Modeling and Simulation 建模与仿真, 2024, 13(6), 5934-5941
Published Online November 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.136541

Hans汉斯

基于生成对抗网络的街景图像修复模型

朱　莹

南京邮电大学理学院，江苏　南京

收稿日期：2024年10月15日；录用日期：2024年11月8日；发布日期：2024年11月14日

摘　要

针对现有街景图像修复方法难以有效处理复杂遮挡和保持语义一致性的问题，本文提出了一种基于生成对抗网络(GAN)的图像修复模型，旨在利用语义信息提升修复效果。该模型能够有效地提取和处理不同类别的语义信息，并将其融入图像修复过程，利用语义信息引导生成器生成更逼真、更符合语义的修复结果。在街景数据集上的实验结果表明，相较于传统方法，本文提出的模型能够生成更加自然逼真的修复结果，可以精准地还原车辆、道路、建筑物等物体的轮廓、纹理细节和颜色，有效消除图像缺失部分带来的视觉突兀感，显著提升图像的整体质量。未来可进一步拓展该模型的应用范围，将其应用于人脸图像以及其他类型场景的图像修复任务中，以期提升图像修复模型的泛化能力和修复效果。

关键词

街景图像修复，语义先验，生成对抗网络

A Streetview Image Inpainting Method Based on Semantic Prior Guidance

Ying Zhu

School of Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Oct. 15th, 2024; accepted: Nov. 8th, 2024; published: Nov. 14th, 2024

Abstract

This paper proposes a novel generative adversarial network (GAN)-based image inpainting model for street view images, aiming to leverage semantic information to enhance inpainting quality. The model effectively extracts and processes different categories of semantic information, integrating them into the inpainting process to guide the generator in producing more realistic and semantically consistent results. Experimental results on street view datasets demonstrate that, compared to traditional methods, our proposed model generates more natural and visually appealing inpainted

文章引用: 朱莹. 基于生成对抗网络的街景图像修复模型[J]. 建模与仿真, 2024, 13(6): 5934-5941.
DOI: 10.12677/mos.2024.136541

images. It accurately restores the contours, textures, and colors of objects such as vehicles, roads, and buildings, effectively eliminating the visual abruptness caused by missing parts and significantly improving the overall image quality. In the future, we plan to extend the application of this model to face images and other types of scenes to improve the generalization ability and inpainting performance of image inpainting models, enabling them to better address diverse image inpainting needs across various scenarios.

Keywords

Streetview Image Inpainting, Semantic Prior, Generative Adversarial Network

1. 引言

图像修复旨在填充图像中缺失或损坏的部分，使其恢复完整性和美观性。它广泛应用于旧照片修复、物体移除、图像编辑等领域。街景图像修复则是将图像修复技术应用于街景图像的一种特定场景。街景图像通常由车辆或人员携带的相机拍摄，由于拍摄环境的复杂性，经常因为遮挡导致图像信息丢失。同时拍摄时也会捕捉到街道上的行人。街景图像修复除了提升图像质量与可用性外，还在保护隐私方面发挥着重要作用，尤其是去除行人，防止个人信息泄露。街景图像修复任务可以看作是图像修复任务的一个子集，继承了图像修复的基本方法与思想，但是同时也面临着场景复杂、图像信息量大等独特问题。

生成对抗网络(GANs) [1]的强大生成能力可以生成逼真的图像内容，填补缺失区域，使修复后的图像更加自然。而他的对抗性训练机制促使生成器不断改进生成图像的质量，从而更接近真实图像。许多研究[2] [3]利用 GANs 在图像修复任务中一系列令人瞩目的成果。对于大面积缺损或复杂结构的缺损，普通方法难以获得令人满意的修复效果。许多方法通过引入结构[4]、纹理[5]或语义等先验信息对图像修复任务进行指导。基于先验信息的修复方法可以学习到图像的通用特征和规律，帮助模型推断缺损区域的内容，使其在面对不同类型的图像和缺损时都能够取得较好的修复效果。先前，语义先验引导的图像修复方法发展受到语义分割图准确性和即时获取的限制。为了克服这一问题，研究者们提出了隐式语义先验的概念，即在图像修复过程中由模型自身模块自动提取的语义信息。然而，这种隐式语义先验通常较为粗糙，且无法可视化，导致难以评估其准确性。本文针对街景图像的大面积缺失与复杂场景两个难题，提出了一个基于生成对抗网络的街景图像修复模型。在语义图的指导下利用图像不同区域的语义特征，在修复过程中更精准地控制纹理和颜色，使修复后的图像更加自然，从而提升街景图像修复任务的性能。

