

中国海洋大学

硕士研究生学位论文

开题报告

学 院 信息科学与工程学院

专 业 电子与通信工程

学 号

研 究 生

指导教师

论文题目 基于 GAN 的多样性图像外推

入学时间 2018 年 09 月

中国海洋大学开题报告申请表

(此页由系统中打印出替换)

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,人脸图像会因为遮挡覆盖而导致部分面部图像缺失,此时通过从有限的信息中还原为人脸的整体图像,可以达到更好的识别效果。因此,以局部的物体或场景图像为条件生成全局图像的图像外推过程,对考古学和公共安全起到重要作用。

图像外推与图像修复有关,不同的是图像修复包括图像复原和图像编辑等,主要对图像内区域进行修复。图像修复作为数字图像处理的一个重要分支,应用在图像噪声去除和修复老照片等许多方面。而图像外推是将图像的内容扩展到边界范围之外,由已知的局部信息恢复整体图像。传统图像检索的方法是通过网络资源和图像数据库检索,找到与给定源图相匹配的图像,然后使两者对齐以保证连接边界部分的一致性。近年来,生成对抗网络(GAN)作为生成模型的出现引起深度学习领域极大的关注,并且在图像修复任务上取得较好的效果,这让图像外推的进一步发展成为可能。

合理的图像外推应该具有多种可能性,对于同样的源图像,由于已知图像信息的局限性,局部信息位置的不确定性以及背景信息的复杂性,结果往往不是唯一的。研究证明,人们能够根据自己的经验预测已知图像以外的内容,比如推测出具有相同眼睛特征的不同人脸形状,或者设计具有相同零件的不同款式汽车。图像外推结果的多样性,不仅为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。但是对于当前的很多基于深度学习的图像修复算法,在网络训练时只有唯一的真值数据与生成图对应,这导致输入网络的一个图像只能生成一个最优的结果,不能满足某些应用场景中样本多样性的要求。除此之外,虽然生成对抗网络(GAN)在图像生成任务上取得惊艳效果,但是在逼近图像数据真实分布时往往会产生模式崩塌问题,生成器会倾向于生成图像数据分布空间中某些位置,结果的类别和风格单一,造成训练困难现象。分析影响多样性的原因,设计合适的生成模型,对解决图像外推结果多样性的问题至关重要。

因此,在图像外推时保证生成样本的真实性和多样性成为了本课题研究的重点。目前基于深度学习的图像修复算法在外推过程中存在生成内容模糊和细节缺失,由于所采用的网络结构和训练策略的约束,往往只生成唯一的结果对应输入图,并不能解决多样性的图像外推问题。但是图像修复算法具有较好的空间结构和语义内容的一致性效果,成为解决图像外推问题的理论基础。

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, 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。

同时本课题在图像外推时能够实现生成结果的多样性, 以解决当前相关的图像修复算法局限于单一样本的多样性不足问题。针对目前生成模型中存在的模式崩塌问题, 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, 控制图像流型空间中的样本距离, 避免模式崩塌问题, 生成多样性的图像外推结果。

实际应用价值：

本课题提出的图像外推算法, 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外推生成多样性的样本, 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：

图像外推 (Image extrapolation) 是一个由部分图像生成更大尺寸图像的问题，已知的部分图像通常认为是含有物体或者场景的部分信息，以此作为条件生成超出已知部分图像边界的内容，旨在合理地将图像的内容扩展到部分图像范围之外，得到更大尺寸的图像。例如四周被遮挡的物体可以恢复整体外观或者构造全景图。图像外推问题面临两个方面的挑战，一个是生成内容与已知的部分图像保持空间结构和语义内容的一致性；另一个是外推生成的图像具有较高的图像质量和近乎真实的效果。

对于有着丰富经验的人类来说，将图像内容合理地扩展到现有范围之外不是一件难事。1989 年特拉华大学的 Helene Intraub 和 Michael Richardson 等人实验证明，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的范围 [1]。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 [2]。这些预测和外推对于合成完整图像起到重要作用并且有着广阔的应用前景。

早期的图像外推通常采用的是图像检索和合成的方法，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于 2013 年首次提出图像外推任务 [3] 并且通过照片得到完整全景图像。通过给定要外推的源图像、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提出数据驱动的方法 [4]，给定要外推的源图像以及要外推的大小和方向，使用库图像为外推区域确定视觉上连续一致的内容。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，达到扩大照片视野效果 [5]。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，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，很难保证外推边界空间结构对齐和语义一致性。

