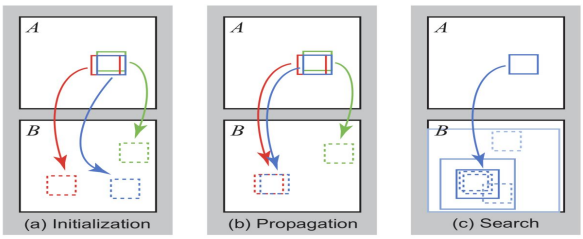


Figure 2: *Phases of the randomized nearest neighbor algorithm: (a) patches initially have random assignments; (b) the blue patch checks above/green and left/red neighbors to see if they will improve the blue mapping, propagating good matches; (c) the patch searches randomly for improvements in concentric neighborhoods.*

具体步骤：

1.随机映射；

2.相邻(上下左右)像素的 Patch 的匹配结果有相似性；

3.空间稍大范围的搜索操作。

在图片数据集内有一些相似的特征时 PatchMatch 这种方法才会取得有效的效果，但这种方法不能产生新的信息。

3 Deep learning methods

3.1 Context encoders: Feature learning by inpainting. (CVPR 2016)

文章中设计了 Encoder 与 Decoder 网络+对抗生成网络的网络架构实现了一定的图像修复功能，Encoder 与 Decoder 网络都是全卷积网络，并没有线性层存在，在 Encoder 的最后一层通过全连接实现信息传递，然后通过五个反卷积层来实现对图像需要修复区域的生成，网络结构如下：

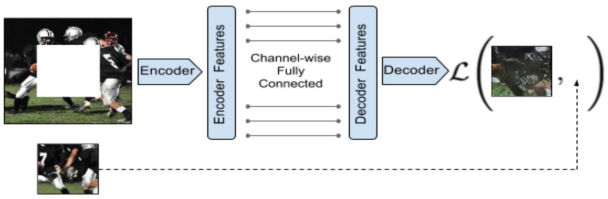


Figure 2: Context Encoder. The context image is passed through the encoder to obtain features which are connected to the decoder using channel-wise fully-connected layer as described in Section 3.1. The decoder then produces the missing regions in the image.

Encoder 的输入是带有 mask（即图像缺失）的图像，解码器输出的是已经将缺失区域补全的图像。再利用生成对抗网络的判别器思想去训练整个网络输出的补全图片。文中使用的损失函数是 L2 loss 和 adversarial loss，L2 loss 可以使得被遮挡区域里面的图像缺失像素信息被恢复,adversarial loss 则使得恢复的区域更加贴近原始未缺失的图像。

Reconstruction Loss：

$$L_{rec} = \|\hat{M} \odot (x - F((1-\hat{M}) \odot x))\|_2^2$$

其中 x 表示输入的图像，M 是遮挡的 mask，数值为 1 表示遮挡，0 表示不遮挡； 计算出来的损失为补全图像区域与真实图像对应区域的 L2 范数。

Adversarial Loss：

$$L_{adv} = \max_D \mathbb{E}_{x \in \chi}[\log(D(x)) + \log(1 - D(F((1-\hat{M}) \odot x)))]$$

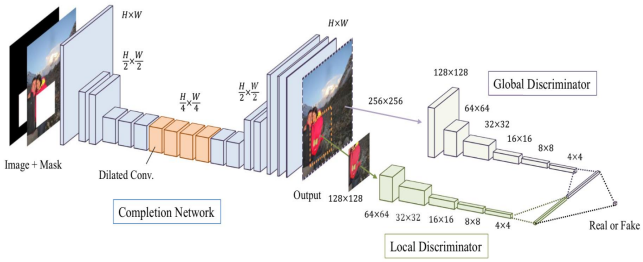
与生成对抗网络的 Loss 基本相同，但这里只训练判别器 Discriminator, 生成器(F)仅通过 Reconstruction Loss 训练，这里的判别器判断的是生成的整张图像。

