

## 一、立论依据

课题来源、选题依据和背景情况、课题研究目的、理论意义和实际应用价值

课题来源:

选题依据和背景情况:

日益发展的人工智能已成为科技进步的重要推动力,计算机视觉作为其主要领域已应用到众多行业,在图像分类、目标检测和图像生成等任务上取得瞩目的成就。但是在实际应用中,由于存在物体破损或者遮挡的情况,有时采集的图像数据只能反映物体或者场景的局部特征信息,难以直观得到整体的面貌,限制了对物体或者场景图像识别能力的进一步提高。这种情况在考古研究中尤为常见,数量庞大的文物历经漫长岁月大部分已经破碎,通过人工手动拼接重建文物不仅费时费力,而且会增加错误率;历史人物图经常会由于材质破损而残缺不全,甚至无法辨识,增大了考古工作的难度。类似的问题也出现在视频监控系统中,人脸画面会因为部分遮挡覆盖而导致面部图像缺失,此时通过有限的信息还原人脸的整体图像,可以达到更好的识别效果。因此,以已知局部的物体或场景图像为条件生成整体图像,在考古学和公共安全方面起到重要作用。

图像内补与外推是图像处理领域的重要问题,基于图像的已知区域和生成区域具有相同的统计特性和几何结构特性的假设,其目标是根据已知的部分图像,生成具有较好感知融合效果的图像,不仅要在整个图像色彩连贯和语义内容保持全局一致,而且在生成区域接近真实的图像,即合成的图像要在局部和全局上都要满足具有相似的颜色、相似的纹理和相似的空间结构,在语义内容和风格上确保统一,使人不易区分出真实图像和生成图像间的差异,满足人类视觉感知一致的效果。图像内补旨在通过已知的图像信息,可以对图像内部较小区域的缺失信息进行近似真实的还原,图像外推则是由已知的局部信息,能够生成已知图像边界范围之外较大的未知区域内容将图像扩展,两者广泛应用于图像修复、图像编辑与全景图合成等方面。

早期图像内补可以通过样本块和扩散方式实现修复,根据图像的纹理外观特征,选取一定大小的图像块在图像未受损的区域搜索,将最优匹配样本块拷贝到图像空缺区域,虽然可以保持合成区域较好的纹理信息,但容易产生边缘结构的不一致,无法实现语义关联,很难准确恢复缺失区域的信息。传统基于数据驱动方法的图像外推,利用已有数据库进行检索并与缺失区域边缘进行匹配,提取不同的图像作为前景和背景得到合成图像,使两者对齐以保证连接边界部分的一致性,但是容易出现颜色差异而有明显边界现象,不能保证空间结构和语义内容的连续性。

近年来深度学习生成模型的发展为图像内补和外推任务提供了新的技术手段,基于生成对抗和卷积神经网络的图像修复成为研究热点,在高分辨率的图像内部缺失区域取得清晰连贯的修复效果,但是存在图像内补的语义信息和颜色不协调,图像外推的背景内容模糊的问题。

合理的图像内补与外推应该保持图像生成结果的局部与整体的空间结构和风格一致性,对于同一源图像而言,由于已知局部图像的局限性和背景图像的复杂性,不能保证图像融合效果。因此,在图像内补和外推时保证生成样本的融合感知效果成为了研究的重点,本工作拟通过在图像内部缺失和外推区域生成语义合理的空间结构内容,对局部和整体的风格和颜色协调进行研究,更好地实现合成图像的前景和背景融合,提高图像内补与外推的感知效果

和图像质量。

#### 课题研究目的：

本课题以局部的物体或者场景图像为条件进行图像生成研究作为入手点,采用改进的生成对抗网络 (GAN) 作为图像生成模型, 提出由含有部分信息的源图像恢复整体外观图像的一种图像外推算法。在保证空间结构和语义内容的一致性同时, 内补和外推区域生成尽可能真实的图像内容, 能够在物体和场景数据集上展示出较好效果, 通过实验验证所提算法在图像内补和外推问题上的有效性和可行性, 与当前最优的相关算法对比, 在视觉效果和评价指标两方面均有提升。