近年来，深度学习的发展为图像生成带来了质的飞跃，以深度学习为基本思想的生成模型不断发展壮大。2014 年，Goodfellow 等人提出生成式对抗网络 (Generative adversarial networks, GAN) [6]。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，训练时固定一方，更新另一个模型的参数，交替迭代，使得对方的错误最大化，最终，G 能估测出样本数据的分布。生成式对抗网络 (GAN) 技术是一种域到域的转换技术，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。生成器以高斯向量为输入，经过反卷积神经网络生成样本。判别器以生成的样本和真实样本作为输入，通过卷积神经网络判断输入的样本是否为真实样本。生成器尽可能生成类似真实的样本让判别器判错，而判别器尽可能做出正确的判断，从而形成了对抗模型，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) [7],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) [8],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、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。WGAN 用一个惩罚函数 f 来对应 GAN 的判别器,而且惩罚函数 f 需要建立在 Lipschitz 连续性假设上。WGAN 开创性地解决 GAN 训练不稳定的问题,基本解决模式崩塌,确保了生成样本的多样性。

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 提出基于 DCGAN 网络架构的深度学习方法用于图像外推任务 [9],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 128×128 大小,超出输入源图边界的外推部分存在模糊。香港中文大学 Yi Wang 等人 [10]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扩后的预测生成图,用生成图与真图成对数据进行训练,在人脸等物体数据集的四周外扩和巴黎街景等场景数据集上的单边外扩实验上得到比较好的效果。谷歌公司的 Piotr Teterwak 等人 [11] 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。此方法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 [12] 在生成器中加入水平跳跃短接 (SHC) 和递归内容传输 (RCT) 模块,并且可以通过多步生成外扩更大的区域。尽管由源图像进行外推理应产生多样合理的结果,但是这些方法在训练时,生成图仅用唯一对应匹配的真图进行约束,对于每一个要外推图像只生成一个最优结果。

为了生成多样的图像外推结果,2020 年宾夕法尼亚大学的张灵志等人提出新的正则化项 [13],在输入部分前景图像时可以得到多个的背景,生成多样性的图像外推结果。

图像补全 (Image completion) 与我们图像外推紧密相关,均可以看作是条件图像生成任务,不同的是图像补全关注于图像局部缺损内容的生成。在提高图像补全结果多样性上,新加坡南洋理工大学 Chuanxia Zheng 提出两个并行路径框架的网络结构实现多样性的图像补全结果 [14]。一个是重建路径,利用给定的真图来得到缺失的部分先验信息,并从该分布重建原始图像;另一种是生成路径,耦合条件先验与重构路径中的分布。此方法在多个数据集的不同掩码位置可以生成多样性的图像补全结果。但是图像补全的算法会依据较多的已知信息补全图像内较少的缺失区域,当用于解决生成已知局部区域外的图像内容,在远离已知局部图像的区域会出现内容模糊和细节缺失的现象,因此很难直接用于解决图像外推问题。

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] Intraub H, Richardson M. Wide-angle memories of close-up scenes[J]. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1989, 15(2): 179.
- [2] Lyle K, Johnson M. Importing perceived features into false memories[J]. Memory, 2006, 14(2): 197-213.
- [3] Zhang Y, Xiao J, Hays J, et al. Framebreak: Dramatic image extrapolation by guided shift-maps[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 1171-1178.
- [4] Wang M, Lai Y, Liang Y, et al. Biggerpicture: data-driven image extrapolation using graph matching[J]. ACM Transactions on Graphics, 2014, 33(6).
- [5] Shan Q, Curless B, Furukawa Y, et al. Photo uncrop[C].European Conference on Computer Vision. Springer, Cham, 2014: 16-31.
- [6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C].Advances in neural information processing systems. 2014: 2672-2680.
- [7] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. International Conference on Learning Representations, 2016.
- [8] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. arXiv preprint arXiv:1701.07875, 2017.
- [9] Sabini M, Rusak G. Painting outside the box: Image outpainting with GANs[J]. arXiv preprint arXiv:1808.08483, 2018.
- [10] Wang Y, Tao X, Shen X, et al. Wide-Context Semantic Image Extrapolation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1399-1408.
- [11] Teterwak P, Sarna A, Krishnan D, et al. Boundless: Generative Adversarial Networks for Image Extension[C].Proceedings of the IEEE International Conference on Computer Vision. 2019: 10521-10530..
- [12] Yang Z, Dong J, Liu P, et al. Very Long Natural Scenery Image Prediction by Outpainting[C].Proceedings of the IEEE International Conference on Computer Vision. 2019: 10561-10570..
- [13] Zhang L, Wang J, Shi J. Multimodal Image Outpainting With Regularized Normalized Diversification[C].The IEEE Winter Conference on Applications of Computer Vision. 2020: 3433-3442..
- [14] Zheng C, Cham T J, Cai J. Pluralistic image completion[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1438-1447.