模型修复效果图：

效果评价：本方法明显好于传统的基于 PatchMatch 的图像补全方法，但是对于细节方面，生成的补全区域效果仍然较差。

3.2 Globally and Locally Consistent Image Completion: Network (SIGGRAPH 2017)

文章中设计的模型架构由三个网络组成：补全网络，全局内容鉴别器和局部内容鉴别器。补全网络被设计成一个全卷积的神经网络用来补全图像，而全局和局部内容鉴别器则是用来专门用来训练判断补全后的图像和原始图像是否能保持一致的判别网络。全局鉴别器的输入为完整的图像，用来判断生成图像的全局一致性，而局部鉴别器则通过生成的图像补全区域周围的小区域来判断更细节的补全区域的质量。在每次训练迭代过程中，首先更新鉴别器，使它们能正确地区分真实和补全的训练图像。之后更新图像补全网络，使它能有效的补全缺失区域以欺骗内容鉴别器网络。网络结构如下图所示：



损失函数

MSE Loss：

$$L(x, M_c) = \|M_c \odot (C(x, M_c) - x)\|^2$$

MSE Loss：

$$\min_C \max_D \mathbb{E}[\log D(x, M_d) + \log(1 - D(C(x, M_c), M_c))]$$

其中，x 是输入的有遮挡的图像，M_c 是遮挡的 mask 区域，C 表示设计的补全网络，损失函数计算的实际上是补全区域的 L2 范数。

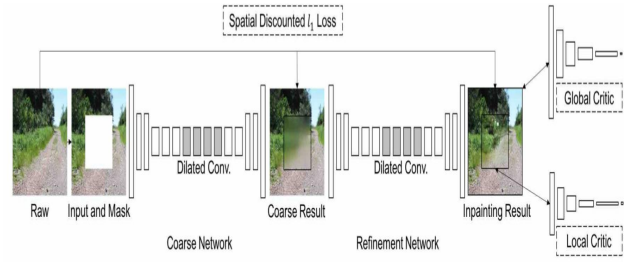
修复效果图：

Fig. 16. Failure cases for our approach where our model is unable to complete heavily structured objects such as people and animals. Photographs courtesy Pete (Public Domain), and brad pierce (Public Domain).

评价：大部分情况下的补全效果都不错，但是对于结构比较强的图片补全效果不好，如上图，这也是很多通用补全网络的通病。

3.3 Generative Image Inpainting with Contextual Attention: Network (CVPR 2018)

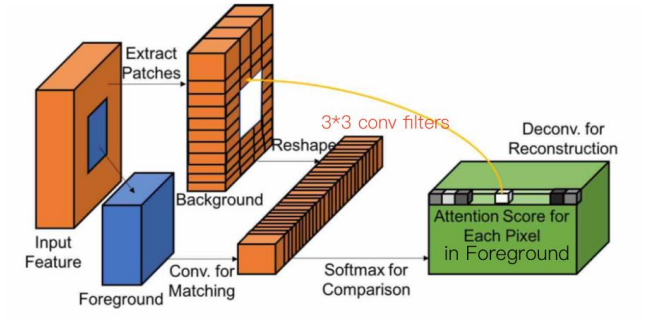
文章设计了一种深度生成模型，能够用来生成新颖的图像，而且可以在网络训练过程中使用修复区域周围的图像像素特征，作为完整图像修复的参考，进而生成更好的补全图像。该网络同样是完全卷积的神经网络，可以修复任意缺失位置区域和形状区域的图像。
网络架构如下图所示：



整个网络架构的图像修复过程包括两个阶段，第一阶段是一个 dilated 卷积神经网络，使用 reconstruction loss 训练，实现对缺失图像粗略的修复；第二阶段使用提出的 contextual attention layer 方法来完成粗略的修复图像到精细修复图像的修复。
contextual attention layer 方法的工作原理：文章使用非缺失 Patches 的特征作为卷积核来优化生成的 Patches，最终使模糊的修复结果变得精细。主要的做法原理是将生成的 patches 与非缺失 patches 进行卷积匹配，沿着通道维度进行 channel-wise softmax 来权衡两张图像的相关性，最后利用相关性进行反卷积得到精细的图像修复结果。
contextual attention layer 也采用了 spatial propagation layer 来提高图像修复的空间一致性。
网络架构的输入：带有区域缺失的 256×256 的图像 + 二进制的 Mask 遮挡。