同时本课题在图像内部和外推时能够实现较好的图像融合效果。针对目前生成模型中存在的训练不稳定和模式崩塌问题, 通过改进网络结构和损失函数, 有效缓解普遍存在的训练困难问题。

#### 理论意义：

通过对生成对抗网络 (GAN) 分析, 改进网络结构, 能够通过训练数据样本更好地学习到图像特征分布, 经过概率分布函数生成近乎真实的内补和外推图像, 提高生成图像的质量。

用改进的损失函数和正则化项对网络训练过程加以约束, 避免模式崩塌问题, 生成具有较好感知融合效果的图像内补和外推结果。

#### 实际应用价值：

本课题提出的图像内补与外推算法, 将含有物体或者场景部分信息的源图像恢复整体图像, 可以利用有限的信息还原出图像完整内容, 对考古研究中的文物碎片重建和历史人物图的复原具有重要意义, 将有助于研究学者还原历史面貌, 展现历史文明风采。

不仅如此, 在视频监控和公共安全领域中存在的人脸等物体遮挡的恢复提供了可行有效的解决方案, 可以在目标追踪和物体识别等应用场景中发挥重要作用。

利用有限的图像数据通过图像内补和外推生成高度感知融合的样本, 不仅为遮挡人脸的身份识别辨认提供丰富的样本参考, 而且在计算机辅助工艺设计中发挥重要作用。

## 二、文献综述

## 国内外研究现状、发展动态；所阅文献的查阅范围及手段

### 国内外研究现状、发展动态：

图像内补与外推是计算机视觉领域里具有挑战性的问题，通过充分挖掘图像完好部分的纹理信息和语义信息，客观真实地还原图像，主要应用于受损照片修复、残缺文物虚拟复原和部分遮挡图像补全等方面。设计和优化图像内补与外推算法，使图像的局部像素和周围像素在色彩的亮度、颜色值等方面均匀变化，语义和空间结构连续且在整体上融合一致，得到符合人类视觉感知的图像效果，一直是国内外研究者们努力的方向。

传统的图像内补技术主要是基于扩散和样本块的方法。2000 年 Marcelo Bertalmio 等人[1]提出偏微分方程的扩散方法，将缺失区域边缘的有效像素从边界平滑地传播到缺失区域内部进行纹理信息填充，但是这种方法只能处理修复任务中的小面积空洞，会造成纹理区域中错误延伸或较大缺失区域中的模糊现象。与基于扩散的方法仅使用缺失区域的邻域像素不同，基于样本块的方法是通过搜索图像中匹配的替换补丁，可以利用较远的图像信息来填补缺失区域。为找到合适的图像样本块，2008 年 Denis Simakov 等人 [2] 提出了双向相似度方法，在计算样本块相似度时可以捕获更多的视觉信息，以减少模糊引起的视觉伪影。2009 年 Connelly Barnes 等人 [3] 提出 PatchMatch 方法，通过从源图像已知的未损坏区域搜索并复制相似的图像样本块来填充缺损区域，使用图像中的自然相干性作为先验信息以及快速的最近邻搜索算法降低计算成本。这些基于样本块的方法假定已知部分图像的语义内容与缺失区域相似，由匹配的局部特征可以在较大的缺失区域中简单可重复的纹理进行填充，但是缺乏对语义内容的理解，对于缺失区域具有复杂背景结构的图像无法生成合理的结果。

人类可以根据自己的生活经验，经过想象补充图像边界范围之外的内容进行图像外推。1989 年特拉华大学的 Helene Introub 和 Michael Richardson 等人实验证明 [4]，如果为实验者提供一些场景图片并且根据自己的记忆绘制场景图时，他们可以绘制出比实际提供的场景图片更大范围的图像内容。人们通过使用视觉上下文联系并将当前场景与记忆经验结合起来，可以预测图像边界之外可能存在的事物 [5]。但是对于计算机而言，图像外推并不容易，由于缺乏先验知识和合理推断，生成图像的背景比较模糊。

