

利用傅立叶卷积进行分辨率可靠的大掩模涂色

Roman Suvorov¹ Elizaveta Logacheva¹ Anton Mashikhin¹ Anastasia Remizova^{3*} Arsenii Ashukha¹
Aleksii Silvestrov¹ Naejin Kong² Harshith Goka² Kiwoong Park² Victor Lempitsky^{1,4}

¹莫斯科三星人工智能中心, ² Samsung Research, ³
EPFL 计算机与通信科学学院

⁴俄罗斯莫斯科斯科尔科沃科技学院



图 1：所提出的方法可以成功地大面积区域进行上色，并能很好地处理各种图像，包括具有复杂重复结构的图像。该方法适用于高分辨率图像，但只能在 256×256 的低分辨率下进行训练。

摘要

* 工作在莫斯科三星人工智能中心完成

尽管现代图像内绘系统取得了长足进步，但在处理大面积缺失区域、复杂几何结构和高分辨率图像时，仍常常力不从心。我们发现，造成这种情况的主要原因之一是内绘网络和损失函数都缺乏有效的感受野。为了缓解这一问题，我们提出了一种新方法，即大型遮罩内绘法 (LaMa)。LaMa 基于 i) 一种新的内绘制网络架构，该架构使用

快速傅立叶卷积 (FFC)，它具有宽图像感受野；ii) 高感受野知觉损失；iii) 大型训练掩码，它释放了前两个组件的潜力。我们的内绘制网络在一系列数据集上证明了其先进性，即使在具有挑战性的场景中（如周期性结构的完成）也能取得优异的性能。我们的模型能出人意料地将分辨率提升到比训练时更高的水平，而且参数和时间成本都低于竞争基准。代码见 <https://github.com/saic-mdal/lama>。

1. 引言

解决图像涂抹问题--逼真地填补缺失部分--需要 "理解" 自然图像的大尺度结构并进行图像合成。这一课题在前深度学习时代就已开始研究[1, 5, 13]，近年来，通过使用深度和广度神经网络[26, 30, 25]以及对抗学习[34, 18, 56, 44, 57, 32, 54, 61]，研究进展加快。通常的做法是在一个自动生成的大型数据集上训练内绘系统，该数据集是通过随机遮挡真实图像而创建的。

，通常使用带有中间预测的复杂两阶段模型，如平滑图像[27, 54, 61]、边缘[32, 48]和分割图[44]。在这项工作中，我们实现了最先进的简单单级网络的结果。

一个大的有效感受野[29]对于理解图像的全局结构，进而解决内绘问题至关重要。此外，在大掩膜的情况下，即使是大但有限的感受野也可能不足以获取生成高质量内绘所需的信息。我们注意到，流行的卷积架构可能缺乏足够大的有效感受野。我们对系统的每个组件进行了精心干预，以缓解这一问题，并释放单级解决方案的潜力。具体来说

i) 我们提出了一种基于最近开发的 *快速傅立叶卷积* (FFCs) [4]的内绘网络。即使在网络的早期层，FFCs 的感受野也能覆盖整个图像。我们的研究表明，FFCs 的这一特性提高了网络的感知质量和参数效率。有趣的是，FFC 的感应偏差允许网络泛化到训练期间从未见过的高分辨率（图 5、图 6）。这一发现带来了显著的实际好处，因为需要的训练数据和计算量更少。

ii) 我们建议使用基于高感受野的语义分割网络的感知损失[20]。这是因为我们发现，感受野不足不仅会影响内绘网络，还会影响感知损失。我们的损耗能促进全局结构和形状的一致性。

iii) 我们引入了一种积极的训练掩码生成策略，以释放前两个组件的高感受野的潜力。该程序可生成宽而大的掩码，从而迫使网络充分利用模型和损失函数的高感受野。

因此，我们提出了 *大型遮罩涂色* (LaMa)--一种新颖

的单阶段图像涂色系统。LaMa 的主要组成部分是高感受野架构 (*i*)、高感受野损失函数 (*ii*) 和生成训练掩码的先进算法 (*iii*)。我们将 LaMa 与最先进的基线进行了细致的比较，并分析了每个建议组件的影响。

通过评估，我们发现 LaMa 只需在低分辨率数据上进行训练，即可泛化到高分辨率图像。LaMa 可以捕捉和生成复杂的周期性结构，并且对大型遮罩具有鲁棒性。此外，与其他同类产品相比，LaMa 的可训练参数和推理时间成本大大降低。

2. 方法

我们的目标是对未知像素的二进制掩码 m 掩码的彩色图像 x 进行内绘，掩码图像表示为 $x \odot m$ 。掩码 m 与掩码图像 $x \odot m$ 叠加，得到四通道输入张量 $x' = \text{stack}(x \odot m, m)$ 。我们使用前馈涂色网络 $f_\theta(-)$ ，我们也称其为生成器。以 x' 为例，内绘网络以全卷积方式处理输入，生成内绘三通道彩色图像 $\hat{x} = f_\theta(x')$ 。训练是在一个数据集上进行的，该数据集包含从真实图像和合成的遮罩中获取的（图像、遮罩）对。

2.1. 早期图层中的全球背景

在具有挑战性的情况下，例如填充大型遮罩时，生成适当的内绘需要考虑全局内容。因此，我们认为，一个好的架构应该在管道的早期就拥有尽可能宽的感受野单元。传统的完全卷积模型，如 ResNet [14]，存在有效感受野增长缓慢的问题 [29]。由于卷积核通常较小（如 3×3 ），感受野可能不足，尤其是在网络的早期层。因此，网络中的许多层都会缺乏全局上下文，并且会浪费计算和参数来创建全局上下文。对于宽掩膜，特定位置上发生器的整个感受野可能都在掩膜内，因此只能看到缺失的像素。这个问题在高分辨率图像中尤为突出。

快速傅立叶卷积（FFC） [4] 是最近提出的一种算子，可以在早期层中使用全局上下文。FFC 基于信道快速傅立叶变换 (FFT)[2]，其感受野覆盖整个图像。FFC 将通道分成两个并行的分支：i) 局部分支使用传统的卷积；ii) 全局分支使用实 FFT 来考虑全局上下文。实值 FFT 只能应用于实值信号，而反实值 FFT 可确保输出为实值信号。与 FFT 相

比，实 FFT 只使用一半的频谱。具体来说，实 FFT 有以下步骤：

a) 将实数 FFT2d 应用于输入张量

$$\text{实数 FFT2d} : \mathbb{R}^{H \times W \times C} \rightarrow \mathbb{C}^{H \times \frac{W}{2} \times C},$$

