

图像修复的研究发展综述

摘要

图像修复是传统计算机视觉问题之一，近年来，随着深度学习技术的发展，许多深度学习技术也被应用到图像修复领域中，同时，随着图像处理工具的改进，人们对图像修复也提出了更高要求，这也使图像修复成为图像处理领域一个重要而富有挑战性的课题。本文首先介绍了图像修复的意义，然后介绍了图像修复中常用的生成对抗神经网络（GAN）；然后我们研究了国内外在图像修复领域的研究现状，总结了现有的修复方法，并将它们分为三类，即基于序列的方法，基于 CNN 的方法，基于 GAN 的方法。此外，我们也对比了每种方法的优点和缺点。

关键词：图像修复，GAN，CNN

一. 介绍

如今图像是最常见的信息形式之一，他们被应用在生活的各个领域。社交网络中的共享图像可以包含许多添加到这些图像中的对象，包括签名、矩形等，这些对象的加入会改变图像的语义；人们在日常生活中拍照，也容易受到路人的影响，造成照片出现拍摄者不希望出现的物体；还有一些图片由于传输，保存不当等原因，会出现一些破损。这些情况都使得图像修复成为近年来人们的研究热点任务之一。

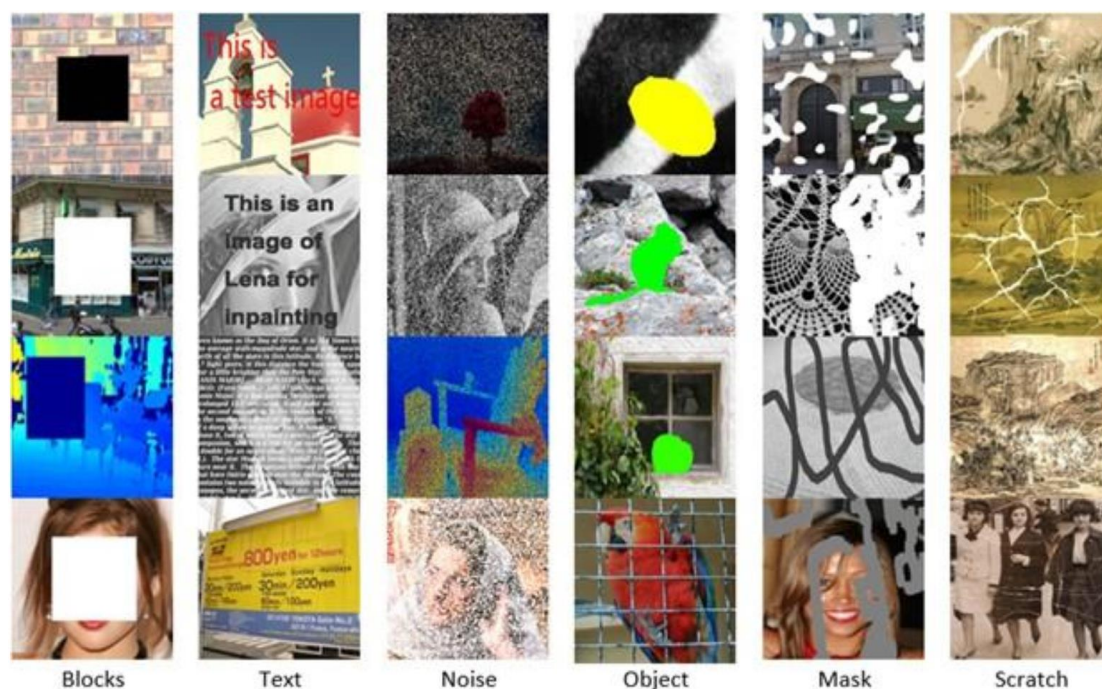


图 1.破损图片类型^[24]

基于图像的已知信息来还原图像中缺失部分是图像修复的核心思想，早期的传统算法是通过外部数据库的搜索方法^[1]，这些方法是通过在包含大量相似的图片中，搜索与待修复图像具有足够的视觉相似性的样本，这种方法是非常有效的，但是当修复的图像与数据库中没有很好的表示时，会出现修复错误，最终使得修复效果非常不理想。

在深度学习的研究中，人们证明了卷积神经网络^[2]（CNN）在高层上提取的图像抽象信息可以作为目标函数的一部分，使得一个生成网络生成的图片与目标图片在语义上更加地相似。加上人们对生成对抗神经网络^[3]的研究不断深入，通过对抗训练可以强化生成图像的视觉效果，使得 GAN 成为图像修复领域的常用技术手段之一。

二. 相关基础

2.1 基于生成对抗的神经网络(GAN)

