

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重要且具有挑战性的问题，两者均是由已知局部信息生成完整的图像内容，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，语义和空间结构等连续而且在整体上融合一致，得到符合人类视觉感知的图像效果，一直是研究者们努力的方向。

图像内补起源于对缺损的文物或艺术品的修复，是在图像内部缺失区域生成合理的结果，为缺失区域生成逼真的纹理细节，并且保持边界的连贯性，传统上主要采用扩散和样本块的方法。2000 年 Marcelo Bertalmio 等人[1]提出扩散方法将周围区域外观特征从边界平滑地传播到缺失区域的内部来合成纹理，利用缺失区域边缘的有效像素信息，使用偏微分方程进行扩散传播来填充缺失区域。但是这种方法只能处理修复任务中的小孔，无法生成有意义的语义结构。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良好的替换补丁，可以利用较远的图像信息来填补缺失区域。为了找到合适的图像补丁，2008 年 Denis Simakov 等人 [2] 为找到合适的图像补丁块，提出了双向相似度方法，在计算补丁相似度时可以捕获更多的视觉信息，引入更少的视觉伪影现象。2009 年 Connelly Barnes 等人 [3] 提出 PatchMatch 方法，通过从源图像的未损坏区域搜索并复制相似的图像补丁来填充目标区域，使用图像中的自然相干性作为先验信息，通过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可以为较大的缺失区域生成逼真的纹理，由匹配局部特征可以改善修复效果。但是，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，在某些具有重复结构的图像中效果较好，但是缺乏对语义内容的理解，对于具有独特结构的图像却无法生成合理的图像结果。

图像外推是将部分图像内容合理地扩展到边界范围外，通过已知的局部信息恢复整体图像。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，以此作为条件生成图像扩充需要的内容，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。例如局部被遮挡的物体可以恢复整体外观或者构造全景图。图像扩充问题面临两个方面的挑战，一个是图像扩充内容与已知的部分图像能够保持空间结构和语义内容融合的一致性；另一个是扩充部分的图像具有较高的图像质量，达到合成图像近乎真实的视觉感知效果。

人类可以根据自己的丰富生活经验，将图像内容补充完整并且合理地扩展到边界范围之外，从而得到更大尺寸的图像。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 [4]。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]。根据预测扩充图像内容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。

早期的图像扩充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于 2013 年首次提出图像外扩任务 [6]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 年清华大学的汪淼等人 [7]提出数据驱动的方法，给定要外扩的源图像以及要外扩的大小和方向，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扩充效果。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。

为使合成图中的局部图与整体背景融合，Daniel Cohen-Or 采用颜色协调的方法 [9]，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的色调值。Patrick Pérez 等人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 [10]。虽然这种方法可以实现无缝拼接，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。

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) [11]。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) [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、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。

许多基于深度学习的方法将修复任务建模为条件生成问题，其中显著的优势是能够从损坏的图像中提取有意义的语义信息，并为图像生成新的内容。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，为丢失的孔生成近乎真实的替代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建立长期相关性方面无效的。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借用特征。原始 (普通) 卷积在所有有效，无效和混合 (例如，那些空洞边界) 像素/特征上应用相同的滤波器，从而导致视觉伪像，例如，在

自由形状上的掩码做测试时,导致视觉的伪影(如颜色差异,模糊和孔周围明显的边缘响应)。为了解决这一局限性,Guilin Liu 等人 [17]部分卷积 (PartialConv) 以强制网络仅使用有效像素(未破坏的像素),将所有位置视为无效或有效,并用 0 或 1 掩码乘以所有层的输入,避免在输入的掩蔽孔中替换像素会为最终结果引入伪影。启发式的将所有的位置归类为无效或有效,会忽略这些重要信息(输入图像的有效像素,掩蔽像素,感受野的神经元没有覆盖输入的有效像素图像)。要在孔中合成像素,这些深层需要有关当前位置在孔内还是孔外的信息。使用掩码进行部分卷积无法提供此类信息。2019 年余家辉等人 [18] 提出 SN-PatchGAN 的门控卷积,适用于自由形状图像修复的门控卷积算法,它学习每个通道和每个空间位置的动态特征门控机制,通过扩展特征选择机制使其在所有层中的每个位置均可学习。Kamyar Nazeri 论文[19]中提出的 EdgeConnect 通过使用其他先验信息来生成合理的结构,首先恢复边缘贴图,然后详细填充缺失的区域。但是由于边缘图的表示能力有限,该方法可能会在对象的边界中生成错误的细节。

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 提出基于 DCGAN 网络架构的深度学习用于图像外推任务 [20],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 [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 [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。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 [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,并且可以通过多步生成外扩更大的区域。

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] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester. Image Inpainting. In Proceedings of the 27th annual conference on Computer graphics and in-teractive 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