整个网络采用窄而深的架构，卷积层使用镜像填充，去除 batch normalization layers，使用 ELU 激活函数，来替代激活函数 tanh 或 sigmoid。
粗修复网络仅使用 reconstruction loss，而精细修复网络采用 reconstruction loss+ 两个 GAN loss。
Loss 损失：
Coarse Result 及 Refinement Result 都与真实图片有 L1 损失
Global 与 Local 对抗损失，采用 WGAN-GP。

Match and attend（匹配和选取）：
文章首先考虑的问题是将图像缺失部分的特征与图像非缺失部分的特征如何进行匹配的问题。
利用 Refinement Network 可以找到图片中距离较远的强相似性的 patch，并且使用卷积的方法计算相似度。
如下图所示：



文章贡献:
创新性的提出了上下文注意力层，使网络可以从距离补全位置较远的区域提取和需要修复图像区域相似的特征；并且介绍了提高训练的速度和稳定性的方法。

修复效果图示：

3.4 Image Inpainting for Irregular Holes Using Partial Convolutions (ECCV 2018)

首先文章指出，大部分研究在进行图片补全操作的图像处理时，通常使用白色或者随机噪声这种没有语义的信息对图像缺失区域进行填充，再使用卷积操作来提取图像特征再进行后续的补全。文章认为对这些不存在语义信息的内容与有效的信息整体进行直接的卷积操作会存在问题，例如下图：

(a) Image with hole (b) PatchMatch (c) Iizuka et al.[1] (d) Yu et al.[2]

于是文章提出了新的卷积方式 Partial Convolution(部分卷积): 只对图像中有效的信息进行卷积，卷积的时候还需要 mask 的信息：

$$x' = \begin{cases} \mathbf{W}^T(\mathbf{X} \odot \mathbf{M}) \frac{1}{\text{sum}(\mathbf{M})} + b, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

卷积后 Feature Map 的长宽会变小，Mask 需要随之更新：

$$m' = \begin{cases} 1, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

网络架构类似于 Unet 网络结构，在图像编码阶段使用 ReLU 函数进行激活，解码阶段使用 LeakyReLU 函数进行激活，批量归一化层在中间层使用，而不在初始和最后层使用。在处理 Padding 时，将 Padding 视作缺失孔洞处理。

本文首次证明了在不规则形状的孔洞上训练图像修复网络的有效性。同时，作者提出了一个大型的不规则 mask 数据集。

修复效果图：

对于复杂的不规则遮挡也有比较好的补全效果

3.5 Free-Form Image Inpainting with Gated Convolution (ICCV2019)

文章提出部分卷积存在的问题：Partial Convolutions 中 Mask 更新的不合理；进而提出新卷积层 Gated Convolution layer，通过为各层中每个空间位置的每个通道提供一种可学习的动态特征选择机制，解决了传统卷积将所有输入像素都视为有效像素的不合理性；提出 SN-PatchGAN loss 来优化图像修复准确性。
文章改进了 Partial Conv 方法的 Mask 更新，不再固定为 0,1 而是取 [0,1] 之间，Gated Convolution 方法和 Partial Convolutions 对比见下图：