2. 方法

2.1. 语义图像修复模型

我们将语义分割标签作为额外信息融入条件生成对抗网络，构建了一种能够修复大面积缺失区域并保持整体语义和场景一致性的新模型。本文提出的基于语义先验引导的街景图像修复模型，主要分为两部分，在第一部分语义图像修复模型中，需要对损坏图像 I_0 的语义分割图 S_0 进行修复。使用如下等式获得语义分割图的修复结果 S_1：

$$S_1 = G_{sm}\left(S_0, M\right) \tag{1}$$

其中 G_{sm} 表示语义图像修复模型中的生成器，M 表示二进制遮掩图像，如图 1 中所示，白色部分表示遮掩区域，黑色部分则表示已知区域。语义图像修复部分的网络模型由编码器、瓶颈层和解码器三部分组成。编码器包含四个下采样块，用于提取图像特征表示；瓶颈层由九个残差块[6]构成的残差网络构成，用于进一步处理和整合特征信息，九个残差块的设置能够有效地平衡模型复杂度和计算资源，从而更好地满足我们的任务需求；解码器则包含四个上采样块和一个激活层，用于生成最终的语义修复图像 S_1。

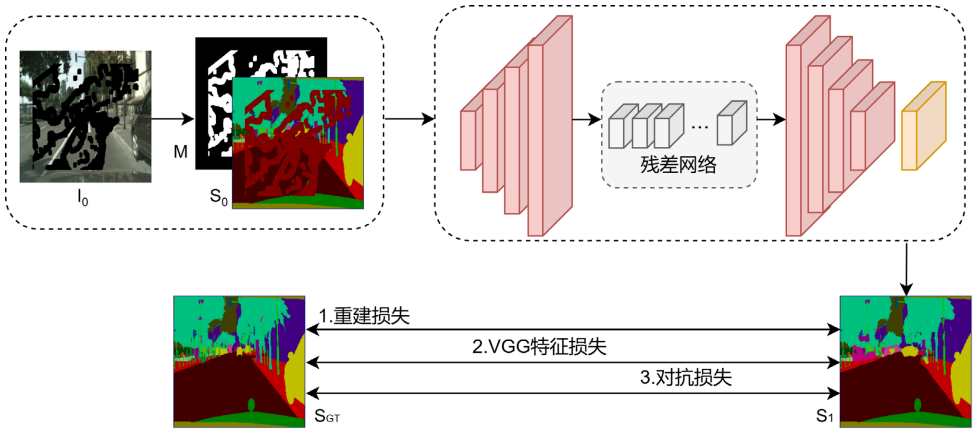

Figure 1. The overall architecture of the semantic inpainting model
图 1. 语义图像修复模型的总体架构

2.2. 生成式图像修复模型

接着在第二部分生成式图像修复模型中，通过利用修复好的语义分割图 S_1 作为先验指导对损坏图像 I_0 进行修复，最后获得修复结果 I_1，如下式所示：

$$I_1 = G_{im}\left(I_0, S_0, M\right) \tag{2}$$

其中 G_{im} 表示生成式图像修复模型中的生成器。在该部分模型生成器训练过程中引入语义信息作为先验知识，使模型能够学习并记住各种语义信息，从而更准确地处理图像中的不同元素。尤其是在处理复杂的图像背景时，模型的性能得到了显著提升，最终能够生成更高质量的修复结果。

