

基于 CycleGAN 的图像风格迁移任务实现

何雪怡¹

摘要: 本研究利用 CycleGAN 实现图像风格迁移任务, 使用 monet2photo、ukiyoe2photo 和 vangogh2photo 数据集完成三种风格的图像转换。为了进一步优化性能, 本文基于 CycleGAN 修改模型训练参数, 更换生成器网络结构, 并在 monet2photo 数据集上进行重新训练。通过实验对比原始模型与改进模型的结果, 验证改进方法的有效性。

关键词: CycleGAN、图像风格迁移、生成网络优化

中图分类号: TP393.1

文献标识码: A

文章编号: 1001-5965(XXXX)XX-XXXX-XX

1 引言

图像风格迁移是计算机视觉中的重要任务, 旨在将一类图像转换为特定风格。近年来, 生成对抗网络 [1] (GAN) 的提出极大推动了风格迁移任务的发展。特别是 CycleGAN [2] 在无监督学习中表现出色, 通过循环一致性约束实现两类图像域之间的风格迁移, 而无需配对数据集。

图像风格迁移的广泛应用包括艺术作品风格模拟、图像增强、图像翻译等。例如, 艺术家莫奈、梵高以及日本浮世绘风格的图像转换, 可以将艺术风格应用于现实图像, 生成富有艺术性的结果。此外, 风格迁移技术也广泛应用于医学图像处理、自动化设计生成和游戏开发等领域。

本文基于 CycleGAN 模型, 针对 monet2photo、ukiyoe2photo 和 vangogh2photo 数据集, 分别实现了莫奈、浮世绘和梵高风格的图像转换。此外, 为了提升模型效果, 本文在生成器网络结构方面进行了优化, 并在 monet2photo 数据集上进行实验与结果对比。

2 相关工作

Pix2Pix: 成对样本的图像到图像转换方法。
Pix2Pix [3] 是 CycleGAN 的前身, 由 Isola 等人提

出, 解决了图像到图像转换任务中的有监督学习问题。该方法通过成对样本 (paired data) 的训练, 学习从输入域到输出域的直接映射, 并引入条件生成对抗网络 (Conditional GAN) 框架, 确保生成结果的真实性和多样性。Pix2Pix 的主要贡献在于提出了适用于多种图像转换任务的通用方法, 例如从语义分割图生成真实图像、从黑白照片生成彩色照片等。然而, 由于该方法需要成对训练样本, 其应用场景受到限制。CycleGAN 在此基础上通过无监督学习框架克服了这一局限性。

DiscoGAN: 发现领域间的图像关联。 DiscoGAN [4] (Discovery GAN) 由 Kim 等人提出, 与 CycleGAN 类似, 也关注无监督条件下的图像到图像转换问题。该方法的核心思想是通过 GAN 学习两个域之间的双向映射关系, 同时保持生成图像的语义一致性。DiscoGAN 与 CycleGAN 的主要区别在于, 其损失函数并未显式引入循环一致性损失, 而是通过构造对应域内的特定约束来实现语义一致性。虽然 DiscoGAN 在某些简单任务 (如物体形状转换或颜色转换) 中表现良好, 但其生成效果在复杂场景下略逊于 CycleGAN。

DualGAN: 双向无监督图像到图像转换。 DualGAN [5] 是 CycleGAN 的另一种变体, 由 Yi 等人提出。其主要思想是利用双生成器和双判别器结构, 学习从源域到目标域以及从目标域到源域的映射关系, 同时引入对抗损失和重构损失来确保生成图像的质量。DualGAN 和 CycleGAN 都依赖循环一致性原则, 但在模型训练和优化上有所不同。DualGAN 的优化过程更倾向于稳定性, 特别是在处理高分辨率图像时具有一定优势。然而,

