

基于傅里叶卷积的分辨率鲁棒大掩模修复

摘要

现代图像修复系统虽然取得了重大进展，但往往难以修复缺失区域较大、几何结构复杂和高分辨率的图像，这主要是由于修复网络和损失函数中缺乏有效感受野。为了解决这个问题，原论文提出了一种名为大掩模修复 (LaMa) 的新方法。LaMa 基于 i) 一种新的修复网络架构，它使用具有图像范围感受野的快速傅里叶卷积 (FFCs)；ii) 高感受野感知损失；iii) 大型训练掩模。其修复网络在一系列数据集上改进了最先进的技术，即使在具有挑战性的场景 (如周期结构的完成) 中也取得了良好的性能。该模型可以很好地推广到分辨率高于训练数据的图像，并在比 baseline 更低的参数和时间成本下实现这一点。

关键词：图像修复；快速傅里叶卷积；高感受野感知损失；大面积掩模生成

1 引言

图像修复问题的解决方案——即如何对缺失部分进行尽量真实的填充——既需要让模型“理解”自然图像的大规模结构，也需要进行图像合成。深度学习时代，该课题已进行了大量研究 [1]，近年来，深度神经网络和对抗性学习的出现加速了研究的进展。图像修复通常的做法是在大型自动生成的数据集上训练修复系统，该数据集是通过随机对真实图像建立掩模创建的。修复模型通常使用具有中间预测的复杂两阶段模型，如平滑图像 [12]、边缘 [16] 和分割图 [24]。而原论文 [25] 使用一个简单的单级网络实现了最先进的结果。一个大的有效感受野 [14] 对于理解图像的全局结构以及解决修复问题至关重要，但在掩模较大的情况下，一个有限的感受野能够访问的信息不足以生成高质量修复。作者注意到流行的卷积架构可能缺乏足够大的有效感受野，于是对系统的每个组件进行调整以缓解问题并解锁单阶段解决方案的潜力。具体来说：

1) LaMa 提出了一种基于快速傅里叶卷积 (FFCs) [13] 的修复网络。FFC 允许网络的早期层中就拥有能覆盖整个图像的感受野。

2) LaMa 使用基于具有高感受野的语义分割网络的感知损失 [10]，因为感受野不足不仅会损害修复网络，还会损害感知损失。

3) LaMa 引入了一种新的训练掩模生成策略，以发挥 FFCs 和感知损失在高感受野中的潜力。该过程产生宽大掩模，迫使网络充分利用模型的高感受野和损失函数。

在仅对低分辨率数据进行训练后，LaMa 可以推广到高分辨率图像，并且可以捕获和生成复杂的周期结构，并且在缺失区域较大的情况下修复质量稳定。

本文基于 LaMa 工作，对损失函数进行调整，实现了较好的修复效果。

2 相关工作

早期的图像修复数据驱动方法依赖于基于补丁的 [3,8,32] 和基于最近邻的 [6] 生成。深度学习时代的第一个修复工作之一是使用带有以对抗方式训练的编码器-解码器架构的卷积网络 [26]。迄今为止，这种方法仍然通常用于深度修复。另一个流行的完成网络选择组是基于 U-Net [19] 的体系结构。

图像修复中一个常见的问题是网络掌握局部和全局上下文的能力。为此，[8] 提出合并扩张卷积 [32] 来扩展感受野，其中的两个鉴别器分别发展全局和局部一致性。在 [28] 中，建议在具有不同接收域的完成网络中使用分支。为了从空间上遥远的补丁中借用信息，[33] 提出了上下文注意层。LaMa 证实了信息在远距离位置之间有效传播的重要性。此方法的一个变体在很大程度上依赖于扩张卷积块，灵感来自 [23]。作为更好的替代方案，我们提出了一种基于频域变换的机制 (FFC) [13]。这也与最近的在计算机视觉中使用变形的趋势一致 [4,5]，并将傅里叶变换视为自我注意的轻量级替换。

