



# 基于感知融合的图像内补与外推研究

Research on Image Interpolation and Extrapolation Based on Perceptual Fusion

—— 毕业论文开题报告

答辩人：刘鸿智

指导老师：郑海永

答辩日期：2020年4月29日



- 选题依据和背景情况
  - 国内外研究现状
  - 研究意义
  - 研究内容
  - 论文安排
-

## 1. 选题依据

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。



图1. 左图和右图分别是佩戴口罩和未戴口罩时的人脸识别和测温。

## 1. 选题依据

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2. 左图和右图分别是待修复的仕女图和修复后的仕女图。





## 1. 选题依据

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，其目标是根据已知图像信息，生成具有较好的感知融合效果的图像结果。

合成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。

## 1. 选题依据

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。

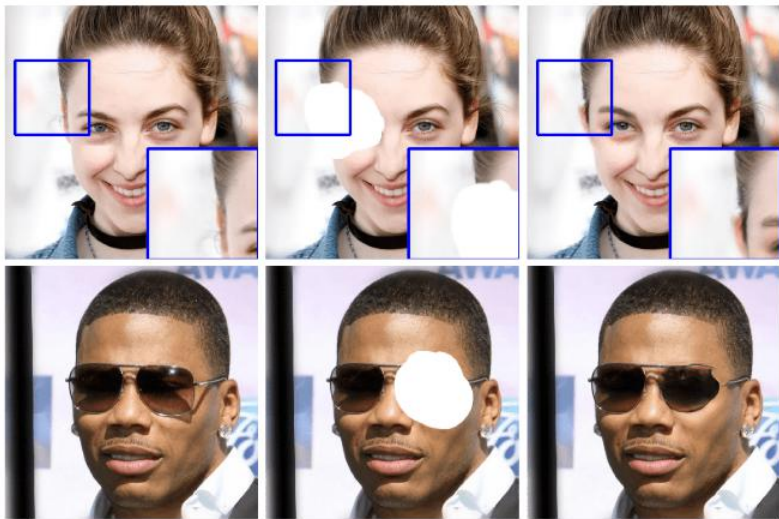


图3. 图像内补效果图。从左到右分别是真图，缺失图，内补图。



图4. 图像外推图。第一行是从左向右单边外推，第二行是全景图。

## 2. 背景情况

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。

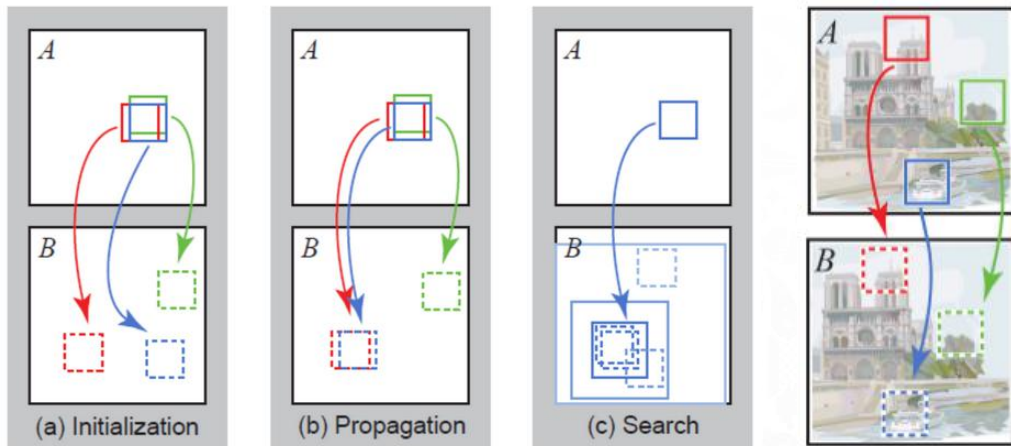


图5. (a) 补丁最初具有随机分配；(b) 蓝色补丁检查上方/绿色邻居和左侧/红色邻居，以查看它们是否会改善蓝色映射，从而促进良好的匹配；(c) 补丁随机搜索同心邻域中的改进。