并连接实部和虚部

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

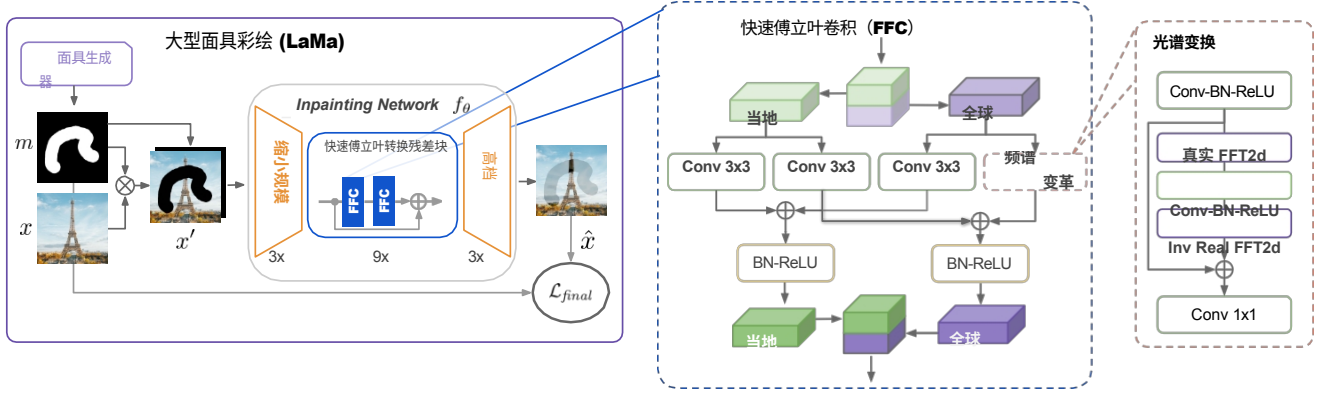


图 2：拟议的大掩模内绘制方法（LaMa）方案。LaMa 基于类似 ResNet 的前馈式内绘网络，该网络使用：最近提出的快速傅立叶卷积 (FFC)[4]、结合了对抗损失和高感受野感知损失的多成分损失，以及训练时间大掩模生成程序。

b) 在频域中应用卷积块

$$ReLU \circ BN \circ Conv_{l \times l} : \mathbb{R}^{H \times \frac{W}{2} \times 2C} \rightarrow \mathbb{R}^{H \times \frac{W}{2} \times 2C} ;$$

c) 应用逆变换恢复空间结构

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

$$反实数 FFT2d : \mathbb{C}^{H \times \frac{W}{2} \times C} \rightarrow \mathbb{R}^{H \times W \times C}。$$

最后，将本地分支（i）和全局分支（ii）的输出融合在一起。图 2 是 FFC 的示意图。

FFC 的威力 FFC 是完全可微分的、易于使用的、可直接替代传统卷积的技术。由于 FFCs 具有宽图像感受野，因此生成器可以从早期层开始考虑全局背景，这对于高分辨率图像内绘至关重要。这也提高了效率：可训练的参数可用于推理和生成，而不是“等待”信息传播。

我们的研究表明，FFCs 非常适合捕捉周期性结构，这些结构常见于砖块、梯子、窗户等人造环境中（图 4）。值得注意的是，在所有频率上共享相同的卷积会使模型趋向于尺度等差性[4]（图 5、图 6）。

2.2. 损失函数

内画问题本质上是模糊的。对于相同的缺失区域，

2.2.1 高感受野知觉损失

假设的监督损失要求发电机重建

精确地重建地面实况。然而，图像的可见部分往往不包含足够的信息来准确重建遮挡部分。因此，使用天真

监督会导致结果模糊，因为平均

可能会有许多似是而非的填充物，尤其是当“洞”变得更大时。我们将讨论所提出的损耗算法的各个组成部分，它们可以共同处理该问题的复杂性。

涂色内容的多种可信模式。

与此相反，感知损失法[20]通过预先训练的基本网络 $j(-)$ ，评估从预测图像和目标图像中提取的特征之间的距离。它不要求精确的重建，允许重建图像的变化。大掩模涂色的重点已转向对全局结构的理解。因此，我们认为使用感受野快速增长的基础网络非常重要。我们引入了 *高感受野知觉损失*（ $HRF PL$ ），它使用高感受野基础模型 $\phi_{HRF}(-)$ ：

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

其中， $[---]^2$ 是元素向运算， M 是顺序两阶段均值运算（层间均值为

层内均值）。 $\phi_{HRF}(x)$ 可以通过傅立叶或扩张卷积来实现。正如消融研究（表 3）所示，HRF 感知损失似乎对我们的大掩模涂色系统至关重要。

前置问题 训练感知损失的基础网络所依据的前置问题非常重要。例如，使用分割模型作为感知损失的骨干，可能有助于将重点放在高级信息上、例如，物体及其部件。相反，众所周知，分类模型更注重纹理[10]，这可能会带来对高层信息有害的偏差。

2.2.2 对抗性损失

我们使用对抗损失来确保涂色模型 $f_\theta(x')$ 能生成自然的局部细节。我们定义了一个判别器 $D_\xi(-)$ ，该判别器在局部斑块级别上工作[19]，对 "真实" 和 "虚假" 斑块进行判别。只有与屏蔽区域相交的斑块才会被贴上 "假" 标签。由于有监督的 HRF 感知损失，生成器很快就能学会复制输入图像的已知部分，因此我们将生成图像的已知部分标记为 "真实"。最后，我们使用非饱和对抗损失：

$$L_D = -E_x \sum_i h \log D_\xi(x_i) - E_{x,m} \sum_i h \log D_\xi(\hat{x}_i) \odot (1 - m) \quad (2)$$

$$L_G = -E_{x,m} \sum_i h \log D_\xi(\hat{x}_i) \quad (3)$$

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

其中， x 是数据集中的样本， m 是合成的掩码， $\hat{x} = f_\theta(x')$ 是 $x' = \text{stack}(x \odot m, m)$ 的绘制结果， sg_{var} 停止梯度变化， L_{Adv} 是要优化的联合损失。

2.2.3 最终损失函数

在最终损失中，我们还使用了 $R_1 = E_x ||\nabla D_\xi(x)||^2$ gradient penalty [31, 38, 7]，以及基于判别器的感知算法。

L_{DiscPL} 是判别网络 L_{DiscPL} 特征上的感知损失[45]。众所周知， L_{DiscPL} 可以稳定训练效果，在某些情况下还能略微提高性能。