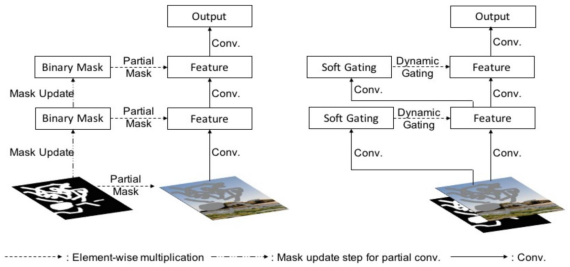
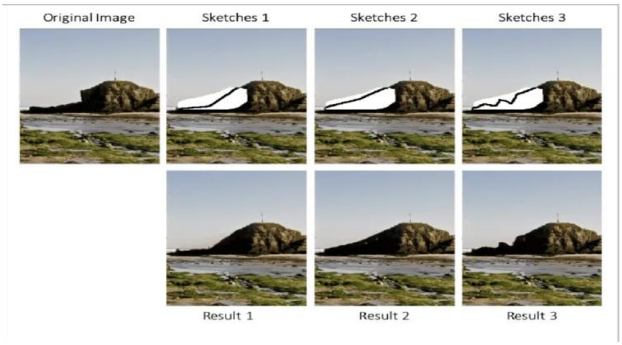


Fig. 2. Illustration of partial convolution (left) and gated convolution (right).

门控卷积并没有采用 hard-gating mask 的 Mask 更新规则，而是采用新的可学习的 soft mask 更新规则。
gated convolution 公式如下：

$$Gating_{y,x} = \sum\sum W_g \cdot I$$
$$Feature_{y,x} = \sum\sum W_f \cdot I$$
$$O_{y,x} = \phi(Feature_{y,x}) \odot \sigma(Gating_{y,x})$$

模型还有与用户的交互功能，网络的输入包含一个与用户交互的 sketch 通道，sketch 通道由真实图像通过边缘检测和阈值分割得到，用户可以在该 sketch 通道绘制引导线，引导修复后图像的走向。在训练的时候，输入为破损图像、mask、sketch。
具体效果见下图所示：



3.6 RePaint: Inpainting using Denoising Diffusion Probabilistic Models (CVPR 2022)

文章提出了使用最近大火的扩散模型实现图像修复工作。将 **Denoising Diffusion Probabilistic Models**（**DDPM**）应用于图像修复工作。以 masked 图像作为输入。它从随机噪声样本开始，迭代去噪，直至产生高质量的输出。由于这个过程是随机，可以得到多种不同的输出样本。并且 DDPM 先验强制协调图像，所以能够从其他区域再现纹理，并修复语义上有意义的内容。

文章指出了目前图像修复面临的两个主要问题，一是用单一类型的 mask 训练限制了模型的泛化能力，二是 pixel-wise 和 perceptual loss 会导致生成模型朝着纹理填充而不是语义修复方向更新。

本文提出了基于去噪扩散概率模型的图像修复方法：**Repaint**，该方法理论上可以使用任何形状的遮掩的 mask 图像来修复，甚至对于极端的 mask 情况（mask 面积很大，几乎遮挡了整幅图像）都适用。我们利用一个与训练的无条件的 DDPM 作为一个生成先验模型。在扩散过程中，文章仅在已知图像区域进行采样来改变逆扩散过程的生成。并且该方法完全不需要调整已经训练好的 DDPM 网络，因此该方法理论上可以在任何图像修复问题上实现良好的效果，生成高质量图像。

由于扩散模型的前向过程通过添加的高斯噪声的马尔可夫链来定义，文章通过定义在任意点上采样中间图像 **x_t**。这使得在任意时间步 **t** 采样已知区域 **m⊙x_t**。因此，通过处理未知区域和已知区域，并最终合并成完整的图片，得到了如下所示的反转步的表达式：

$$
\begin{aligned}
x_{t-1}^{\text{known}} &\sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}) \\
x_{t-1}^{\text{unknown}} &\sim \mathcal{N}(\mu_\theta(x_t,t), \Sigma_\theta(x_t,t)) \\
x_{t-1} &= m \odot x_{t-1}^{\text{known}} + (1-m) \odot x_{t-1}^{\text{unknown}}
\end{aligned}
$$