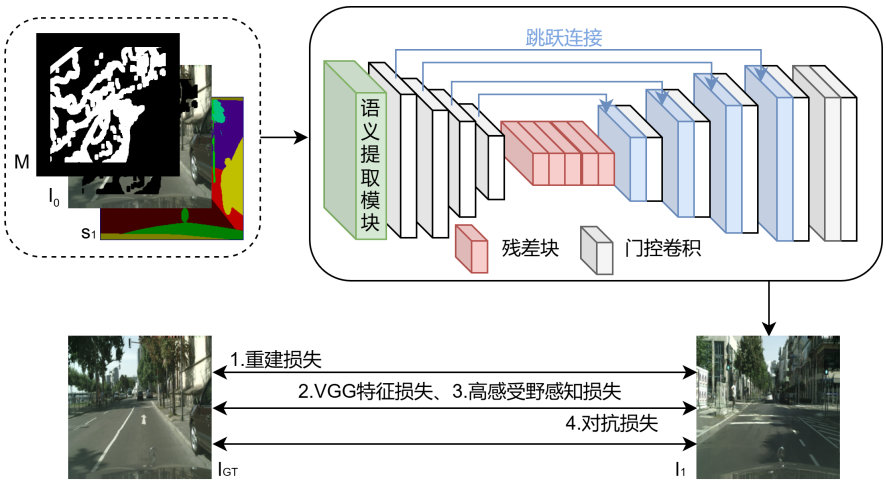

Figure 2. The overall architecture of the generative inpainting model
图 2. 生成式图像修复模型的总体架构

生成式图像修复模型中，本文采用的网络结构如图 2 所示：在网络的开始部分，本文设计了一个语义提取模块，该模块能够从语义图中提取不同类别物体(例如建筑物、天空、道路、树木等)的语义信息，并将其存储下来，以便后续使用。该模块语义提取与存储的可视化效果如图 3 所示，在该图中输入图像首先根据语义图将不同类别的语义信息分别学习存储，在图像修复过程中再通过存储的语义类别特征的指导生成该模块结果 $I_{a/}$。

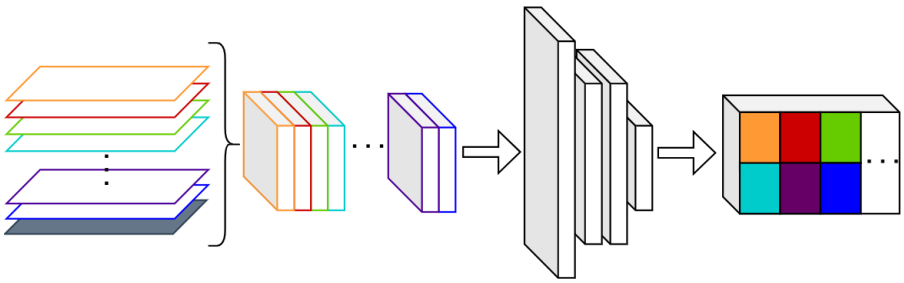

Figure 3. The visualization effect of semantic extraction module
图 3. 语义提取模块的可视化效果

在经过语义提取模块后，后面的网络主要分为编码器、瓶颈层以及解码器三部分。其中编码器解码器主要由带有跳过连接的四组上采样和下采样块构成。长跳过连接可以将信息从编码器传播到解码器，有助于恢复在下采样期间丢失的信息。在编码器的末端，感受野比输入图像的分辨率(256 × 256)大得多。感受野的扩大有利于提升整个结构的实现效果。在该模块的网络构成中，本文用门控卷积替换了所有普通卷积，旨在利用其门控机制控制信息流动，选择性传递有用特征。与普通卷积相比，门控卷积在图像修复方面具有明显的优势。普通卷积可能导致信息丢失，缺乏长期依赖性，难以处理非线性修复问题，并且在感知动态变化和处理不规则区域方面表现不佳。门控卷积更有利于处理复杂、非线性或依赖上下文一致性的图像修复任务。因此，我们在模型中使用门控卷积来代替普通卷积。在门控卷积模块中，

首先将卷积操作 $\mathrm{Conv}\left(I_{\varepsilon}\right)$ 的输出沿通道维度(用 $d=1$ 表示)分成两部分：

$$I_{g1}, I_{g2} = \mathrm{Split}\left(\mathrm{Conv}\left(I\varepsilon\right),\ d=1\right) \tag{3}$$

接下来，我们分别对 I_{g1} 和 I_{g2} 应用激活函数，然后将它们相乘。最后，最终输出经过批量归一化层和 ReLU 层。这个过程可以用以下公式表示：

$$I_{gated} = \mathrm{Re}\,LU\left(\mathrm{BatchNorm}\left(ELU\left(I_{g1}\right)\odot\sigma\left(I_{g2}\right)\right)\right) \tag{4}$$

其中，ELU 代表指数线性单元激活函数，而 σ 代表 sigmoid 激活函数。

3. 模型训练

如图 2、图 3 所示，在本文提出的基于语义先验引导的街景图像修复模型中，两部分主要使用了四种损失函数，分别是重建损失、对抗损失、VGG 特征损失以及高视野感知损失，并使用超参数来平衡不同损失。