三、研究内容

1. 学术构想与思路、主要研究内容及拟解决的关键技术或问题

学术构想与思路：

首先，针对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学术成果。其次分析图像外推的特殊性，结合图像修复等相关工作，探索和设计出适合图像外推的生成模型的网络结构，并提出改进和创新思路。最后设计实验方案，根据实验结果验证自己提出的假设和创新点，根据实验提出新的创新点。

主要研究内容：

- (1) 针对图像外推的具体问题做全面的国内外先进成果调研。
- (2) 研究图像外推中生成内容模糊以及空间结构和语义内容一致性的原因。
- (3) 研究生成式对抗网络（GAN）模式崩塌的原因。

拟解决的关键技术或问题：

- (1) 由已知的部分图像作为条件，生成更大的外推结果，保证生成部分与已知部分在空间结构和语义内容上的一致性。
- (2) 解决生成模型中存在的模式崩塌现象，提高在图像外推问题上的多样性问题。

2. 拟采取的研究方法、技术路线或研究步骤、实施方案及可行性分析

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模型技术。了解生成模型的发展历史和现状，进而帮助确定研究课题。

（2）比较法。对现有的经典的深度估计算法进行概括总结，并比较其优缺点。

（3）探索性研究法。针对图像生成的具体问题的具体特性以及目前先进的图像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。

研究步骤与实施方案：

（1）设计实验方案，处理数据集。

本文使用卷积神经网络解决图像生成问题，将生成模型考虑为编码-解码结构，选择使用生成式对抗网络。数据集选择CelebA-HQ人脸数据集和Paris StreetView场景数据集。

图1. CelebA-HQ数据集，图像格式为JPEG，图像大小是256×256，一共30000张图像，其中27000张作为训练集，600张作为验证集，2400张作为测试集。



图1. CelebA-HQ数据集样本

（2）分析问题，设计网络模型。目前简单设计的网络模型如图2所示。

该模型由两部分构成：生成器网络和判别器网络。生成器采用类似encoder-decoder编码-解码器结构，用来生成图像外推数据来欺骗判别器。判别器用来区分生成的图像是否属于真实的图像分布，网络通过生成器和判别器之间的博弈进行训练。

图2. 模型网络结构图

(3) 训练模型，并测试当前模型的效果。根据测试结果提出新的问题并改进当前模型。

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N 卡的服务器上训练测试，满足算力要求。

3. 论文的创新之处

- (1) 本文提出了一种新的图像外推算法, 生成高质量和多样性的图像数据。
- (2) 本文采用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。
- (3) 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。

4. 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) 改进出生成高质量与高多样性的图像外推结果的模型方法。
- (3) 通过实验定性和定量的分析当前方法的优缺点。

四、论文安排

序号	阶段及内容	起迄日期	阶段成果形式
1	基础学习阶段： 学习基本图像处理和深度学习知识，阅读前沿论文，掌握计算机视觉发展动向；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。
2	课题准备阶段： 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。
3	关键技术学习、研究阶段： 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— 2020 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。
4	实验阶段： 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— 2020 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。
5	论文撰写阶段： 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
6	论文完善阶段： 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。	2021 年 1 月— 2021 年 3 月	论文定稿

五、评审意见

导师意见:

导师签字: 年 月 日

审核小组意见:

(论文内容与该生业及选定的研究方向相关)

审核小组成员签字: _____

年 月 日

学位评定分委员会对研究生开题报告的意见:

学位评定分委员会主席签字:

(学院盖章)

年 月 日