GAN 是 Goodfellow^[3]在 2014 年提出的一种新型生成网络，并于 2016 年，Pathak 等人^[4]提出 Context Encoder 模型，首次将对抗生成网络引入到图像修复中去，给图像修复领域带来新的技术手段。经过最近几年的研究，主流的 GAN 修复模型如图 2 所示。

模型主要分为 2 部分，第一部分采用编码-解码器结构作为生成器，第二部分使用多层卷积神经网络作为判别器。第一部分中，首先使用卷积神经网络进行编码，将图像映射到高层语义层中，并逐步修复。然后采用反卷积^[5]的方式，将图像从高层语义层中还原到低层语义中。在解码器与编码器的过渡层，研究者们经常使用空洞卷积^[6]的方式，以便于增大卷积的视野域，从而能够减少修复的痕迹。第二部分是由多层卷积层搭建的判别器模型，为了提高训练的稳定性，Iizuka 等人^[7]提出了多尺度判别器模型，即在采用两个判别器，一个是全局判别器，用于提高图像的整体效果，另一个是局部判别器，用于提高图像的细节效果，这种技术手段后来被人们广泛应用于自己的模型中。

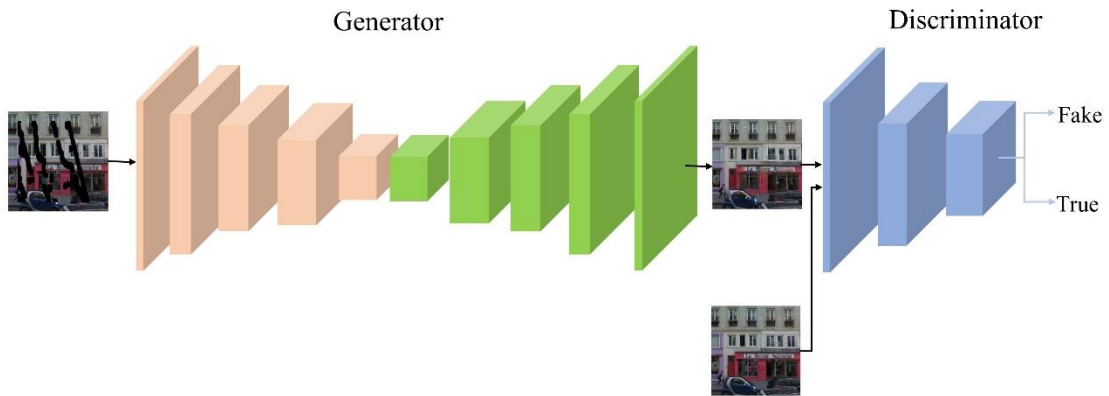


图 2.GAN 结构图

2.2 生成对抗神经网络的训练

GAN 是一种包含两个前馈网络的网络结构，分别是生成器 G 和判别器 D，生成器把损失的图像作为输入，生成无限接近原图的生成图像，判别器把原图和生成图像作为输入，判断样本是否是真和假。

GAN 中的重构损失（reconstruction loss）用于训练生成器，使生成器能够生成和目标图像无限接近的图片，GAN 的对抗损失（adversarial loss）用于训练判别器，使判别器能够识别出哪张图片是真实图片，哪张图片是生成图片，整个 GAN 网络在上述的博弈中训练，其目标函数如公式 1 所示。

$$\min_G \max_D E_{x \in \chi_{data}} [\log D(x)] + E_{z \in Z} [\log (1 - D(G(z)))] \quad (1)$$

王等人^[8]引用图像风格转换中的直觉损失（perceptual loss），这种损失计算了高层次语义的相似度，即

$$L_{perc} = E \left[\sum_i \frac{1}{N_i} \|\phi_i(I_{gt}) - \phi_i(I_{pred})\|_1 \right] \quad (2)$$

提高了图片的修复质量。Liu 等人^[9]在 2018 年使用 PC 卷积的方式进行图像修复，在他们的工作中，使用风格损失（style loss）和直觉损失共同训练生成器，大大的提高了图像修复的质量。

三. 方法

3.1 基于序列的算法

基于序列的方法可以分为两类：一类是基于仿射的方式，另一类是基于补丁的方式。

基于仿射的方式^[10, 11]填充损失区域，主要是利用周围的已知区域，采用平滑的方式填充未知区域。李等人^[10]提出了一种基于仿射的图像修复方法，首先对修复区域进行局部扩散，然后根据变化在通道内和通道间的局部方差构建一个特征集来识别修复区域；而 Sridevi 等人^[11]提出了一种基于分数阶导数和傅里叶变换的仿射图像绘制方法。但是这种方式仅仅能够处理损失区域较小的情况。