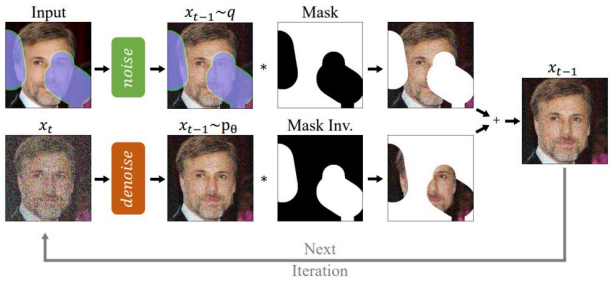
使用给定图像 m⊙x_0 中的已知像素对 t-1 时刻的已知图像分布进行采样，同时在训练好的 DDPM 迭代 x_t 时得到 x_t-1 时刻的未知区域。将两个区域合并起来进行下一次迭代。

具体的算法伪代码如下图所示：

Algorithm 1 Inpainting using our RePaint approach.

$x_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \dots, 1$ **do**
  for $u = 1, \dots, U$ **do**
    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\epsilon = \mathbf{0}$
    $x_{t-1}^{\text{known}} = \sqrt{\bar{\alpha}_t}x_0 + (1-\bar{\alpha}_t)\epsilon$
    $z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
    $x_{t-1}^{\text{unknown}} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(x_t, t)\right) + \sigma_t z$
    $x_{t-1} = m \odot x_{t-1}^{\text{known}} + (1-m) \odot x_{t-1}^{\text{unknown}}$
    if $u < U$ and $t > 1$ **then**
      $x_t \sim \mathcal{N}(\sqrt{1-\beta_{t-1}}x_{t-1}, \beta_{t-1}\mathbf{I})$
    end if
  end for
end for
return x_0

RePaint 方法的概览如下图所示：



同时文章提出了一个新的问题：

当直接应用上述方法时，只有内容类型与已知区域匹配。生成的遮掩区域图像，与原始图像未遮掩区域理论上毫无关联。比如，在下图 中，当 n 为 1 时，图像修复的区域是与原始输入图像皮毛相匹配的皮毛纹理。尽管图像修复的区域与邻近区域的纹理相匹配，但是却在语义信息上完全不相关。因此，虽然 DDPM 利用了已知区域的上下文，但它并没有很好地协调生成的补全图像的相关性。

所以作者为解决这种问题，提出了增加生成图像区域与原始图像之间的约束，即生成的遮掩区域图像需要通过重新添加噪声还原回上个采样时刻的原始图像。

具体的修复效果如图所示：

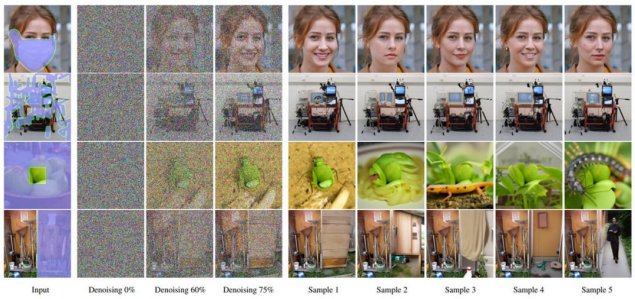


Figure 1. **We use Denoising Diffusion Probabilistic Models (DDPM) for inpainting.** The process is conditioned on the masked input (*left*). It starts from a random Gaussian noise sample that is iteratively denoised until it produces a high-quality output. Since this process is stochastic, we can sample multiple diverse outputs. The DDPM prior forces a harmonized image, is able to reproduce texture from other regions, and inpaint semantically meaningful content.