在更全局的层面上，[33] 引入了一个从粗到细的框架，该框架涉及两个网络，第一个网络在孔中完成粗略的全局结构，然后在第二个网络中使用它作为指导来细化局部细节。这种遵循结构-纹理分解 [1] 相对古老思想的两阶段方法在后续工作中变得普遍。一些研究 [22] 修改了框架，以便同时而不是顺序地获得粗略和精细的结果组件。一些工作提出了两阶段方法，使用完成其他结构类型作为中间步骤：[16] 中的显着边、[24] 中的语义分割图、[29] 中的前景对象轮廓、[30] 中的梯度图和 [18] 中的边缘保留平滑图像。与所有这些工作相比，LaMa 证明了精心设计的单阶段方法可以实现非常强大的结果。

为了处理不规则的掩码，一些研究修改了卷积层，引入了部分 [11]、门控 [33]、轻量级门控 [31] 和区域 [15] 卷积。探索了各种形状的训练掩模，包括随机 [8]、自由形式 [33] 和物体形状的掩模 [31]。我们发现，只要训练掩码的轮廓足够多样化，掩码生成的确切方式不如掩码的宽度重要。

为了训练修复网络，研究人员提出了许多损失函数。通常，使用像素级（例如 l_1 、 l_2 ）和对抗性损失。一些方法对像素级损失应用空间折扣加权策略 [17,33]。简单的卷积鉴别器 [17,30] 或 PatchGAN 鉴别器 [8] 被用来实现对抗损失。而 LaMa 使用了一个 r_1 梯度惩罚补丁鉴别器。感知损失也经常被应用，通常使用在 ImageNet 分类 [21] 上预训练的 VGG-16 [11] 或 VGG-19 [24] 骨干。这种感知损失对于图像修复是次优的，与这些工作相比，LaMa 提出了一种更好的替代方案，并将特征匹配损失用于模型中。

3 本文方法

3.1 本文方法概述

论文的目标是填充由未知像素 m 的二进制掩码屏蔽的彩色图像 x ，掩码图像表示为 $x \odot m$ 。掩码 m 与掩码图像 $x \odot m$ 堆叠，从而产生四通道输入张量 $x = \text{stack}(x \odot m, m)$ 。论文使用前馈网络 $f_\theta(\cdot)$ ，也称为生成器。以 x' 为例，修复网络以完全卷积的方式处理输入，并产生内嵌的三通道彩色图像 $\hat{x} = f_\theta(x')$ 。训练是在从真实图像和合成生成的掩码获得的（图像、掩码）对的数据集上执行的。

3.2 早期层中的全局上下文

在填充大缺失区域这样具有挑战性的情况下，正确修复的生成需要考虑全局上下文。因此，论文认为一个好的架构应该尽早拥有尽可能宽的感受野的单位。传统的全卷积模型，例如 ResNet [7]，存在有效感受野增长缓慢 [14]。由于卷积核通常很小（例如 3×3 ），感受野可能不够，尤其是在网络的早期层中。因此，网络中的许多层将缺乏全局上下文，并将浪费计算和参数来创建一个。对于宽掩码，生成器在特定位置的整个感受野可能位于掩码内，因此仅观察缺失的像素。对于高分辨率图像，这个问题变得尤为明显。

快速傅里叶卷积 (FFC) [13] 是最近提出的允许在早期层使用全局上下文的算子。FFC 基于通道快速傅立叶变换 (FFT) [2]，并且具有覆盖整个图像的接受场。FFC 将通道分成两个并行分支：i) 本地分支使用传统的卷积，以及 ii) 全局分支使用真实 FFT 来解释全局上下文。实 FFT 只能应用于实值信号，逆实 FFT 保证了输出是实值的。与 FFT 相比，真正的 FFT 只使用一半的频谱。具体来说，FFC 执行以下步骤：

- (a) 将 *RealFFT2d* 应用于输入张量

$$\text{RealFFT2d} : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{C}^{H \times \frac{W}{2} \times C},$$

并连接实部和虚部

$$\text{ComplexToReal} : \mathbb{C}^{H \times \frac{W}{2} \times C} \rightarrow \mathbb{R}^{H \times \frac{W}{2} \times 2C};$$

- (b) 在频域应用卷积块

$$\text{ReLU} \circ \text{BN} \circ \text{Conv} I \times I : \mathbb{R}^{H \times \frac{W}{2} \times 2C} \rightarrow \mathbb{R}^{H \times \frac{W}{2} \times 2C};$$