基于补丁的方式^[12, 13]是在未损失的部分寻找和损失部分相似的补丁，然后将这些补丁复制到损失区域，这种方式能够修复非常大的图像缺失区域。为了找到最相似的补丁，Simakov 等人^[14]提出了双向相似度来获取更多的视觉信息，在计算补丁的相似度时引入较少的视觉伪影。同时，为了降低计算成本，Nazeri^[15]设计了一种图像中的自然连续性作为先验信息的快速最近邻搜索算法。然而，在一些任务像人脸图像的绘制，这种假设损失区域和未损失区域具有相似的语义内容方法是不可取的。

3.2 基于 CNN 的方法

基于 CNN 的方法在修复图像细节上，有着传统方法不可比拟的优势。ShiftNet^[24]采用 U-Net^[16]结构，利用一种叫做 shift 操作和一个指导损失。这种 shift 操作可以推测编码器层中的已知区域和解码器层中的相关损失区域之间的关系，能够生成比较精细的结构和图像纹理；刘等人^[17]提出了一种叫做连续语义注意力机制（CSA），这种方法不仅考虑到了损失区域与未损失区域像素之间的相关性，也考虑到了损失区域像素之间的相关性，这种方法的网络结构如图 3 所示；曾等人^[18]提出了一种金字塔结构叫做 PEN-NET 用于高分辨率图像的修复。

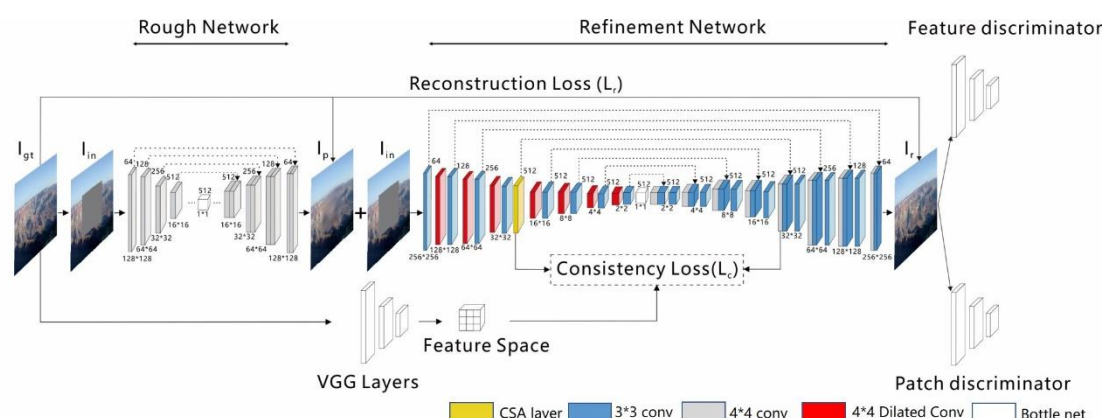


图 3.CSA 模型结构图

3.3 基于 GAN 的方法

Context-encoder^[4]第一次把对抗训练的方式引入到图像修复领域，但是这种方式修复的图像分辨率较低，对于细节的修复也不能令人满意；Iizuka 等人^[7]发展了这项工作，提出了多尺度判别器提高了图像的修复质量，但是这种方式需要对修复后的图像进行后处理；杨等人^[19]将 Context-encoder 的结果作为输入，逐步增加纹理细节，得到了高分辨率预测，但是这种方法的计算量很大。

近年来，随着人们对深度学习的研究不断深入，人们提出了许多修复质量更好的模型。Yu 等人^[20]提出一种双阶段修复图像过程，将图像修复过程分为两个阶段，第一阶段是粗修复，先大致修复图像的全局，第二阶段是精修复，根据第一阶段的结果，继续修复图像的具体细节，这种方式极大地提高了图像的修复质量；Kamyar 等人^[15]提出 EdgeConnect 网络，这种网络依然采用两个阶段进行图像修复，不过在第一阶段只修复图像的轮廓，在第二阶段进行图像的上色，如图 4 所示；Ren 等人^[21]根据 Kamyar 等人的思想，提出了 structureflow 网络，他们在第一阶段先修复图像的光流图，再根据图像的光流图进行图像的具体细节的修复。虽然这些方法的修复质量得到了极大的提升，但是网络的测试速度比较慢。Sagong 等人^[22]提出一种效率更快的网络，叫做 PEPSI，这种网络在解码器部分

分为两个路径，一个路径做粗修复，另一个路径采用 Contextual attention 模块做精修复，两个路径的权重值共享，极大的加快了模型的运行速度，但是与此相对应的，也丢失了一部分的图像修复质量。