嵌画系统的最终损失函数

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

是所讨论损失的加权总和，其中 L_{Adv} 和

L_{DiscPL} 负责生成自然的局部细节，而 L_{HRFPL} 则负责监督信号和全局结构的一致性。

2.3. 在培训期间生成面具

我们系统的最后一个组成部分是遮罩生成策略。每个训练示例 x' 都是训练数据集中的一张真实照片，上

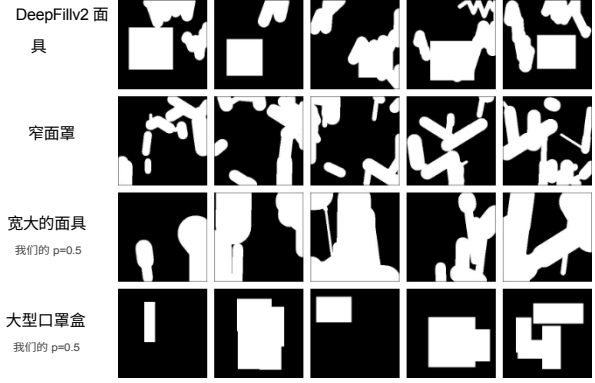


图 3：来自不同训练掩码基因组的样本

生成策略。我们认为，掩码的生成方式会极大地影响系统的最终性能。

我们采用了更激进的大掩膜生成策略。与 DeepFillv2 等传统做法不同，我们采用了更激进的大型掩膜生成策略。其中的掩码统一来自宽掩码或箱掩码策略。来自大掩膜策略的掩膜具有

面叠加了一个合成生成的遮罩。与数据增强对最终性能影响较大的判别模型类似，我们发现遮罩生成策略也会明显影响涂色系统的性能。

因此，我们选择了积极的大型掩膜生成策略。该策略统一使用由高随机宽度扩张的多边形链（宽掩膜）和任意长宽比的矩形（盒掩膜）的样本。图 3 展示了我们的掩码示例。

更重要的是，它的面积更大（直方图见补充材料）。使用我们的策略进行训练，可以帮助模型在宽掩膜和窄掩膜上都有更好的表现（表 4）。在准备测试数据集时，我们避免使用覆盖图像 50% 以上的遮罩。

我们测试了几种方法的大掩码训练和窄掩码训练，发现使用大掩码策略训练一般都能提高狭行掩码和宽掩码的性能（表 4）。这表明，增加掩码的多样性可能对各种内绘系统都有好处。采样算法见补充材料。

3. 实验

在本节中，我们将证明所提出的技术在标准低分辨率上优于一系列强大的基线技术，而在对更宽的孔进行内绘时，差异会更加明显。然后，我们进行了消融研究，显示了 FFC、高感受野知觉损失和大掩膜的重要性。令人惊讶的是，该模型可以泛化到从未见过的高分辨率，而与大多数通用基线相比，其参数要少得多。

实现细节 对于 LaMa Inpainting 网络，我们使用了类似 ResNet [14] 的架构，其中有 3 个下采样块、6-18 个残差块和 3 个上采样块。在我们的模型中，残差块使用 FFC。有关鉴别器架构的更多详情，请参见补充材料。我们使用 Adam [23] 优化器，平滑网络和判别网络的固定学习率分别为 0.001 和 0.0001。除非另有说明，否则所有模型都进行了 100 万次迭代训练，批量大小为 30。在所有实验中，我们都选择了超参数。

# 参数 ×10 ⁶	地点 (512 × 512)				CelebA-HQ (256 × 256)								
	窄掩		窄		方法	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓				
	↓LPIPS ↓FID	↓LPIPS ↓FID	↓LPIPS ↓FID	↓LPIPS ↓FID	↓ LaMa-Fourier (我们的)	27.0	0.630	0.090	2.21	0.135	5.35	0.058	7.26
	0.085	6.96		0.098									
CoModGAN [64]	109▲ 0.82▲30% 0.111▲23%	1.82▼18% 0.147▲9%	6.40▲20% 0.066▲14			16.8▲131%	0.079▼7%	24.4▲250%	0.102▲4%				
MADF [67]	85▲ 0.57▼10% 0.085▼5%	3.76▲70% 0.139▲3%	6.51▲22% 0.061▲5%			-	—	-	-				
AOT GAN [60]	15▼ 0.79▲25% 0.091▲1%	5.94▲169% 0.149▲11%	7.34▲37% 0.063▲10%			6.67▼8% 0.081▼4%	10.3▲48%	0.118▲20%					
GCPR [17]	30▲ 2.93▲363% 0.143▲59%	6.54▲196% 0.161▲19%	9.20▲72% 0.073▲27%			-	-	-	-				
高填充[54] ₃	▼ 9.24▲1361% 0.218▲142%	12.8▲479% 0.180▲34%	12.7▲137% 0.085▲49			-	—	-	-				
RegionWise [30]	47▲ 0.90▲42% 0.102▲14%	4.75▲115% 0.149▲11%	7.58▲42% 0.066▲14%	11.1▲53%	0.124▲46%	8.54▲23%	0.121▲23%	DeepFill v2 [57]					
	4▼ 1.06▲68% 0.104▲16%	5.20▲135% 0.155▲15%	9.17▲71% 0.068▲18%	12.5▲73%	0.130▲53%	11.2▲61%	0.126▲28%	EdgeConnect					
[32]	22▼ 1.33▲110% 0.111▲23%	8.37▲279% 0.160▲19%	9.44▲76% 0.073▲27%	9.61▲32%	0.099▲17%	9.02▲30%	0.120▲22%	RegionNorm [58]	12▼ 2.13▲236%				
	0.120▲33%	15.7▲613%	0.176▲31%	13.7▲156%	0.082▲42%	-		-	-				

表 1：在 Places 和 CelebA-HQ 数据集上对内绘法进行的定量评估。我们报告了 *学习感知图像补丁相似度* (LPIPS) 和 *弗雷谢特起始距离* (FID) 指标。与我们的 LaMa-Fourier 模型（第一行）相比，▲ 表示恶化，▼ 表示得分提高。这些指标针对不同的测试掩码生成策略（即窄掩码、宽掩码和分割掩码）进行了报告。LaMa-Fourier 始终优于各种基线。CoModGAN [64] 和 MADF [67] 是仅有的两个接近的基线。不过，这两个模型都比 LaMa-Fourier 重得多，平均性能也更差，这表明我们的方法能更有效地利用可训练参数。

采用坐标式波束搜索策略。这

该方案导致权重值 $\kappa = 10$, $\alpha = 30$, $\beta = 100$, $\gamma = 0.001$ 。除了损耗消融研究（见第 3.2 节）中描述的模型外，我们使用这些超参数对所有模型进行训练。在所有情况下，超参数搜索都是在单独的验证子集中进行的。有关数据集拆分的更多信息，请参见补充材料。

数据和指标 我们使用 Places [66] 和 CelebA-HQ [21] 数据集。我们沿用了最近的图像 2 图像文献中的既定做法，使用了 *学习感知图像补丁相似度* (LPIPS) [63] 和 *弗雷谢特起始距离* (FID) [15] 指标。与像素级的 L1 和 L2 距离相比，LPIPS 和 FID 更适合在多个自然补全是可信的情况下测量大型遮罩涂色的性能。实验管道使用 PyTorch [33]、PyTorch-Lightning [9] 和 Hydra [49] 实现。代码和模型可在 github.com/saic-mdal/lama 上公开获取。