- (c) 应用逆变换来恢复空间结构

$$\text{RealToComplex} : \mathbb{R}^{H \times \frac{W}{2} \times 2C} \rightarrow \mathbb{C}^{H \times \frac{W}{2} \times C},$$

$$\text{InverseRealFFT2d} : \mathbb{C}^{H \times \frac{W}{2} \times C} \rightarrow \mathbb{R}^{H \times W \times C},$$

最后，将局部和全局分支的输出融合在一起。FFC 的说明如图1所示。

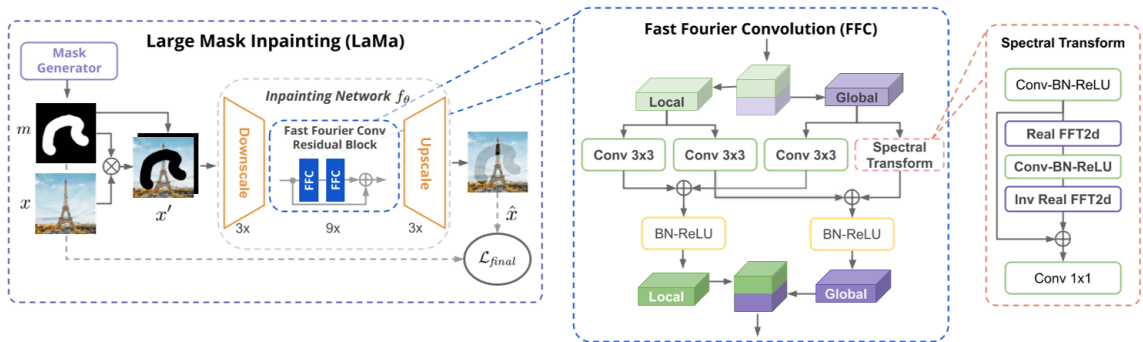


图 1. LaMa 方案示意图。LaMa 基于前馈 ResNetlike 修复网络，该网络使用最近提出的快速傅里叶卷积 (FFC)，一种结合了对抗性损失和高感受野感知损失的多分量损失，以及训练时间大掩码生成过程。

快速傅里叶卷积的作用：FFC 是完全可微的，易于使用的下降来代替传统的卷积。由于图像范围的感受野，FFCs 允许生成器从早期层开始考虑全局上下文，这对于高分辨率图像修复至关重要。这也让效率更高，可训练参数可用于推理和生成，而不是“等待”用于信息传播。

3.3 损失函数定义

修复问题本质上是模棱两可的。对于相同的缺失区域，可能有许多合理的填充，尤其是当“洞”变得更宽时。

3.3.1 高感受野感知损失

监督损失要求生成器精确地重建基本事实。然而，图像的可见部分往往不包含足够的信息来精确重建掩码部分。因此，由于对内嵌内容的多种合理模式进行平均，使用朴素监督会导致结果模糊。

相比之下，感知损失 [10] 通过基础预训练网络 $\Phi(\cdot)$ 评估从预测图像和目标图像中提取的特征之间的距离。它不需要精确的重建，允许重建图像的变化。大掩码修复的重点是转向对全局结构的理解。因此，我们认为使用接受域快速增长的基础网络是很重要的。论文引入了高感受野感知损失 ($HRFPL$)，它使用高感受野基础模型 $\Phi_{HRF}(\cdot)$ ：

$$L_{HRFPL}(x, \hat{x}) = M(([\Phi_{HRF}(x) - \Phi_{HRF}(\hat{x})])^2) \quad (1)$$

其中 $[\cdot - \cdot]^2$ 是逐元素操作， M 是顺序两阶段平均操作（层内均值的层间均值）。 $\Phi_{HRF}(x)$ 可以使用傅里叶或扩张卷积来实现。HRF 感知损失对 LaMa 来说至关重要。