CycleGAN 在生成图像的质量和细节保留上表现更为优秀。

UNIT: 基于共享潜在空间的无监督图像翻译。UNIT [6] (Unsupervised Image-to-Image Translation) 由 Liu 等人提出, 是一种基于共享潜在空间 (shared latent space) 的图像到图像翻译方法。与 CycleGAN 不同, UNIT 假设源域和目标域共享一个潜在表征, 通过变分自编码器 (VAE) 和生成对抗网络 (GAN) 的结合, 实现从潜在空间到图像空间的映射。UNIT 在某些任务中生成的图像具有较好的语义一致性, 但其共享潜在空间的假设在某些复杂场景中可能不成立, 限制了模型的泛化能力。相较而言, CycleGAN 在不对潜在空间作强假设的情况下更具适应性。

这些相关工作展示了 CycleGAN 在图像到图像转换领域的重要性, 同时也反映了该领域的研究在生成质量、训练效率和适用性方面的多样化探索。

3 方法

3.1 CycleGAN

CycleGAN, 即循环一致性生成对抗网络 (Cycle-Consistent Generative Adversarial Network), 是一种深度学习模型, 广泛应用于图像到图像转换任务。该模型的核心优势在于无需成对样本 (paired data), 即可在不同领域间实现高质量的风格迁移和图像转换, 这使得它在无监督学习场景中具有重要的研究与应用价值。

CycleGAN 的设计基于生成对抗网络 (GAN) 的基本框架, 同时引入了循环一致性 (cycle consistency) 的概念, 这一机制极大地缓解了无配对数据中映射不唯一的问题。传统的图像转换任务通常需要成对的数据集, 即源领域 (source domain) 和目标领域 (target domain) 中图像具有一一对应的关系。然而, 在实际场景中, 这种成对数据的获取往往极其困难甚至不可能。CycleGAN 的无监督学习框架解决了这一瓶颈问题, 使得仅凭两组无配对的图像数据即可训练出具有可靠映射能力的生成器。

CycleGAN 的网络结构由两个生成器 (generators) 和两个判别器 (discriminators) 构成。生成器的任务是实现领域间的图像映射: 一个生成器将源领域的图像转化为目标领域的图像, 而另一个生成器完成逆向的映射。判别器则分别对目标领域和源领域的生成结果进行真实性的判别, 从

而通过对抗机制提高生成器的生成质量。在此基础上, CycleGAN 提出了循环一致性损失 (cycle consistency loss) 这一关键机制, 即在图像完成从源领域到目标领域的映射后, 生成器能够通过逆向映射将图像还原为源领域, 并尽量保持原始内容的核心特征不变。循环一致性损失的引入有效减少了映射过程中的信息丢失, 确保生成结果在风格转换的同时, 仍能保持语义层面的完整性。

3.2 选择原因

在应用层面, CycleGAN 展现出广泛的跨领域适用性。在图像风格迁移任务中, CycleGAN 可以将自然照片转换为特定的艺术风格, 例如梵高、莫奈或毕加索的画风; 还可以实现自然场景的季节转换, 例如将夏季风景变为冬季风景。此外, 在医学影像处理中, CycleGAN 可用于不同模态图像间的转换, 例如将 CT 图像转化为 MRI 图像, 以增强医疗数据分析的多样性与鲁棒性。在数据增强领域, CycleGAN 可生成多样化的数据样本, 用以缓解数据稀缺问题, 尤其在需要模拟特定条件 (如低光照或特殊天气) 的场景中具有显著优势。更广泛地, 它还被用于视频风格化处理、建筑草图转效果图生成等任务, 为艺术创作和工业设计提供了高效解决方案。