3.1. 与基线的比较

我们将提议的方法与表 1 中列出的一些强大基线进

行了比较。我们仅使用公开的预训练模型来计算这些指标。对于每个数据集，我们验证了窄掩码、宽掩码和基于分割的掩码的性能。LaMa-Fourier 的性能始终优于大多数基线模型，而参数数量却少于最强的竞争者。仅有的两个竞争基线 CoModGAN [64] 和 MADF [67] 使用的参数分别多出 ≈ 4 倍和 ≈ 3 倍。这种差异在宽掩码时尤为明显。

用户研究 为了减少所选研究对象可能存在的偏差

模型 ↓	窄面罩		宽面罩		# 参数	块
	Convs	FID ↓ LPIPS ↓	LPIPS ↓	FID		
BaseFourier	27 9	0.63	0.090	2.21	0.135	
BaseDilated	46 9	0.66▲4% 0.089▼1%	2.30▲4% 0.136▲1%	1		
基础 常规	46 9	0.60▼5% 0.089▼1%	3.51▲59% 0.139▲3			
浅层傅立叶	19 6	0.72▲13% 0.094▲4%	2.31▲5% 0.138▲2			
深层常规	74 15	0.63	0.	0902.62▲18% 0.137▲2%		

表 2：该表展示了不同 LaMa 架构的性能，其他组件保持不变。与基础傅里叶模型（第一行）相比，▲ 表示性能下降，▼ 表示性能提高。基于 FFC 的模型在窄掩膜上可能会牺牲一点性能，但在宽掩膜上却明显优于更大的规则卷积模型。如图 4 所示，在视觉上，基于 FFC 的模型能更好地恢复复杂的视觉结构。

我们还进行了一项众包用户研究。用户研究的结果与定量评估结果非常吻合，并证明与其他方法相比，我们的方法产生的内绘效果更佳，更不易被察觉。用户研究的协议和结果见补充材料。

3.2. 消融研究

本研究的目的是仔细研究该方法不同组成部分的影响。在本节中，我们将介绍 Places 数据集的结果；CelebA 数据集的其他结果见补充材料。

F 的感受野 θ (-) FFC 增加了系统的有效感受野。添加 FFC 大幅提高了宽掩膜涂色的 FID 分数（表 2）。

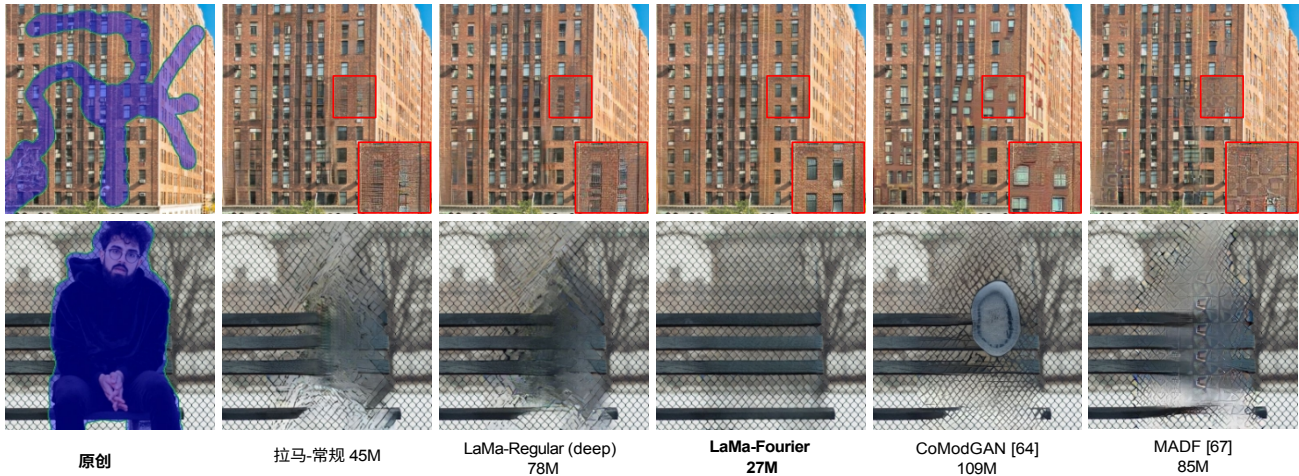


图 4：各种内绘系统在 512×512 图像上的并排比较。众所周知，窗户和链条栅栏等重复性结构很难上色。FFC 能够更好地生成这类结构。有趣的是，在参数较少的情况下，LaMa-Fourier 在比较中表现最佳，同时提供了可行的推理时间，即 LaMa-Fourier 平均只比 LaMa-Regular 慢 20%。

感受野的重要性最为明显

当模型应用的分辨率高于训练时的分辨率时，就会产生明显的伪影。如图 5 所示，当分辨率超过训练时所使用的分辨率时，具有规则卷积的模型就会产生明显的伪影。同样的效果得到了定量验证（图 6）。FFCs 还能影响图 4）。有趣的是，LaMa-Fourier 只比 LaMa-Regular 慢 20%，却比 LaMa-Regular 小 40%。

扩张卷积[55, 3]是允许感受野快速增长的另一种选择。与 FFCs 相似，扩张卷积提高了我们的图像绘制系统的性能。这进一步支持了我们的假设，即有效感受野的快速增长对图像绘制的重要性。然而，扩张卷积具有更多限制性感受野，并且严重依赖比例，导致对更高分辨率的泛化能力较差（图 6）。稀释卷积在大多数框架中都能广泛实现，在资源有限的情况下，例如在移动设备上，稀释卷积可作为傅立叶卷积的实用替代品。我们在补充材料中提供了有关 LaMa-Dilated 架构的更多细节。

损失 我们验证了使用稀释卷积实现的感知损失的高感受野确实提高了内绘质量（表 3）。事实证明，借口问题和使用稀释层以外的设计选择也很重要。对于每种损失变量，我们都进行了权重系数搜索，以确保评估的公平性。

模型	问题	扩展	FID	分段掩码	
				↓LPIPS	↓ L_{HRFL}
Segm.			+	5.69	0.059
RN50	Chf.	+	5.87▲3%	0.059	
L_{ChfPL} RN50	Chf.	-	6.00▲5%	0.061▲3%	

VGG19 Clf.