3.1. 重建损失

在训练网络的过程中，我们需要计算真实图像 S_{GT}/I_{GT} 和修复图像 S_1/I_1 之间的重建损失，使用的是 l_1 损失函数。该损失函数包含两部分，分别对应缺失区域和非缺失区域的重建误差，并通过不同的权重系数进行调节。当使用规则的矩形掩码进行训练时，我们更加关注缺失区域的重建效果；而当使用不规则

掩码进行训练时，我们则需要兼顾缺失区域和非缺失区域的重建效果，以获得更好的整体修复质量。如公式(5)所示：

$$L_{re}=\lambda_r^1\left\|I_1-I_{GT}\right\|_1\odot M+\lambda_r^2\left\|I_1-I_{GT}\right\|_1\odot\left(1-M\right)。\tag{5}$$

3.2. 对抗损失

本文在 pix2pix-HD [7]的基础上，将条件生成对抗网络(CGAN)引入到图像修复任务中，并利用语义信息提升修复效果。为了进一步提升模型的训练效果和生成图像的质量，本文在判别器中采用了多尺度对抗损失。多尺度对抗损失是指在多个不同的图像尺度上计算对抗损失，从而促使生成器在各个尺度上都生成逼真的图像。具体来说，我们将判别器设计成多尺度的结构，例如包含 D_1,D_2,D_3 三个判别器，分别接收不同分辨率的图像作为输入。每个判别器都会计算生成图像与真实图像之间的对抗损失，并将这些损失加权求和得到最终的多尺度对抗损失如下式所示：

$$\min_G\max_{D_1,D_2,D_3}\sum_{k=1,2,3}L_{GAN}\left(G,D_k\right)=\sum_{k=1,2,3}E\left[\log\left(D_k\left(I_1\right)_k\right)+\log\left(1-D_k\left(\left(I_1\right)_k,\mathrm{G}\left(I_1\right)_k\right)\right)\right]\tag{6}$$

3.3. VGG 特征损失

为了进一步提升模型的训练稳定性和修复效果，我们引入了一种感知损失函数。不同于传统的像素级损失函数(例如损失)，感知损失函数关注图像在视觉特 l_2 征层面的差异，而非简单地比较像素值。我们选择使用经典的 VGG [8]损失函数，它利用预训练的 VGG 网络提取图像的深层特征，并比较修复图像与真实图像在这些特征上的差异。通过最小化感知损失，模型能够学习生成更符合人类视觉感知的修复结果，从而提升图像的整体质量和自然度。VGG 损失函数如下式所示：

$$L_{vgg}=\sum_{l=1}^{L}\frac{1}{C_lH_lW_l}\left\|f_{vgg}^l\left(I_1\right)-f_{vgg}^l\left(I_{GT}\right)\right\|_1\tag{7}$$

其中 $f_{vgg}^l\left(I_1\right)$ 和 $f_{vgg}^l\left(I_{GT}\right)$ 是通过 VGG 网络提取的特征图，第 l 层特征图的尺寸大小为 $C_lH_lW_l$。

3.4. 高视野特征损失

最后，我们引入了高感受野感知损失(HRFPL)来进一步提升修复效果。HRFPL 类似于 VGG 损失，也是一种感知损失函数，但它能够捕捉更大范围的图像上下文信息。具体来说，HRFPL 首先将修复图像和真实图像映射到更高层的特征空间，然后计算它们在特征层面的欧氏距离。通过最小化 HRFPL，模型能够学习生成更符合全局语义和结构的修复结果。HRFPL 如下式所示，$H_l^{I_1}$ 代表在高感受视野中提取的 I_1 的特征图，N 表示的是特征图中所有特征点的数量。

$$L_{hrfpl}=\sum_{l=0}^{l-1}\left[\frac{\left\|H_l^{I_1}-H_l^{I_{gt}}\right\|_2}{N}\right]\tag{8}$$

4. 实验

4.1. 实验验证

本文在 Cityscapes 街景数据集上做了实验验证，Cityscapes 是一个专注于城市街道场景理解的大规模数据集，包含来自 50 个城市的 5000 张精细标注和 20,000 张粗略标注的高分辨率图像。它提供了像素级别的语义和实例分割标注，涵盖 19 个类别，如道路、车辆和行人，以及部分图像的立体视觉信息。Cityscapes 被广泛应用于自动驾驶、机器人导航和城市规划等领域。在实验中，我们使用 Adam 算法对生成器和判