取得的效果：对于六种掩模分布中的至少五种，repaint 优于最先进的 Autoregressive 和 GAN 方法。

CelebA-HQ	Wide		Narrow		Super-Resolve 2×		Altern. Lines		Half		Expand	
Methods	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]
AOT [44]	0.104	11.6 ± 2.0	0.047	12.8 ± 2.1	0.714	1.1 ± 0.6	0.667	2.4 ± 1.0	0.287	9.0 ± 1.8	0.604	8.3 ± 1.7
DSI [27]	0.067	16.0 ± 2.3	0.038	22.3 ± 2.6	0.128	5.5 ± 1.4	0.049	5.1 ± 1.4	0.211	4.5 ± 1.3	0.487	4.7 ± 1.3
ICT [35]	0.063	27.6 ± 2.8	0.036	30.9 ± 2.9	0.483	4.2 ± 1.2	0.353	0.7 ± 0.5	0.166	12.7 ± 2.1	0.432	8.8 ± 1.8
DeepFillv2 [40]	0.066	23.9 ± 2.6	0.049	21.0 ± 2.5	0.119	9.8 ± 1.8	0.049	10.6 ± 1.9	0.209	4.1 ± 1.2	0.467	13.1 ± 2.1
LaMa [33]	0.045	41.8 ± 3.1	0.028	33.8 ± 3.0	0.177	5.5 ± 1.4	0.083	20.6 ± 2.5	0.138	35.6 ± 3.0	0.342	24.7 ± 2.7
RePaint	0.059	*Reference*	0.028	*Reference*	0.029	*Reference*	0.009	*Reference*	0.165	*Reference*	0.435	*Reference*

ImageNet	Wide		Narrow		Super-Resolve 2×		Altern. Lines		Half		Expand	
Methods	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]	LPIPS↓	Votes [%]
DSI [27]	0.117	31.7 ± 2.9	0.072	28.6 ± 2.8	0.153	26.9 ± 2.8	0.069	23.6 ± 2.6	0.283	31.4 ± 2.9	0.583	9.2 ± 1.8
ICT [35]	0.107	42.9 ± 3.1	0.073	33.0 ± 2.9	0.708	1.1 ± 0.6	0.620	6.6 ± 1.5	0.255	51.5 ± 3.1	0.544	25.6 ± 2.7
LaMa [33]	0.105	42.4 ± 3.1	0.061	33.6 ± 2.9	0.272	13.0 ± 2.1	0.121	9.6 ± 1.8	0.254	41.1 ± 3.1	0.534	20.3 ± 2.5
RePaint	0.134	*Reference*	0.064	*Reference*	0.183	*Reference*	0.089	*Reference*	0.304	*Reference*	0.629	*Reference*

Table 1. **CelebA-HQ (*top*) and ImageNet (*bottom*) Quantitative Results.** Comparison against the state-of-the-art methods. We compute the LPIPS (lower is better) and *Votes* for six different mask settings. *Votes* refers to the ratio of votes with respect to ours.

主要贡献：
文章的工作为解决图像修复目前的局限性开辟了有趣的研究方向。
待改进的地方：
每个图像的 DDPM 优化过程明显慢于基于 GAN 和自回归模型的对应过程。这使得模型难以应用于实时应用的领域。
对于极端的 mask 情况，RePaint 可以生成与 GT 图像非常不同的图像。

4 Discussion and conclusions

本文介绍了关于图像修补领域顶尖期刊的一部分重点模型和算法，梳理了近年来图像修补领域研究的历史研究脉络，基于深度学习的图像修补的效果也越来越好，我相信在未来，基于深度学习的图像修补方法还会有很大的进步空间，仍然具有很好的前景。

5 References

[1] Bertalmio, Marcelo, et al. 2000. Image inpainting. In SIGGRAPH

[2] Bertalmio, Marcelo, et al. 2000. Image inpainting. In SIGGRAPH

[3] Barnes, Connelly, et al. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. In ToG

[4] Pathak, Deepak, et al. 2016. Context encoders: Feature learning by inpainting. In CVPR

[5] Iizuka, Satoshi, Edgar Simo-Serra, and Hiroshi Ishikawa. 2017. Globally and locally consistent image completion. In SIGGRAPH

[6] Yu, Jiahui, et al. 2018. Generative image inpainting with contextual attention. In CVPR

[7] Liu, Guilin, et al. 2018. Image inpainting for irregular holes using partial convolutions. In ECCV.

[8] Andreas Martin Danelljan. 2022. RePaint: Inpainting using Denoising Diffusion Probabilistic Models In CVPR

[9] Yu Jiahui et al. 2018. Free-From Image Inpainting with Gated Convolution.arvix preprint arXiv:1806.03589 ICCV 2019