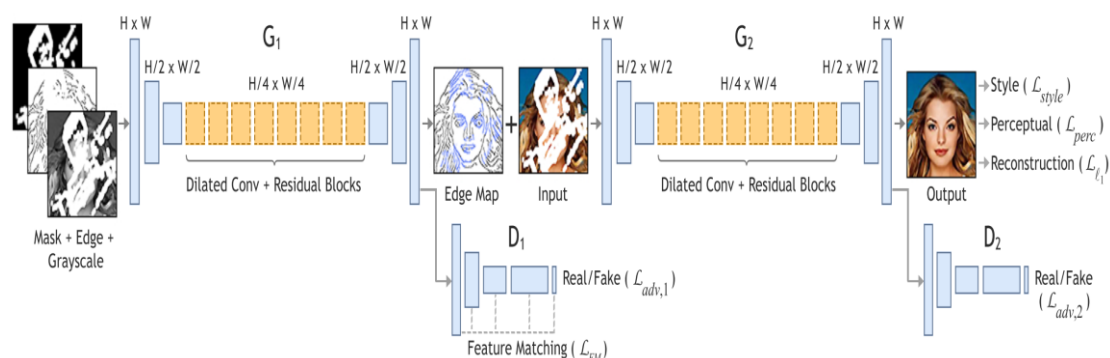


图 4.Edgeconnect 网络结构图

四. 结论

图像修复随着深度学习技术的发展，已经愈加成熟，但是基于对上述方法的讨论，我们也可以看到，目前深度学习在图像修复的领域依然不成熟，在未来会有更多更广泛的尝试，下面基于对图像修复领域的研究提出几点展望。

(1) 虽然随着网络复杂程度的加深，图像的修复质量逐渐提高，但是网络的测试时间也逐渐加上，虽然人们对于缩短时间也进行了积极探索，但是低耗时的网络结构带来的是修复质量的下降。因此，能否有设计一种网络，使其图像的修复质量好的同时能够保证时间消耗少。

(2) 当前的图像修复全部集中在 256×256 的图像种，对于更高级别的分辨率图像修复的探索还是一个盲区，因此，能否将网络应用到高分辨率像素依然是有待解决的问题。

参考文献

- [1] Hays J, Efros A A. Scene completion using millions of photographs[J]. ACM Transactions on Graphics (TOG), 2007, 26(3): 4.
- [2] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
- [4] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2536-2544.
- [5] Pan J, Sayrol E, Giro-i-Nieto X, et al. Shallow and deep convolutional networks for saliency

prediction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 598-606.

[6] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.

[7] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics (ToG), 2017, 36(4): 107.

[8] Wang C, Xu C, Wang C, et al. Perceptual adversarial networks for image-to-image transformation[J]. IEEE Transactions on Image Processing, 2018, 27(8): 4066-4079.

[9] Liu G, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 85-100.

[10] Guo Q, Gao S, Zhang X, et al. Patch-based image inpainting via two-stage low rank approximation[J]. IEEE transactions on visualization and computer graphics, 2017, 24(6): 2023-2036.

[11] Sridevi G, Kumar S S. Image inpainting based on fractional-order nonlinear diffusion for image reconstruction[J]. Circuits, Systems, and Signal Processing, 2019: 1-16.

[12] Duan J, Pan Z, Zhang B, et al. Fast algorithm for color texture image inpainting using the non-local CTV model[J]. Journal of Global Optimization, 2015, 62(4): 853-876.

[13] Lu H, Liu Q, Zhang M, et al. Gradient-based low rank method and its application in image inpainting[J]. Multimedia Tools and Applications, 2018, 77(5): 5969-5993.

[14] Simakov D, Caspi Y, Shechtman E, et al. Summarizing visual data using bidirectional similarity[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.

[15] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Generative image inpainting with adversarial edge learning[J]. arXiv preprint arXiv:1901.00212, 2019.

[16] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[17] Liu H, Jiang B, Xiao Y, et al. Coherent Semantic Attention for Image Inpainting[J]. arXiv preprint arXiv:1905.12384, 2019.

[18] Zeng Y, Fu J, Chao H, et al. Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 1486-1494.

[19] Yang C, Lu X, Lin Z, et al. High-resolution image inpainting using multi-scale neural patch

synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6721-6729.

[20] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5505-5514.

[21] Ren Y, Yu X, Zhang R, et al. StructureFlow: Image Inpainting via Structure-aware Appearance Flow[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 181-190.

[22] Sagong M, Shin Y, Kim S, et al. Pepsi: Fast image inpainting with parallel decoding network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 11360-11368.

[23] Yan Z, Li X, Li M, et al. Shift-net: Image inpainting via deep feature rearrangement[C]//Proceedings of the European Conference on Computer Vision. 2018: 1-17.

[24] Elharrouss O, Almaadeed N, Al-Maadeed S, et al. Image inpainting: A review[J]. Neural Processing Letters, 2019: 1-22.