早期的图像外推通常采用的是图像检索和合成的方法，假定相似边界外区域可能具有类似的内容，使用描述相似场景样本的图像数据库，合成已知部分边界外的图像。新加坡国立大学的张寅达等人 [6]于 2013 年首次提出图像外扩任务并且通过照片得到完整全景图像。通过给定要外扩的源图像、引导性的全景图像以及源图像与全景图中的关注区域（ROI）之间的严格定位，他们的方法可以生成新全景图，其中合成的外扩内容来自源图像。此方法依赖于位置关系以及在引导性全景图中存在的重复图像。2014 年清华大学的汪森等人 [7]提出数据驱动的方法，给定要外推的源图像以及指定的大小和方向，使用库图像为外推区域确定视觉上连续一致的内容。此方法使用图像数据库来搜索合适的匹配图，并将其与给定的输入源图像对齐以保证连接边界部分的一致性，在风景图上展示出较好的图像外推效果。华盛顿大学的 Shan Qi 等人使用搜索网络资源的图像，达到扩大照片视野的效果 [8]。此方法通过给定输入图像的底层场景几何对网络得到的图像进行搜索、映射和重组，合成到输入参考图像外，将图像外扩问题公式化为马尔可夫随机场，能够处理较大的图像数据库信息，在视野、外观和布局变化方面具有较高的鲁棒性。但是，早期基于图像检索方法在外扩图像内容时候受限于参考数据库，只能用已有的库图像用于扩展源图像的边界，很难保证外扩图像边界的空间语义一致性。不仅如此，合成图中源图像和扩充的图像可能存在不同的颜色分布，无法实现较好的融合效果。

生成逼真的合成图像需要前景和背景区域之间的内容和风格很好的匹配。传统方法使用梯度域和统计全局颜色信息来确保风格一致。2003 年 Patrick Pérez 等人[10]提出泊松方程的梯度方法,重新整合得到融合后局部区域的颜色,虽然可以实现无缝拼接,但是当粘贴前景图和背景的颜色分布不均匀时,通常会产生明显的伪影。2006 年 Daniel Cohen-Or 等人[9]采用颜色协调的方法,根据一组编码颜色协调的预定义模板来调整图像颜色映射值,但是直接颜色匹配忽略了图像语义内容的合理性,无法实现较好的感知融合效果。

近年来,以深度学习为基本思想的图像生成模型迅速发展。2014 年,Goodfellow 等人 [11] 提出生成式对抗网络 (Generative adversarial networks, GAN)。GAN 启发自博弈论中的二人零和博弈,包含一个生成模型 (Generative model G) 和一个判别模型 (Discriminative model D)。生成模型捕捉样本数据的分布,判别模型是一个二分类器,判别输入是真实数据还是生成的样本。这个模型的优化过程是一个“二元极小极大博弈 (Minimax two-player game)”问题,通过生成器和判别器的相互对抗,学习样本的概率分布,从而基于概率分布生成各式各样的域样本形成了对抗模型,称为生成式对抗网络。

自 Goodfellow 等人提出 GAN 以来,各种基于 GAN 的衍生模型被提出,以改进模型结构,进一步进行理论扩展及应用。2016 年,Radford 等人提出了改进 GAN 架构的 DCGAN(深度卷积生成式对抗网络)[12],把有监督学习的 CNN 与无监督学习的 GAN 整合到一起,生成器和判别器分别学到对输入图像层次化的表示。为了解决训练梯度消失问题, Martin Arjovsky 等人在 2017 年提出了 Wasserstein GAN (WGAN)[13],用 Earth-Mover 代替 Jensen-Shannon 散度来度量真实样本和生成样本分布之间的距离。相比 KL 散度和 JS 散度, Earth-Mover 距离是平滑的,即便两个分布没有重叠,仍能反映它们的远近,从而提供有意义的梯度,开创性地解决 GAN 训练不稳定的问题。