蒙板生成 更宽的训练蒙板可以改善 LaMa（我们的）和 RegionWise [30] 的宽孔和窄孔绘画效果（表 4）。但是，更宽的掩码可能会使结果更糟，DeepFill 就是这种情况。

L_{PL}	-	-	-	6.46▲13%	0.065▲9%
----------	---	---	---	----------	----------

表 3：使用不同感知损失训练的 LaMa-Regular 模型的比较。▲表示得分下降，▼表示与使用基于扩张卷积的分割 ResNet50 的 HRF 感知损失训练的模型相比得分提高（见第一行）。扩张卷积和借口问题都提高了得分。

v2 [57] 和 EdgeConnect [32] 在窄掩膜上的应用。我们推测，造成这种差异的原因是特定的设计选择（如发生器的高接收场或损耗函数），使得一种方法或多或少地适合于同时嵌合窄掩膜和宽掩膜。

3.3. 推广到更高分辨率

直接进行高分辨率训练不仅速度慢，而且计算成本高。不过，现实世界中的大多数图像编辑场景都需要在高分辨率下进行内绘。因此，我们使用 512×512 图像中的 256×256 裁剪图像来训练模型，并在更大的图像上对其进行评估。我们以完全卷积的方式应用模型，即对图像进行一次性处理，而不是分块处理。

基于 FFC 的模型向更高分辨率的转移效果明显更好（图 6）。我们推测，FFC 在不同尺度下具有更强的鲁棒性，这是由于 i) 图像范围内的感受性；ii) 图像范围内的感受性；iii) 图像范围内的感受性。

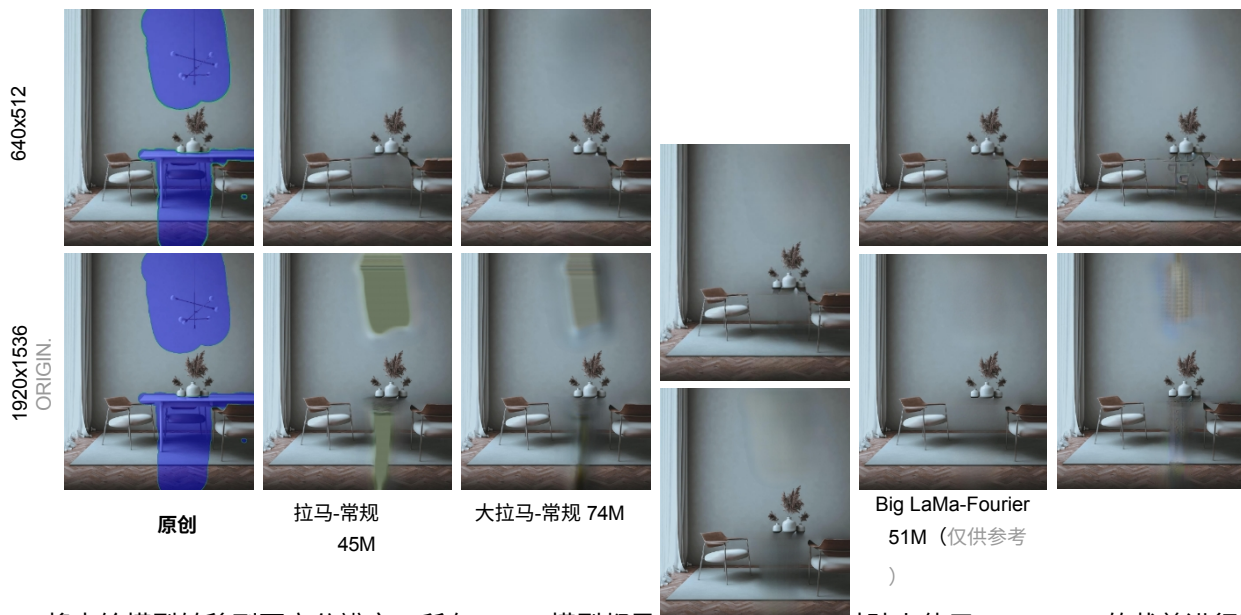
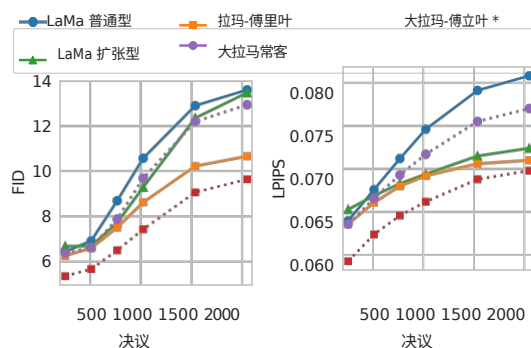


图 5：将内绘模型转移到更高分辨率。所有 LaMa 模型都是在 512×512 的基础上使用 256×256 的裁剪进行训练的，而 MADF [67] 则是直接在 512×512 的基础上进行训练的。随着分辨率的提高，采用常规卷积的模型很快就会产生关键的伪影，而基于 FFC 的模型则能继续生成语义一致、细节精细的图像。有关 51M 模型的更多正面和负面示例，请访问 bit.ly/3k0gaIK。

方法	窄面罩		宽面罩	
	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓
拉马常规	0.68	0.091	5.41	0.144
深度填充 v2	1.06	0.104	5.20	0.155
EdgeConnect	1.33	0.111	8.37	0.160
区域智慧	0.90	0.102	4.75	0.149
LaMa-Regular	0.60▼12%	0.089▼2%	3.51▼54%	0.139▼4%
DeepFill v2	1.35▲21%	0.107▲3%	4.34▼20%	0.148▼4%
EdgeConnect	2.78▲52%	0.141▲27%	7.94▼5%	0.160
地区智慧	0.74▼21%	0.095▼7%	3.56▼33%	
	0.144▼3			

表 4：该表显示了使用窄掩膜或宽掩膜训练不同内绘方法的性能指标。表中的 ▲ 表示性能下降，▼ 表示相应方法在宽掩膜训练下得分提高。LaMa 和 RegionWise inpainting 明显受益于宽掩码训练。这从经验上证明，积极的遮罩生成可能对其他内绘系统有益。

ii) 在尺度变化时保留频谱的低频；iii) 频域中 1×1 卷积固有的尺度等差性。虽然所有模型在 512×512 分辨率下都有相当好的通用性，但与所有其他模型相比，启用 FFC 的模型在 1536×1536 分辨率下保持了更高的质量和一致性（图 5）。值得注意的是，与普通基线相比，这些模型以更低的参数成本获得了更高的



质量。

3.4. 预告片模型：大拉玛

为了验证我们的方法对真实高分辨率图像的可扩展性和适用性，我们训练了一个大型的

图 6：基于 FFC 的内绘模型可以转换到更高分辨率，而这种分辨率在训练中从未出现过，同时质量下降幅度明显较小。所有 LaMa 模型均以 256×256 分辨率进行训练。*Big LaMa Fourier--我们的最佳模型--在不同条件下进行了训练（第 3.4 节），以供参考。