## 2. 背景情况

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。

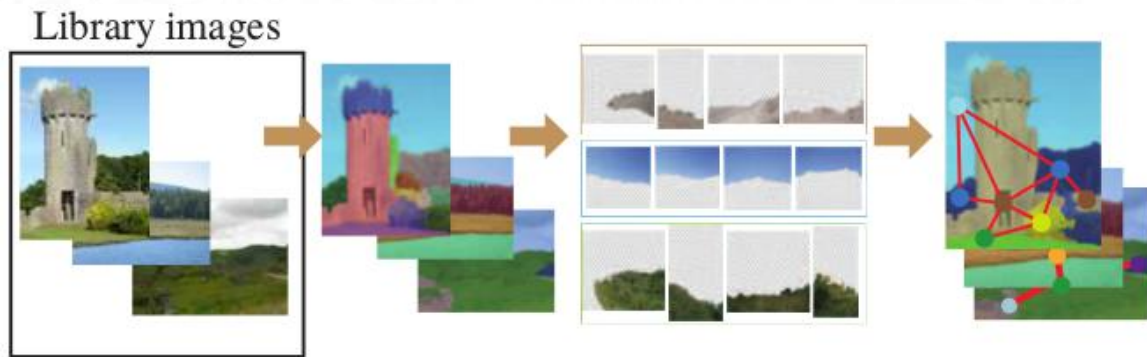


图6. 通过从已有数据库检索，提取不同图像进行边界对齐，用于前景和背景的合成，得到最右结果图。



## 2. 背景情况

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局部与整体的空间结构和风格一致性，由于已知局部图像的局限性和背景图像的复杂性，现有方法不能保证图像具有较好的感知融合效果。

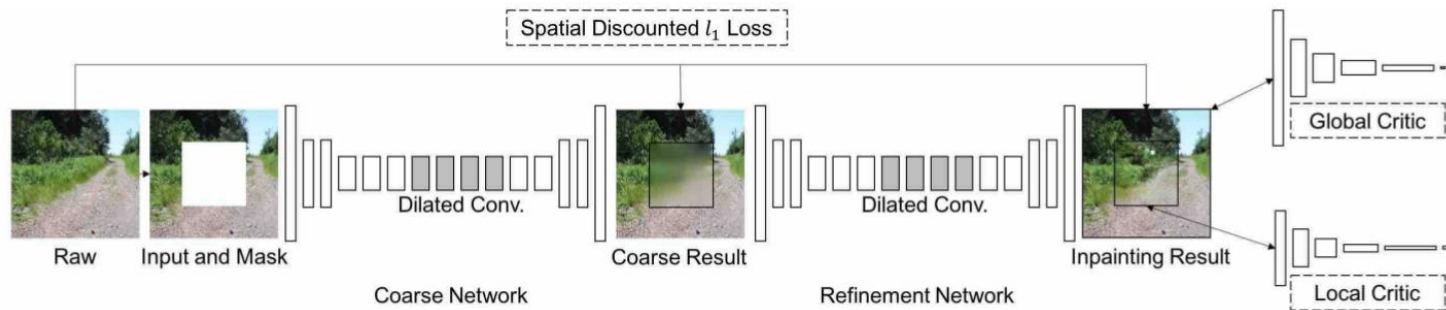
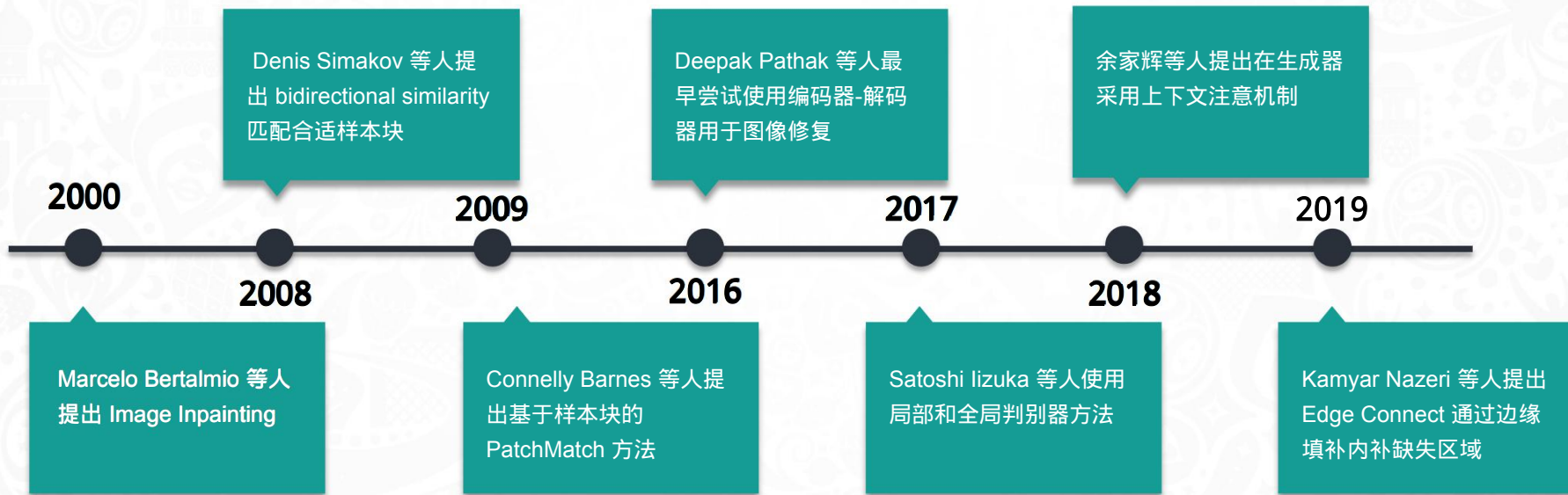