许多基于深度学习的方法将图像内补建模为条件生成问题,其中显着的优势是在纹理修复的基础上对语义内容进行预测,弥补了传统图像内补算法的不足,使得整体效果更加真实。2016 年 Deepak Pathak 等人 [14]最早尝试使用编码器-解码器体系结构来首先提取特征然后重建输出,但是该网络努力保持全局一致性,并会生成带有视觉伪影的结果。2017 年 Satoshi Iizuka 等人[15]通过使用局部和全局判别器,为内部缺失区域生成近乎真实的图像内容并分别保持完整图像的一致性。2018 年余家辉等人[16]发现卷积神经网络在无法建立长期相关性。为了解决这个问题,他们提出在上下文中注意机制从远离缺失区域提取特征。普通卷积在所有有效、无效和空洞边界的像素/特征上应用相同的滤波器,从而引起视觉伪影。为了解决这一局限性, Guilin Liu 等人 [17]提出部分卷积 (Partial Conv) 在生成器网络的每一层使用上一层更新后的掩码,让网络仅使用掩码中的有效像素,避免在输入的掩码位置中替换像素引起结果的伪影现象。但是通过输入掩码将图像中的像素位置归类为无效或有效,部分卷积会忽略感受野的神经元没有覆盖输入的有效图像像素信息。因此 2019 年余家辉等人 [18]提出适用于自由形状图像修复的门控卷积算法 SN-PatchGAN,通过扩展动态特征门控选择机制使其在生成网络所有层中的每个位置均可学习,避免了部分卷积忽略有效像素信息的问题。Kamyar Nazeri 等人[19]提出的 Edge Connect 通过使用图像已知部分的先验信息来生成合理的结构,首先恢复图像边缘信息,然后填充内部缺失的区域。但是由于边缘草图的表示能力有限,该方法可能会在图像缺失区域的边界生成错误的细节。

GAN 在图像外推的应用也逐渐发展起来。2018 年斯坦福大学的 Mark Sabini 和 Gili Rusak [20]提出基于 DCGAN 网络架构的深度学习方法用于图像外推任务,将给定需要外推图像和掩码图像输入生成器 G 中,生成得到图像中不仅含有输入源图还包括外推部分图像,判别器 D 用于评判生成图像是否属于真实样本,在 Places 365 场景数据集上证明了方法的可行性,但是生成图只有  $128 \times 128$  大小,解码重构阶段部分细节纹理信息无法恢复,超出输入源图



边界的外推部分表现出过度平滑和模糊的现象。2019 年香港中文大学 Yi Wang 等人 [21] 提出语义再生网络 (Semantic regeneration network, SRN) 用于图像外推。此方法通过特征扩展 (Feature expansion) 和图像上下文预测 (Context prediction) 两个阶段, 分别生成得到图像整体特征和外推后的预测生成图, 在图像单边外推和四周外推实验上得到比较好的效果。谷歌公司的 Piotr Teterwak 等人 [22] 采用 Wasserstein GAN 网络结构的生成器以及含有预训练特征分类网络的判别器, 有助于生成内容和真图相匹配, 提高图像外推的语义效果。悉尼大学的杨宗欣等人 [23] 在生成器中加入水平跳跃短接 (SHC) 和递归内容传输 (RCT) 模块, 通过多步迭代生成外推更大的区域。虽然目前图像外推在保持生成的图像与原始输入之间的空间和内容一致性取得较大改进, 但是当已知较少的图像前景信息需要生成复杂的背景信息时, 往往存在语义不合理和色斑现象, 影响融合后图像的感知效果。