别器进行优化，并直接采用了原数据集的训练集和测试集划分方式。本文利用所提出的模型对街景图像进行了修复，并在图 4 中展现了修复效果。图中，第一列为待修复的受损街景图像，第二列为对应的真实街景图像，第三列为使用生成式图像修复模型获得的修复结果，第四列为真实街景图像对应的语义分割图，最后一列则为语义图修复模型获得的语义分割图修复结果。

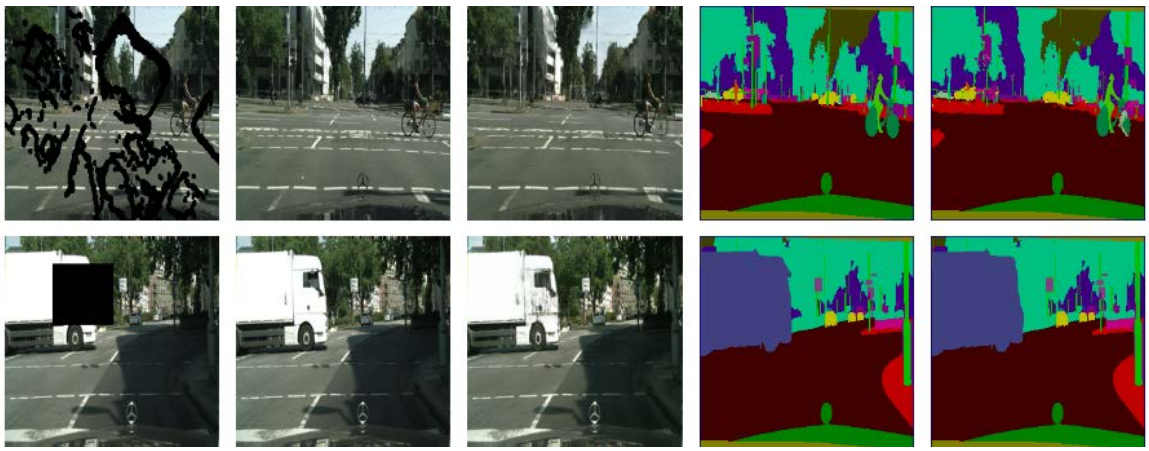

Figure 4. Validation results of the proposed model
图 4. 模型验证结果展示

在本研究中，我们提出的模型在街景数据集上展现出卓越的图像修复能力，尤其在各类语义的修复方面，实现了令人印象深刻的精度和真实感。无论是建筑物、道路、车辆，还是行人、树木等元素，该模型都能精准地还原其缺失信息，有效去除噪声和伪影，并确保修复后的图像在视觉上与真实场景高度一致。

4.2. 对比试验

Figure 5. Comparative experiment on irregular mask image inpainting
图 5. 不规则形状遮掩对比试验展示

Figure 6. Comparative experiment on regular mask image inpainting
图 6. 规则形状遮掩对比试验展示

在本文对比试验设计中，本文分别在规则形状遮掩与不规则形状遮掩上进行了对比，旨在更全面地评估模型在不同遮掩情况下的修复性能，并深入探究模型对复杂遮掩模式的泛化能力。规则形状遮掩，

例如矩形或圆形遮挡，通常用于简化实验设置和初步评估模型的基本修复能力。然而，真实世界中的图像破损或缺失往往呈现出不规则、复杂的形态。为了更贴近实际应用场景，我们引入了不规则形状遮掩，模拟更加多样化的遮挡情况，例如物体遮挡、随机涂鸦或划痕等。通过对比模型在规则形状遮掩和不规则形状遮掩下的修复效果，我们可以更全面地了解模型的性能表现，避免模型在特定遮掩模式下表现良好，但在其他情况下表现不佳的情况。更重要的是，不规则形状遮掩能够更好地评估模型对复杂遮掩模式的泛化能力。一个优秀的图像修复模型应该具备强大的泛化能力，能够有效应对各种类型的遮挡，并将其推广到更广泛的应用场景中。

同时本文选择 SPG [9]与 W-Net [10]这两种方法与街景图像修复模型进行比较。SPG 与本文提出的模型同样是基于生成对抗网络的街景图像修复模型，W-Net 综合利用了结构和纹理特征等信息，并通过 W 形网络结构、TSA 模块、SCE 模块和两阶段网络等方法来实现高质量的图像修复。但是与本文提出的针对街景图像的修复模型相比，修复效果均有一定的差距。