3.3.2 对抗性损失

论文使用对抗性损失来确保修复模型 $f_\theta(x')$ 生成自然查看局部细节。它定义了一个鉴别器 $D_\xi(\cdot)$ ，它适用于局部补丁级 [9]，区分“真实”和“假”补丁，只有与掩码区域相交的补丁才会得到“假”标签。由于有监督的 HRF 感知损失，生成器快速学习复制输入图像的已知部分，因此将生成的图像的已知部分标记为“真实”。最后，使用非饱和对抗性损失：

$$L_D = -\mathbb{E}_x [\log D_\xi(x)] - \mathbb{E}_{x,m} [\log D_\xi(\hat{x}) \odot m] - \mathbb{E}_{x,m} [\log (1 - D_\xi(\hat{x})) \odot (1 - m)] \quad (2)$$

$$L_G = -\mathbb{E}_{x,m} [\log D_\xi(\hat{x})] \quad (3)$$

$$L_{Adv} = sg_\theta(L_D) + sg_\xi(L_G) \rightarrow \min_{\theta, \xi} \quad (4)$$

其中 x 是数据集中的样本， m 是合成生成的掩码， $x = f_\theta(x)$ 是 $x = stack(x \odot m, m)$ 的修复结果， sg_{var} 停止梯度 w.r.t var ， L_{Adv} 是优化的联合损失。

3.3.3 最终的损失函数

在最终损失中，论文还使用 $R1 = E_x(\|\nabla D_\xi(x)\|)^2$ 梯度惩罚 [20] 和基于鉴别器的感知损失或所谓的特征匹配损失——鉴别器网络 L_{DiscPL} [27] 特征的感知损失。众所周知，用 L_{DiscPL} 来稳定训练在某些情况下会略微提高性能。

修复模型的最终损失函数：

$$L_{funal} = \kappa L_{Adv} + \alpha L_{HRFPL} + \beta L_{DiscPL} + \gamma R_1 \quad (5)$$

是讨论损失的加权和，其中 L_{Adv} 和 L_{DiscPL} 负责生成自然查看局部细节，而 L_{HRFPL} 负责全局结构的监督信号和一致性。

3.4 在训练期间生成掩码

模型的最后一个组件是一个掩码生成策略。每个训练示例 x' 是来自合成生成的掩码叠加的训练数据集的真实照片。与数据增强对最终性能的影响很大的判别模型类似，掩码生成的策略显着影响修复系统的性能。

因此，论文选择了激进的大型掩码生成策略。该策略统一使用来自高随机宽度（宽掩码）和任意纵横比（框掩码）矩形扩展的多边形链的样本。

我们针对几种方法的窄掩码训练测试了大型掩码训练，发现使用大掩码策略进行训练通常会提高窄掩码和宽掩码的性能。

4 复现细节

4.1 与已有开源代码对比

使用了 LaMa 的 github 给出的代码，链接 <https://github.com/advimman/lama>。对模型进行分析后我认为它的损失函数部分还有可以改进的地方，所以将损失函数中的双线性插值改为了双三次插值。

作者将所有模型打包放到了包 saicinpainting 中，在其中的 losses-training 文件夹下 feature_matching 文件中将求双线性插值的部分改为求双三次插值，如2所示。

```
#cur_mask = F.interpolate(mask, size=fake_feat.shape[-2:], mode='bilinear', align_corners=False)
cur_mask = F.interpolate(mask, size=fake_feat.shape[-2:], mode='bicubic', align_corners=False)
```

图 2. 修改损失函数

4.2 实验环境搭建

作者在 github 项目中给出了 yaml 文件，安装 tiny-conda 后使用 conda 创建环境即可。

但此项目需要使用 tensorflow，需要提前安装。并需于创建环境后安装 python-lightning，用于在训练模型时尽可能多地使用 GPU 多线程进行训练。

4.3 使用说明

数据集中需要包含训练集、验证集和测试集。首先需要对验证集和测试集的数据通过 `get_gen_mask.py` 求随机掩码。接着通过 `train.py` 对训练集图片进行训练，得出模型存储于 `model` 当中。然后使用 `predict.py`，并设置模型路径、测试集路径以及输出路径，测试模型的修复效果。最后使用 `evaluate_predicts.py` 对生成的修复图片进行评估即可。

4.4 创新点

原文中对掩码进行双线性插值，使用插值后的遮罩 `cur_mask` 用于加权计算特征图之间的误差。在这里，误差权重是由 $1 - \text{cur_mask}$ 计算得到的，即遮罩值越大（接近 1），权重越小，表示该位置的信息是缺失的。这样，通过双线性插值，代码实现了在特征匹配损失中对缺失信息的考虑。