使用更多资源对 Big LaMa 模型进行内绘。

Big LaMa-Fourier 在三个方面与 LaMa-Fourier 不同：生成器的深度、训练数据集和批处理的大小。它有 18 个残差块，全部基于 FFC，产生 5100 万个参数。该模型是在 Places-Challenge 数据集 [66] 中的 450 万张图像子集上训练的。与我们的标准基础模型一样，Big LaMa 也只在大约 512×512 图像的 256×256 低分辨率裁剪上进行训练。Big LaMa 使用的批次规模更大，为 120（而不是其他模型的 30）。尽管我们认为该模型相对较大，但它仍小于某些基线模型。该模型在 8 个 NVidia V100 GPU 上进行了约 240 小时的训练。图 1 和图 5 展示了 Big LaMa 模型的入画示例。

4. 相关工作

早期数据驱动的内绘方法依赖于基于补丁的生成[5]和基于近邻的生成[13]。深度学习时代的首批内绘作品之一[34]使用了以对抗方式[11]训练的具有编码器-解码器架构的 convnet。这种方法是迄今为止深度内绘常用的方法。完成网络的另一种流行选择是基于 U-Net 的架构[37]，如[26, 50, 59, 27]。

一个共同关注的问题是网络把握局部和全局背景的能力。为此，文献[18]建议将扩张卷积[55]用于扩展感受野；此外，还提出了两个判别器，以分别鼓励全局和局部一致性。文献[46]建议在完成网络中使用具有不同感受野的分支。为了从空间上遥远的斑块中借用信息，[56] 提出了情境注意层。文献[28, 47, 65]还提出了其他的注意机制。我们的研究证实了在遥远位置之间有效传播信息的重要性。我们的方法的一个变体在很大程度上依赖于扩张卷积块，这是由文献[41]提出的。作为一个更好的替代方案，我们提出了一种基于主频变换（FFC）的机制[4]。这也符合最近在计算机视觉中使用变换器（Transformers）的趋势[6, 8]，以及将傅立叶变换作为自我关注的轻量级替代品的趋势[24, 35]。

在更全局的层面上，[56] 提出了一个从粗到细的框架，其中涉及两个网络。在他们的方法中，第一个网络在洞中完成粗略的全局结构，而第二个网络则以此为指导细化局部细节。这种遵循结构-纹理分解[1]这一相对古老思想的两阶段方法在随后的研究中得到了广泛应用。一些研究[40, 42]修改了这一框架，使粗略和精细结果成分同时获得，而不是按顺序获得。有几项研究提出了两阶段方法，将完成其他结构类型作为中间步骤：[32] 中的突出边缘、[44] 中的语义分割图、[48] 中的前景物体轮廓、[52] 中的梯度图和[36] 中的边缘保留平滑图像。另一个趋势是渐进式方法 [62, 12, 25, 61]。与所有这些研究相反，我们证明了精心设计的单阶段方法可以获得非常好的结果。

为了处理不规则的掩码，一些研究对卷积层进行了

修改，引入了部分卷积[26]、门控卷积[57]、轻量级门控卷积[54]和区域卷积[30]。我们还探索了各种形状的训练掩码，包括随机掩码[18]、自由形状掩码[57]和物体形状掩码[54, 61]。我们发现，只要训练掩码的轮廓足够多样化，掩码生成的具体方法就不那么重要了。

与面罩的宽度一样重要。

人们提出了许多损失来训练内绘网络。通常使用像素损失（如 ℓ_1 、 ℓ_2 ）和对抗损失。有些方法会对像素损失采用空间折扣加权策略 [34, 53, 56]。使用简单的连通判别器 [34, 52] 或 PatchGAN 判别器 [18, 59, 36, 28] 来实现对抗损失。其他流行的选择还有带有梯度惩罚判别器的 Wasserstein 对抗性损失 [56, 54] 和光谱归一化判别器 [32, 57, 27, 61]。根据之前的研究成果 [31, 22]，我们在系统中使用了 r_1 梯度惩罚补丁判别器。感知损失也是常用的方法，通常使用 VGG-16 [26, 47, 25, 27] 或 VGG-19 [51, 43, 32, 52]。

在 ImageNet 分类[39]上预先训练的骨骼。与这些工作相反，我们发现这种感知损失对于图像内绘来说并不理想，并提出了更好的替代方案。内绘框架通常包含样式 [26, 30, 30, 47, 32, 25] 和特征匹配。

51, 44, 32, 16] 损失。我们的系统也采用了后者。

5. 讨论

在这项研究中，我们对使用简单的单阶段方法进行大掩膜涂色进行了研究。结果表明，这种方法非常有竞争力，在适当选择架构、损失函数和掩码生成策略的情况下，可以推动图像内绘技术的发展。所提出的方法在生成重复的视觉结构（图 1、图 4）方面有很好的表现，这似乎是许多内绘方法所面临的问题。然而，当涉及到强烈的透视变形时，LaMa 通常会陷入困境（见补充材料）。我们想指出的是，对于不属于数据集的互联网复杂图像，通常会出现这种情况。FFC 能否解释这些周期性信号的变形仍然是个问题。有趣的是，傅立叶变换允许该方法适用于从未见过的高分辨率，而且与最先进的基线方法相比，参数效率更高。傅立叶或扩张卷积并不是接收高感受野的唯一选择。例如，高感受野可以通过 Vision 变换器获得[6]，这也是未来研究的一个令人兴奋的课题。我们相信，具有大感受野的模型将为开发高效的高分辨率计算机视觉模型

带来新的机遇。

感谢 Nikita Dvornik、Gleb Sterkin、Aibek Alanov、Anna Vorontsova、Alexander Grishin 和 Julia Churkina 提供的宝贵意见。

补充材料 详情和样本请参阅项目页面 <https://saic-mdal.github.io/lama-project/> 或补充材料 <https://bit.ly/3zhv2rD>。