图7. 图像内补网络结构图。第一阶段网络生成粗略结果。第二阶段网络生成图像内补的细节纹理。

## 1. 图像内补



## 2. 图像外推





## 3. 传统图像内补与外推

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，只能处理修复任务中的小面积空洞，较大缺失区域中的内容模糊现象。

与基于扩散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。为找到匹配合适的图像样本块，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] 提出基于样本块的PatchMatch 方法。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填充简单可重复的纹理，但是由于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合成效果并不理想。

[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.





## 3. 传统图像内补与外推

早期的图像外推通常采用的是图像检索和合成的方法。新加坡国立大学的张寅达等人[1]于 2013 年首次提出图像外推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,该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 年清华大学的汪淼等人[2]提出数据驱动的方法,给定要外推的源图像以及指定外推的大小和方向,使用库图像为外推区域确定视觉上连续一致的内容。华盛顿大学的 Shan Qi 等人[3]使用搜索网络资源的图像,达到扩大照片视野的效果。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,将图像外推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。

但是,早期基于图像检索方法在外推图像内容时候受限于所使用的参考数据库,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连贯一致性。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。

[1] Y. Zhang, J. Xiao, J. Hays, P. Tan. Framebreak: Dramatic image extrapolation by guided shift-maps. In CVPR, 2013.

[2] M. Wang, Y. Lai, Y. Liang, R. R. Martin. Biggerpicture: data-driven image extrapolation using graph matching. ACM Transactions on Graphics, 2014.

[3] Q. Shan, B. Curless, Y. Furukawa, C. Hernandez. Photo uncrop. In ECCV, 2014.



## 4. GAN

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[1]提出生成式对抗网络 (Generative adversarial networks, GAN), 包含一个生成模型 (Generative model, G) 和一个判别模型 (Discriminative model, D)。

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) [2], 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) [3], 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。

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza. Generative adversarial nets. In NIPS, 2014.

[2] A. Radford, L. Metz, S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In ICLR, 2016.

[3] M. Arjovsky, S. Chintala, L. Bottou. Wasserstein GAN. arXiv preprint arXiv:1701.07875, 2017.