我使用双三次插值替代双线性插值，考虑了更多的相邻像素，通过一个更复杂的插值核心，得到了更平滑的插值结果。

5 实验结果分析

使用 `get_gen_mask.py` 求出的随机掩码如图3所示。

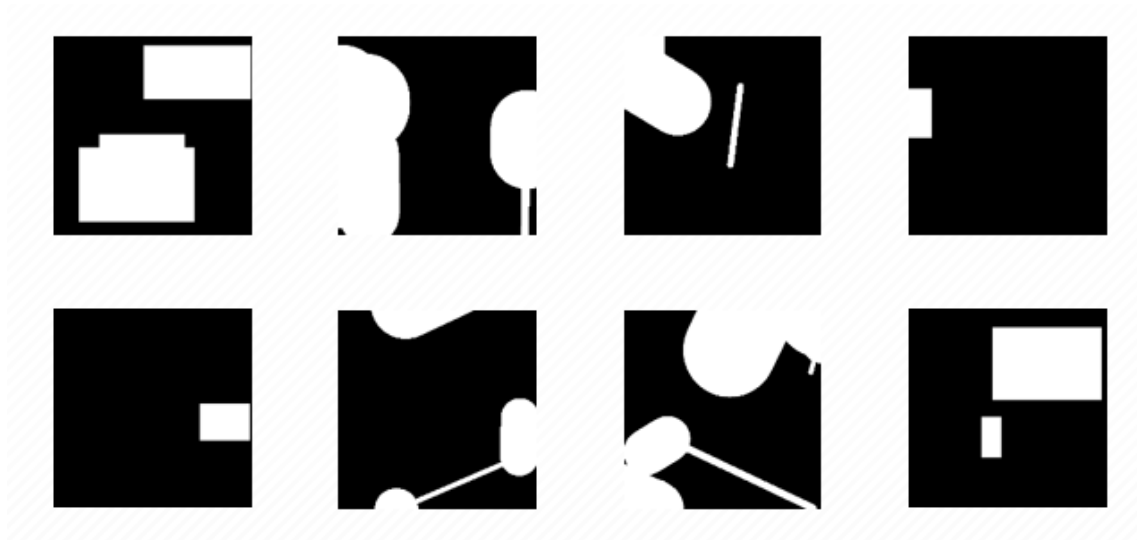


图 3. 生成的随机掩码

根据原文给出的精细掩码以及已经训练好的模型得出如图4所示的结果。

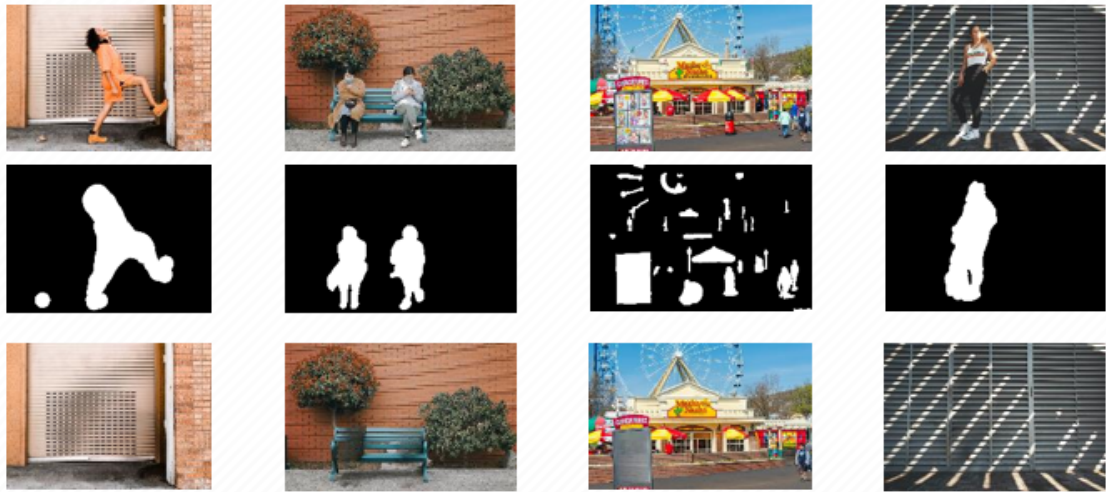