<https://saic-mdal.github.io/lama-project/> 或补充材料 <https://bit.ly/3zhv2rD>。

参考资料

- [1] Marcelo Bertalmio, Luminita A. Vese, Guillermo Sapiro 和 Stanley J. Osher. 同时进行结构和纹理图像绘制。In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, 16-22 June 2003, Madison, WI, USA, pages 707-712. IEEE 计算机协会, 2003 年。
- [2] E Oran Brigham 和 RE Morrow. 快速傅立叶变换。《*电气和电子工程师学会频谱*》, 4 (12) : 63-70, 1967 年。
- [3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 利用深度卷积网和全连接 crfs 进行语义图像分割。In Yoshua Bengio and Yann LeCun, editors, *Proc. ICLR*, 2015.
- [4] Lu Chi, Borui Jiang, and Yadong Mu. 快速傅立叶卷积。见 H. Larochelle、M. Ranzato、R. Hadsell、M. F. Balcan 和 H. Lin 编辑的《*神经信息处理系统进展*》第 33 卷第 4479-4488 页。Curran Associates, Inc., 2020。
- [5] Antonio Criminisi、Patrick Pe'rez、Kentarō Toyama。通过基于示例的涂色去除物体。In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, 16-22 June 2003, Madison, WI, USA, pages 721-728. IEEE 计算机协会, 2003 年。
- [6] Alexey Dosovitskiy、Lucas Beyer、Alexander Kolesnikov、Dirk Weissenborn、Xiaohua Zhai、Thomas Unterthiner、Mostafa Dehghani、Matthias Minderer、Georg Heigold、Sylvain Gelly 等。一幅图像胜过 16x16 个单词：大规模图像识别的转换器。 *ArXiv 预印本 arXiv:2010.11929*, 2020。
- [7] H.H. Drucker 和 Y. Le Cun. 使用双反向传播提高泛化性能。 *IEEE Transactions on Neural Networks*, 3(6):991-997, 1992.
- [8] Patrick Esser、Robin Rombach 和 Bjorn Ommer. 用于高分辨率图像合成的驯服变换器。 *IEEE/CVF 计算机视觉和模式识别会议论文集* , 第 12873-12883 页, 2021 年。
- [9] WA Falcon and .al. Pytorch 闪电。 *GitHub*. 注 : <https://github.com/PyTorchLightning/pytorch-lightning>, 3, 2019.
- [10] Robert Geirhos、Patricia Rubisch、Claudio Michaelis、Matthias Bethge、Felix A. Wichmann 和 Wieland Brendel。 Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*, 2019.
- [11] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 生成对抗网络。 *arXiv preprint arXiv:1406.2661*, 2014.
- [12] Zongyu Guo, Zhibo Chen, Tao Yu, Jiale Chen, and Sen Liu. 利用全分辨率残差网络进行渐进式图像着色。 *第 27 届 ACM 国际多媒体大会论文集* , 第 2496-2504 页, 2019 年。
- [13] James Hays 和 Alexei A. Efros. 使用 百万张照片完成场景补全。 *ACM Trans. Graph.*, 26(3):4, 2007.
- [14] 何开明、张翔宇、任少清和孙健。图像识别的深度残差学习。 In *Proceed-*

- ings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016.
- [15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter.通过双时间尺度更新规则训练的甘斯收敛于局部纳什均衡 rium。arXiv preprint arXiv:1706.08500, 2017.
- [16] 郑辉、李杰、王秀梅、高新舟。图像细粒度嵌画。arXiv preprint arXiv:2002.02609, 2020.
- [17] Haakon Hukkelar, Frank Lindseth 和 Rudolf Mester。用可学习的特征归因进行图像绘制。Pattern Recognition : 42nd DAGM German Conference, DAGM GCPR 2020, Tu'bingen, Germany, September 28-October 1, 2020, Proceedings 42, pages 388-403.Springer, 2021.
- [18] Satoshi Iizuka、Edgar Simo-Serra、Hiroshi Ishikawa。全局和局部一致的图像补全。ACM Transactions on Graphics (ToG), 36(4):1-14, 2017.
- [19] 菲利普-伊索拉、朱俊彦、周廷辉和阿列克谢-A-埃弗罗斯。利用条件广告网络实现图像到图像的翻译。计算机视觉与模式识别 (CVPR), 2017 IEEE Conference on, 2017.
- [20] 贾斯汀-约翰逊、亚历山大-阿拉希和李菲菲。实时风格转移和超分辨率的感知损失。欧洲计算机视觉会议, 第 694-711 页。Springer, 2016.
- [21] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen.为提高质量、稳定性和 和变异而进行的甘斯渐进生长。arXiv preprint arXiv:1710.10196, 2017.
- [22] Tero Karras、Samuli Laine 和 Timo Aila.基于风格的生成式对抗网络生成器架构。In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4401-4410, 2019.
- [23] Diederik P Kingma 和 Jimmy Ba.Adam: 一种随机优化方法。arXiv preprint arXiv:1412.6980, 2014.
- [24] James Lee-Thorp、Joshua Ainslie、Ilya Eckstein 和 Santiago Ontanon。Fnet: 用傅立叶变换混合标记。arXiv preprint arXiv:2105.03824, 2021.
- [25] 李静远、王宁、张乐飞、杜波、陶大成。用于图像内绘的循环特征推理。IEEE/CVF 计算机视觉与模式识别/大会论文集, 第 7760-7768 页, 2020 年。
- [26] Guilin Liu、Fitsum A Reda、Kevin J Shih、Ting-Chun Wang、Andrew Tao 和 Bryan Catanzaro。使用部分卷积对不规则孔洞进行图像内绘。欧洲计算机视觉会议 (ECCV) 论文集, 第 85-100 页, 2018 年。
- [27] Hongyu Liu, Bin Jiang, Yibing Song, Wei Huang, and Chao Yang.通过带特征均衡的相互编码器-解码器反思图像绘制》。ArXiv 预印本 arXiv:2007.06929, 2020.
- [28] Hongyu Liu, Bin Jiang, Yi Xiao, and Chao Yang.图像内绘的协同语义关注。In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4170-4179, 2019.
- [29] 罗文杰、李宇佳、拉奎尔-乌塔松和理查德-泽梅尔。理解深度卷积神经网络中的有效感受野。见 D. Lee、M. Sugiyama、U. Luxburg、

- I.Guyon 和 R. Garnett 编辑, 《神经信息处理系统进展》, 第 29 卷。Curran Associates, Inc., 2016。
- [30] Yuqing Ma, Xianglong Liu, Shihao Bai, Lei Wang, Aishan Liu, Dacheng Tao, and Edwin Hancock. 针对大面积缺失的区域性对抗图像绘制 (Region-wise generative adversarial image inpainting for large missing areas)。 *arXiv preprint arXiv:1909.12507*, 2019.
- [31] Lars Mescheder、Sebastian Nowozin 和 Andreas Geiger。哪些甘斯训练方法会真正收敛? *国际机器学习大会 (ICML)*, 2018。
- [32] Kamyar Nazeri、Eric Ng、Tony Joseph、Faisal Z Qureshi。用对抗边缘学习进行内绘。 *arXiv:1901.00212*, 2019。
- [33] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Rai-son、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、吕方、白俊杰和 Soumith Chintala。Pytorch: 一个模仿风格的高性能深度学习库。H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alche'-Buc、E. Fox 和 R. Garnett 编辑, 《神经信息处理系统进展》第 32 期, 第 8024-8035 页。Curran Associates, Inc., 2019。
- [34] Deepak Pathak、Philipp Krahenbuhl、Jeff Donahue、Trevor Darrell 和 Alexei A Efros。上下文编码器: 通过内绘制进行特征学习。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536-2544, 2016。
- [35] 饶永明、赵文亮、朱铮、吕继文、周杰。用于图像分类的全局滤波网络。 *arXiv preprint arXiv:2107.00645*, 2021。
- [36] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li。结构流: 通过结构感知外观流进行图像内绘。In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 181-190, 2019。
- [37] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox: 用于生物医学图像分割的卷积网络。 *医学图像处理和计算机辅助干预国际会议*, 第 234-241 页。Springer, 2015。
- [38] 安德鲁-斯拉文-罗斯和菲娜尔-多希-维莱斯。通过正则化输入梯度提高深度神经网络的对抗鲁棒性和可解释性。In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1660-1669, 2018。
- [39] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211-252, 2015。
- [40] Min-cheol Sagong、Yong-goo Shin、Seung-wook Kim、Seung Park 和 Sung-jea Ko。Pepsi: 利用并行解码网络快速绘制图像。在