如图 5 和图 6 分别展示了在规则形状遮掩和不规则形状遮掩下，本文方法与其他方法的对比实验结果。在每组对比实验中，第一列为输入街景图像修复模型的受损图像，第二列为对应的真实图像，第三列为本文提出的街景修复模型的修复结果，第四列为 SPG 方法的修复结果。SPG 方法同样引入了语义信息作为修复任务的先验指导，但与之相比，本文方法能够有效修复图像中缺失的各类语义信息、物体结构轮廓等重要信息，提升了修复结果的视觉质量和真实感，使修复后的图像更加自然流畅，更接近真实场景。第五列为 W-Net 方法结果。在图 5 中本文方法的结果 PSNR 值为 27.28，SPG 结果的 PSNR 值为 20.21 以及 W-Net 的 PSNR 值为 21.63。图 6 中本文方法的结果 PSNR 值为 24.31，SPG 结果的 PSNR 值为 18.64 以及 W-Net 的 PSNR 值为 22.48。定量指标说明本文方法的结果显著优于其他方法，在 PSNR 值上实现了很大的提升。

5. 结论与展望

本文提出了一种基于生成对抗网络的街景图像修复模型。该模型有效利用语义信息，对不同类别的语义进行提取和处理，并将其融入图像修复过程。实验结果表明，利用语义信息引导的生成对抗网络能够显著提升街景图像的修复效果，例如，可以精准地还原车辆、道路、建筑物等物体的轮廓、纹理细节和颜色等，使修复后的图像更加自然逼真，有效地消除了图像缺失部分带来的视觉突兀感，提升了图像的整体质量。总结来说，将语义信息融入图像修复模型，特别是生成对抗网络，能够有效提升修复效果，使修复后的图像更加逼真、自然，并提升图像的整体质量。

未来，我们将致力于拓展该模型的应用范围，将其应用于人脸图像以及其他类型场景的图像修复任务中，旨在提升图像修复模型的泛化能力和修复效果，使其能够更好地应对不同场景下的图像修复需求。

基金项目

江苏省研究生科研与实践创新计划项目(SJCX23_0250)。

参考文献

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., *et al*. (2014) Generative Adversarial Nets. *Proceedings of the* 27*th International Conference on Neural Information Processing Systems*, **2**, 2672-2680.

[2] 朱立忠, 佟昕. 基于深度学习的图像双分支修复算法[J]. 通信与信息技术, 2024(5): 14-18+55.

[3] 徐嘉悦, 赵建平, 李冠男, 等. 级联式生成对抗网络的全景图像修复[J]. 重庆理工大学学报(自然科学), 2024, 38(8): 154-163.

[4] Nazeri, K., Ng, E., Joseph, T., Qureshi, F. and Ebrahimi, M. (2019) EdgeConnect: Structure Guided Image Inpainting

Using Edge Prediction. 2019 *IEEE/CVF International Conference on Computer Vision Workshop* (*ICCVW*), Seoul, 27-28 October 2019, 3265-3274. https://doi.org/10.1109/iccvw.2019.00408

[5] Xiang, H., Min, W., Han, Q., Zha, C., Liu, Q. and Zhu, M. (2024) Structure-Aware Multi-View Image Inpainting Using Dual Consistency Attention. *Information Fusion*, **104**, Article ID: 102174. https://doi.org/10.1016/j.inffus.2023.102174

[6] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2017) Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**, 4278-4284. https://doi.org/10.1609/aaai.v31i1.11231

[7] Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J. and Catanzaro, B. (2018) High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8798-8807. https://doi.org/10.1109/cvpr.2018.00917

[8] Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O. (2018) The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 586-595. https://doi.org/10.1109/cvpr.2018.00068

[9] Song, Y., Yang, C., Shen, Y., *et al*. (2018) SPG-Net: Segmentation Prediction and Guidance Network for Image Inpainting.

[10] Zhang, R., Quan, W., Zhang, Y., Wang, J. and Yan, D. (2023) W-Net: Structure and Texture Interaction for Image Inpainting. *IEEE Transactions on Multimedia*, **25**, 7299-7310. https://doi.org/10.1109/tmm.2022.3219728