图 4. 根据已给出的测试图集，得出的测试结果

在 Places365 数据集上训练模型，并对分割出 2000 张图片进行测试，效果如图5所示。

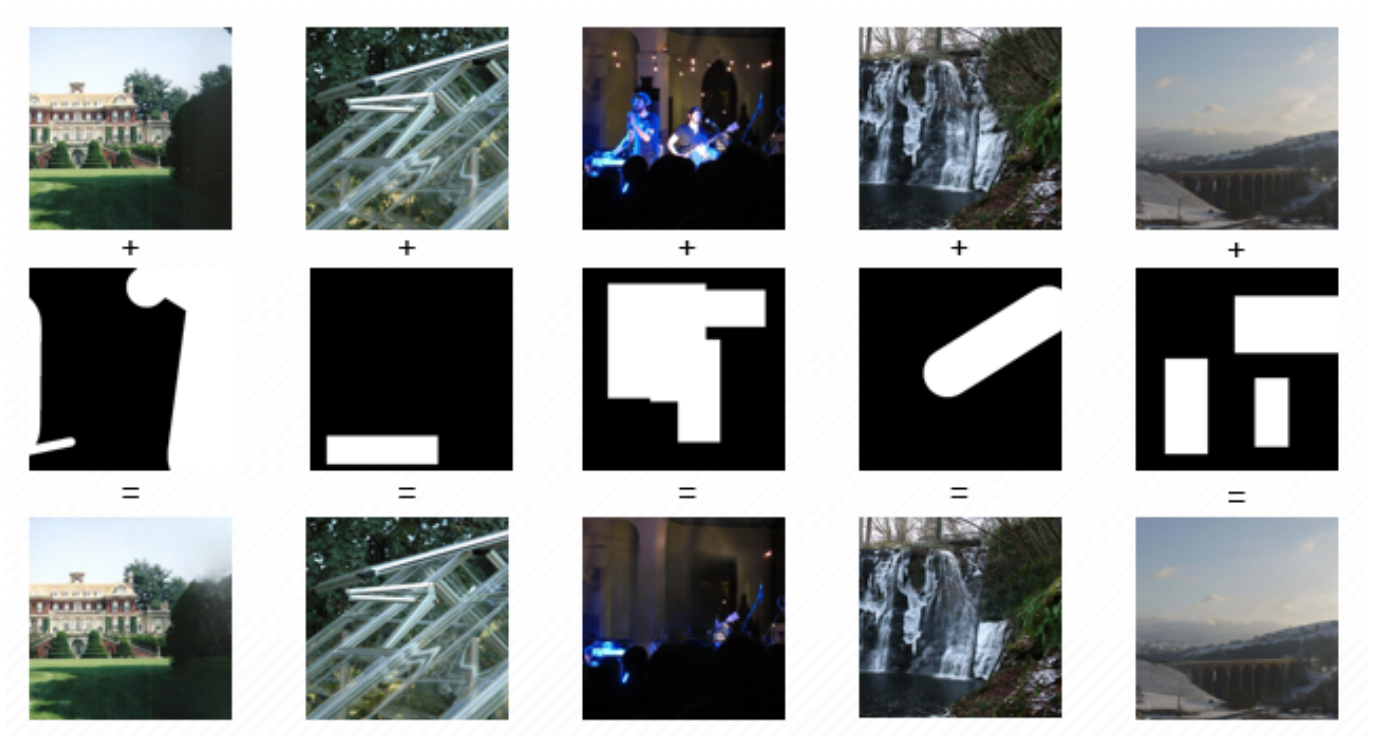


图 5. 使用训练模型得出的修复结果

对训练好的模型进行评估，将得出的评估与原文模型进行对比，如表1所示。效果不如原文模型，我分析推测是因为我的模型训练集较小。训练模型时由于训练集过大训练时间较长，我将训练集减少到原本的五分之一，这也许是性能下降的原因。