## 5. 基于深度学习的图像内补与外推

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，使得整体效果更真实。2016 年 Deepak Pathak 等人[1]最早尝试使用编码器-解码器体系结构来首先提取图像特征然后重建输出，但生成结果带有视觉伪影。2017 年 Satoshi Iizuka 等人[2]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[3]提出在上下文中注意机制从远离缺失区域提取特征。为了解决普通卷积的局限性，Guilin Liu 等人 [4]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。

2019 年Kamyar Nazeri 等人[5]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。

[1] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. Efros. Context encoders: Feature learning by inpainting. In CVPR, 2016.

[2] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and Locally Consistent Image Completion. ACM Transactions on Graphics (ToG), 2017.

[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In CVPR, 2018.

[4] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In ECCV, 2018.

[5] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and Mehran Ebrahimi. Edgeconnect: Generative image in-painting with adversarial edge learning. In ICCVW, 2019.



## 5. 基于深度学习的图像内补与外推

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 [1]提出基于 DCGAN 网络架构的深度学习方法用于图像外推任务,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[2]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[3]采用 Wasserstein GAN 网络结构的生成器以及含有预训练分类网络的判别器,有助于生成内容和真值图相匹配,提高图像外推的语义效果。悉尼大学的杨宗欣等人[4]在生成器中加入水平跳跃短接 (SHC)和递归内容传输 (RCT)模块,通过多步迭代外推生成更大的区域。

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,存在外推语义不合理和色斑现象,影响合成图像的真实性。

[1] M. Sabini, G. Rusak. Painting outside the box: Image outpainting with GANs. arXiv preprint arXiv:1808.08483, 2018.

[2] Y. Wang, X. Tao, X. Shen, J. Jia. Wide-Context Semantic Image Extrapolation. In CVPR, 2019.

[3] P. Teterwak, A. Sarna, D. Krishnan. Boundless: Generative Adversarial Networks for Image Extension. In ICCV, 2019.

[4] Y. Z. Dong J, Liu P, et al. Very Long Natural Scenery Image Prediction by Outpainting. In ICCV, 2019.





## 1. 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,提出基于感知融合的图像内补与外推算法。在网络结构设计中,充分利用图像已知部分的语义和风格等图像感知信息,保证空间结构和语义内容的一致性同时,使内补和外推区域生成尽可能真实的图像内容。

本课题所提方法能够在图像内部和外推任务中实现较好的感知融合效果,通过对目前图像内补算法进行改进,使图像局部与整体的颜色和风格融合;缓解图像外推任务中出现背景内容模糊的问题,生成具有合理语义信息和纹理细节的外推结果。

## 2. 理论意义

通过对生成对抗网络(GAN)分析,改进现有内补与外推算法的生成模型的网络结构,能够通过数据训练样本更好地学习到已知图像的语义和风格特征,融合感知信息,生成近乎真实的内补和外推图像结果,提高生成图像的真实性。