IEEE/CVF 计算机视觉与模式 识别会议, 第 11360-11368 页, 2019 年。

- [41] Rene' Schuster、Oliver Wasenmuller、Christian Unger 和 Didier Stricker。Sdc-stacked dilated convolution: 用于密集匹配任务的统一描述符网络。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2556-2565, 2019.
- [42] Yong-Goo Shin、Min-Cheol Sagong、Yoon-Jae Yeo、Seung-Wook Kim 和 Sung-Jea Ko。Pepsi++: 用于图像着色的快速轻量级网络。*IEEE 神经网络与学习系统事务*, 32 (1): 252-265, 2020 年。
- [43] Yuhang Song, Chao Yang, Zhe Lin, Xiaofeng Liu, Qin Huang, Hao Li, and C-C Jay Kuo. 基于上下文的图像绘制: 推断、匹配和翻译。*欧洲计算机视觉会议 (ECCV) 论文集*, 第 3-19 页, 2018 年。
- [44] 宋宇航、杨超、沈业骥、王鹏、黄钦和 C-C Jay Kuo. Spg-net: 用于图像内绘的分割预测和引导网络。*arXiv preprint arXiv:1805.03356*, 2018.
- [45] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 高分辨率图像同步与条件矢量语义处理。*电气和电子工程师学会计算机视觉和模式识别会议论文集*, 2018 年。
- [46] Yi Wang, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. 通过生成式多列会话神经网络进行图像绘制。*ArXiv preprint arXiv:1810.08771*, 2018.
- [47] 谢超豪、刘少辉、李超、程明明、左旺萌、刘晓、温世磊、丁二瑞。利用可学习的双向注意力图进行图像年龄内绘。In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8858-8867, 2019.
- [48] Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, and Jiebo Luo. 前景感知图像绘制。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5840-5848, 2019.
- [49] 奥姆里-亚丹 Hydra - 用于优雅配置 复杂应用程序的框架。Github, 2019 年。
- [50] 颜昭仪、李晓明、李牧、左汪萌、单世光。Shift-net: 通过深度特征重排进行图像绘制。*欧洲计算机视觉会议论文集 (ECCV)*, 第 1-17 页, 2018 年。

。

- [51] 杨超、卢昕、林哲、Eli Shechtman、Oliver Wang 和李浩。使用多尺度神经补丁合成的高分辨率图像绘制。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6721-6729, 2017.
- [52] Jie Yang, Zhiquan Qi, and Yong Shi. 学习结合结构知识进行图像内绘。*美国人工智能学会会议论文集*, 第34卷, 12605-12612页, 2020年。
- [53] Raymond A Yeh、Chen Chen、Teck Yian Lim、Alexander G Schwing、Mark Hasegawa-Johnson、Minh N Do。利用深度生成模型进行语义图像绘制。在

- IEEE 计算机视觉与模式识别会议论文集》*，第 5485-5493 页，2017 年。
- [54] Zili Yi、Qiang Tang、Shekoofeh Azizi、Daesik Jang 和 Zhan Xu。用于超高分辨率图像绘制的上下文残差聚合。2020年*IEEE/CVF 计算机视觉与模式识别大会论文集*，7508-7517页。
- [55] Fisher Yu 和 Vladlen Koltun。通过扩张卷积实现多尺度上下文聚合。*ArXiv 预印本 arXiv:1511.07122*, 2015.
- [56] 余佳慧、林哲、杨继美、沈晓辉、卢昕、黄轶翔。用文本注意力生成图像涂色。In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5505-5514, 2018.
- [57] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang。使用门控卷积的自由形式图像绘制。In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4471-4480, 2019.
- [58] Tao Yu, Zongyu Guo, Xin Jin, Shilin Wu, Zhibo Chen, Weip- ing Li, Zhizheng Zhang, and Sen Liu。用于图像着色的区域归一化。2020年，*AAAI*，第12733-12740页。
- [59] 曾艳红、付建龙、赵红阳和郭百宁。用于高质量图像内绘的金字塔-上下文编码器网络学习。In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1486-1494, 2019.
- [60] 曾艳红、付建龙、赵红阳和郭百英。用于高分辨率图像绘制的聚合上下文变换。In *Arxiv*, pages -, 2020.
- [61] Yu Zeng, Zhe Lin, Jimei Yang, Jianming Zhang, Eli Shecht- man, and Huchuan Lu。采用迭代置信反馈和引导上采样的高分辨率图像内绘。*欧洲计算机视觉会议》*，第 1-17 页。Springer, 2020.
- [62] 张浩然、胡珍珍、罗长志、左旺萌、王萌。利用渐进生成网络进行语义图像绘制。*第 26 届 ACM 多媒体国际会议论文集》*，第 1939- 1947 页，2018 年。
- [63] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shecht- man, and Oliver Wang。深度特征作为感知度量的不合理有效性。In *Proceedings of the IEEE conference on computer vision and pattern recogni- tion*, pages 586-595, 2018.
- [64] 赵胜宇、崔乔纳森、盛轶伦、董玥、梁晓、Eric I Chang 和 Yan Xu。通过共调制生成式对抗网络完成大规模图像补全。*In- ternational Conference on Learning Representations (ICLR)*, 2021.
- [65] 郑传霞、湛达仁、蔡剑飞。多元图像补全。In *Proceedings of the IEEE/CVF Con- ference on Computer Vision and Pattern Recognition*, pages 1438-1447, 2019.
- [66] Bolei Zhou、Agata Lapedriza、Aditya Khosla、Aude Oliva 和 Antonio Torralba。地点用于场景识别的千万级图像数据库。*IEEE patterns analysis and machine intelligence*, 40(6):1452-1464, 2017.

- [67] Manyu Zhu, Dongliang He, Xin Li, Chao Li, Fu Li, Xiao Liu, Errui Ding, and Zhaoxiang Zhang. 具有遮罩感知的端到端级联细化图像绘制 *IEEE 图像处理论文集* , 30:4855-4866, 2021 年。