深度学习的发展同样促进了人们对图像感知融合的研究, 为了生成逼真的合成, 需要调整前景和背景的上下文语义信息和颜色风格使其兼容。2015 年 Jun-Yan Zhu 等人 [24] 提出 CNN 模型来对图像进行真实感评估, 改善合成图像真实感。2017 年 Yi-Hsuan Tsai 等人 [25] 提出了生成协调一致图像的端到端卷积神经网络, 其中使用了额外分支来合并语义信息。2020 年 Wenyan Cong 等人 [26] 提出 DoveNet 和域验证辨别器用于合成图像融合, 在图像协调的数据集评估了图像协调算法的有效性。

#### 所阅文献的查阅范围及手段:

使用学校图书馆的书籍等文献资料、电子档案资料, 通过各大论文网站获取本研究领域的相关论文, 同时关注图像处理领域的一些会议和期刊, 以跟踪国内外的的发展趋势, 主要参考范围如下:

- <http://ieeexplore.ieee.org>
- <http://www.engineeringvillage.com>
- <http://apps.webofknowledge.com>
- <http://www.csdn.net/>
- <http://www.cnki.net/>
- <http://www.sciencedirect.com/>
- <https://scholar.google.com/>
- <https://arxiv.org/>

#### 参考文献:

- [1] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester. Image Inpainting. In Proceedings of the 27th Annual Conference on Computer Graphics and in-Interactive Techniques, 2000.
- [2] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In CVPR, 2008.
- [3] C. Barnes, E. Shechtman, A. Finkelstein, and D. B Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics (ToG), 2009.
- [4] Introub H, Richardson M. Wide-angle memories of close-up scenes[J]. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1989, 15(2): 179.
- [5] Lyle K, Johnson M. Importing perceived features into false memories[J]. Memory, 2006, 14(2): 197-213.
- [6] Zhang Y, Xiao J, Hays J, et al. Framebreak: Dramatic image extrapolation by guided shift-maps. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1171-1178.

- [7] Wang M, Lai Y, Liang Y, et al. Biggerpicture: data-driven image extrapolation using graph matching[J]. ACM Transactions on Graphics, 2014, 33(6).
- [8] Shan Q, Curless B, Furukawa Y, et al. Photo uncrop[C].European Conference on Computer Vision. Springer, Cham, 2014: 16-31.
- [9] D Cohen-Or, O Sorkine, R Gal, T Leyvand. Color Harmonization. ACM Transactions on Graphics, 2006.
- [10] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. ACM Transactions on Graphics (SIGGRAPH), 2003.
- [11] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C].Advances in neural information processing systems. 2014: 2672-2680.
- [12] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. International Conference on Learning Representations, 2016.
- [13] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. arXiv preprint arXiv:1701.07875, 2017.
- [14] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting. In CVPR, 2016.
- [15] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and Locally Consistent Image Completion. ACM Transactions on Graphics (SIGGRAPH 2017), 2017.
- [16] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In CVPR, 2018.
- [17] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In ECCV, 2018.
- [18] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In ICCV, 2019.
- [19] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image in-painting with adversarial edge learning. In ICCVW, 2019.
- [20] Sabini M, Rusak G. Painting outside the box: Image outpainting with GANs[J]. arXiv preprint arXiv:1808.08483, 2018.
- [21] Wang Y, Tao X, Shen X, et al. Wide-Context Semantic Image Extrapolation . In CVPR, 2019.
- [22] Teterwak P, Sarna A, Krishnan D, et al. Boundless: Generative Adversarial Networks for Image Extension. In ICCV, 2019:
- [23] Yang Z, Dong J, Liu P, et al. Very Long Natural Scenery Image Prediction by Outpainting. In ICCV, 2019.
- [24] Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman, and Alexei A Efros. Learning a discriminative model for the perception of realism in composite images. In ICCV, 2015.
- [25] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. Deep image harmonization. In CVPR, 2017.
- [26] Wenyan Cong, Jianfu Zhang, Li Niu, et al. DoveNet: Deep Image Harmonization via Domain Verification. In CVPR, 2020.