用改进的损失函数和正则化项对网络训练过程加以约束,生成具有较好感知融合效果的内补和外推的图像结果。

## 3. 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。

为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。



## 1. 学术构想与思路

首先，针对生成对抗网络和生成模型做前期调研，分析和学习目前世界上最先进的生成模型用于图像内补与外推任务上的学术成果。

其次，归纳总结图像内补和外推问题的本质和共同点，分析各自目前存在的问题，探索和设计出适合图像内补和外推的生成模型的网络结构，并提出改进和创新的思路。

最后设计实验方案进行验证，根据实验结果证明本工作所提出方法的有效性。

## 2. 主要研究内容

(1)针对图像内补，合成缺失区域的内容与已知部分的语义和风格一致性问题进行研究，通过颜色协调和风格转换达到局部与整体的协调融合。

(2)针对图像外推任务中，生成较大背景区域时细节信息不足和模糊问题进行研究，提高前景和背景空间结构的一致性，能够得到符合语义信息、具有纹理细节的外推结果。

(3)对图像内补和外推任务中的补全区域，生成风格一致的语义内容和色彩一致的融合图像进行研究，提高生成图的真实性和感知效果。

## 3. 拟解决的关键技术或问题

(1)在维持图像已知部分内容的固有特征的前提下，使图像内补结果在全局具有颜色、语义等方面的一致性，实现较好的融合效果。

(2)在图像外推区域，生成具有细节和语义信息的背景内容，保持前景和背景的风格一致性，符合视觉感知效果。





## 4. 拟采取的研究方法

本课题拟使用的研究方法包括：文献研究法、比较法、探索性研究法、实验法、定量分析法和定性分析法。

(1)文献研究法。通过大量查阅文献来获得资料，从而全面地、正确地了解掌握生成对抗网络的生成模型技术。了解图像内补与外推技术的发展历史和现状，进而帮助确定研究课题。

(2)比较法。对现有经典的图像内补与外推算法进行概括总结，并比较其优缺点。

(3)探索性研究法。针对图像条件生成网络以及目前先进的图像内补与外推算法所存在的问题进行改进。

(4)实验法。设计实验方案，通过实验验证所提出算法的有效性。

(5)定性与定量评价法。对实验结果进行定性与定量评价，从而判断所提出的算法是否有效。



## 5. 研究步骤

- (1)设计实验方案，确定数据集。
- (2)分析问题，设计网络模型。
- (3)训练并测试网络结构模型的效果。
- (4)根据实验结果，总结所提方法。

## 6. 可行性分析

- (1)经过研究生阶段的学习，对本课题方向的研究有一定的了解，有足够的知识储备。
- (2)实验室具有完善的生成模型研究实验平台和充足的计算资源。
- (3)后期实验主要在带有NVIDIA显卡的服务器上对算法进行训练和测试，满足算力要求。



## 7. 论文创新点

- (1)本文提出了一种新的图像内补与外推算法，提高生成图像的真实性。
- (2)本文对已有图像内补与外推算法进行改进，在补全区域生成风格一致的语义内容和色彩一致的融合图像，提升图像的感知融合效果。
- (3)用 spectral normalization 技术对判别器的权重进行归一化操作，加强训练的稳定性。改进新的损失函数，并设计新的正则化项，将不同尺度的生成器和判别器的损失函数加入总体损失函数中，加强网络的训练效果。

## 8. 预期研究结果

- (1)找到适合于本研究的训练数据集预处理方法。
- (2)提出基于感知融合的图像内补与外推模型，提升图像的真实性。
- (3)通过实验定性和定量的分析当前方法的优缺点。

阶段及内容	起讫日期	阶段成果形式
了解深度学习基本方法与前学习基本图像处理和深度学习知识，阅读前沿论文，掌握计算机模型算法。学习生成式模型，了解生成式对抗网络最前沿成果等。	2019 年 7 月— 2019 年 10 月	了解深度学习基本方法与前沿发展，掌握流行的生成式模型算法。
选定课题，进行有关本课题的相关资料的查询和收集，并详细阅读和学习。	2019 年 11 月— 2019 年 12 月	根据课题要求，阅读有关图像外推的文献，理解相关的原理和应用，找到可行的方法。
针对拟解决的问题进行关键技术问题的研究学习。	2020 年 1 月— 2019 年 6 月	设计网络型并程序实现，完成程序的初步设计和总体规划。
设计实验方案，将涉及到的算法一一实现并做对比分析，寻找最适合算法组合。	2020 年 7 月— 2019 年 9 月	对数据进行处理，并训练自己设计算法，分析实验结果。
对本课题的研究应用工作进行总结、提炼，归集资料、确定大纲、完成草稿，并最终形成系统的完整的学位论文，提交学校准备盲评。	2020 年 10 月— 2020 年 12 月	论文初稿。
对批阅、评审后的论文进行有针对性的调整修改，进一步完善课题的体系结构，丰富课题的内容，并最终完稿，为毕业答辩做好充分的准备。	2020 年 1 月— 2020 年 3 月	论文定稿。





# Q & A