6 总结与展望

在本研究中，我们尝试通过某种方法训练模型以达到原文中的效果。然而，由于硬件及时间的限制，我们未能成功训练出与原文相同效果的模型。

模型（均为窄掩码情况）	LPIPS(↓)	FTD(↓)
原文	0.60	0.089
我的	0.74	0.097

表 1. 三行四列的表格

在实验过程中，我们遇到了一些挑战，包括资源有限以及训练时间过长。这些因素共同影响了我们对模型的调优和优化。尽管我们取得了一些进展，但模型性能仍未达到预期水平。

未来的工作将集中在改进模型的训练过程，以提高模型的性能。我们计划通过寻找更有效的模型架构、优化超参数以及使用更强大的硬件资源来解决当前面临的问题。此外，我们也将考虑使用更先进的训练技术和数据增强方法，以进一步提升模型的泛化能力。

总体而言，尽管当前的结果不尽如人意，但这些挑战将为未来的研究提供宝贵的经验。我们将持续努力改进并追求更好的性能，以便更好地满足实际应用的需求。

参考文献

- [1] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, Dec 2003.
- [2] E. O. Brigham and R. E. Morrow. The fast fourier transform. *IEEE Spectrum*, page 63–70.
- [3] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, Nov 2003.
- [4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition*, Oct 2020.
- [5] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2021.
- [6] James Hays and Alexei A. Efros. Scene completion using millions of photographs. *Communications of the ACM*, page 87–94, Oct 2008.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

- [8] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics*, page 1–14, Aug 2017.
- [9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.
- [10] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*, page 694–711. Jan 2016.
- [11] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. *Image Inpainting for Irregular Holes Using Partial Convolutions*, page 89–105. Jan 2018.
- [12] Hongyu Liu, Bin Jiang, Yibing Song, Wei Huang, and Chao Yang. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations.
- [13] Chenyang Lu, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Neural Information Processing Systems, Neural Information Processing Systems*, Jan 2020.
- [14] Wenjie Luo, Yujia Li, Raquel Urtasun, and RichardS. Zemel. Understanding the effective receptive field in deep convolutional neural networks. *arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition*, Jan 2017.
- [15] Yuqing Ma, Xianglong Liu, Shihao Bai, Lei Wang, Aishan Liu, Dacheng Tao, and Edwin R. Hancock. Regionwise generative adversarial image inpainting for large missing areas. *IEEE Transactions on Cybernetics*, page 1–14, Jan 2022.
- [16] Kamyar Nazeri, Eric Ng, Tony Joseph, FaisalZ. Qureshi, and Mehran Ebrahimi. Edge-connect: Generative image inpainting with adversarial edge learning. *Cornell University - arXiv, Cornell University - arXiv*, Jan 2019.
- [17] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.
- [18] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct 2019.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science, Lecture Notes in Computer Science*, Jan 2015.
- [20] Andrew Ross and Finale Doshi-Velez. Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients. *Proceedings of the AAAI Conference on Artificial Intelligence*, Jun 2022.

- [21] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, page 211–252, Dec 2015.
- [22] Min-cheol Sagong, Yong-goo Shin, Seung-wook Kim, Seung Park, and Sung-jea Ko. Pepsi : Fast image inpainting with parallel decoding network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019.
- [23] René Schuster, Oliver Wasenmüller, Christian Unger, and Didier Stricker. Sdc - stacked dilated convolution: A unified descriptor network for dense matching tasks. *arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition*, Apr 2019.
- [24] Yuhang Song, Chao Yang, Yeji Shen, Peng Wang, Qin Huang, and C.-C. Jay Kuo. Spg-net: Segmentation prediction and guidance network for image inpainting. *arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition*, May 2018.
- [25] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions, 2021.
- [26] Amey Thakur. Generative adversarial networks. *International Journal for Research in Applied Science and Engineering Technology*, page 2307–2325, Aug 2021.
- [27] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
- [28] Yi Wang, Tao Xin, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. Image inpainting via generative multi-column convolutional neural networks. *Cornell University - arXiv, Cornell University - arXiv*, Oct 2018.
- [29] Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, and Jiebo Luo. Foreground-aware image inpainting. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019.
- [30] Jie Yang, Zhiquan Qi, and Yong Shi. Learning to incorporate structure knowledge for image inpainting. Feb 2020.
- [31] Zili Yi, Qiang Tang, Shekoofeh Azizi, Daesik Jang, and Zhan Xu. Contextual residual aggregation for ultra high-resolution image inpainting. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020.

- [32] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition*, Nov 2015.
- [33] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative image inpainting with contextual attention. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.