CycleGAN 的成功依赖于其独特的设计理念及无监督学习的高适配性。一方面, 通过对抗损失 (adversarial loss) 和循环一致性损失的结合, 模型能够生成高质量且风格逼真的转换结果; 另一方面, 其无需成对数据的特性显著降低了对数据标注的依赖, 使其在实际应用中具备了更强的普适性与经济性。然而, CycleGAN 也面临一些局限性, 例如生成结果在某些场景下可能会出现伪影 (artifacts) 或细节丢失, 同时, 训练过程需要较高的计算资源。尽管如此, 作为一种无监督图像到图像转换的代表性方法, CycleGAN 为跨领域图像生成问题提供了重要的理论基础与实践启示。

4 实验

4.1 实验设置

本实验选取莫奈、梵高和浮世绘风格图像到照片的转换作为应用场景。实验数据集来自公开的数据集, 实验采用 PyTorch 框架实现, 所有代码均在 Intel(R) Xeon(R) Silver 4215 CPU @ 2.50GHz 16GB 以及双卡 NVIDIA GeForce RTX 3080 10GB 硬件环境下运行。

实验过程中，模型生成器的结构选用 ResNet 包含 9 个残差块，判别器采用基础网络结构，包含 3 层卷积网络。生成器和判别器的特征图数分别设定为 64。为了提升模型的训练稳定性，初始化方法为正态分布初始化，标准差为 0.02)。损失函数选用 lsgan 作为 GAN 目标函数，且模型训练中未使用 dropout 机制。

输入图像通道数为 3 (RGB)，输入图像后，首先将图像调整为 286×286 的尺寸，然后随机裁剪为 256×256。训练过程中，输入和输出图像的方向设置为 AtoB，并启用了图像水平翻转的数据增强。训练样本采用非顺序加载，加载线程数为 4。

CycleGAN 模型的优化器使用 Adam 优化器，设置初始学习率为 0.0002，动量参数为 $\beta_1=0.5$ 。由于硬件资源的限制，共运行 50 个 epoch。学习率采用线性衰减策略，从第 25 个 epoch 开始逐步减小学习率，训练共进行 50 个 epoch，其中前 25 个 epoch 保持固定学习率，后 25 个 epoch 进行线性衰减。每 5 个 epoch 保存一次模型权重，并在训练日志中每隔 100 步打印一次训练信息并显示生成结果。此外，设置了一个大小为 50 的历史缓冲池用于生成样本的存储。

损失函数的权重超参数中，Cycle Consistency Loss 的权重分别为 10.0，Identity Loss 的权重设置为 0.5。

4.2 模型结构更改

本实验对 CycleGAN 模型的生成器结构进行了调整，将原始使用的 9 残差块 (ResNet-9blocks) 生成器更换为包含 6 残差块 (ResNet-6blocks) 和 128 通道 U-Net 结构的生成器，并在 Monet 数据集上重新进行训练，训练轮数为 50。以下对几种生成器结构的特点进行详细分析：

ResNet-9blocks 生成器采用包含 9 个残差块的结构，其特点是通过残差连接 (Residual Connections) 有效缓解了深层网络的梯度消失问题，同时保留输入图像的细节信息和语义特征。它的优势在于生成图像的风格保真度较高，特别适用于复杂纹理和细节较多的任务。然而，9 个残差块的深度使得网络的计算量较大，训练和推理速度相对较慢。

ResNet-6blocks 生成器则减小了残差块的数量，从 9 个减少到 6 个。这种调整降低了模型的复杂度，训练速度和内存占用得到显著提升，更适合计算资源有限的场景。然而，由于网络深度

的减少，生成图像的风格细腻度可能稍逊于 9 残差块的模型。

128 通道 U-Net 生成器在传统 U-Net 结构的基础上增加了特征通道数 (128 通道)，进一步增强了网络的特征表达能力。U-Net 通过编码器和解码器结构以及特征跳跃连接 (Skip Connections) 实现对图像局部和全局信息的融合。在图像到图像翻译任务中，U-Net 能够更好地保留源图像的结构信息，同时实现风格的迁移。相比 ResNet 结构，U-Net 的优势在于生成的图像具有更强的空间一致性，尤其是在需要严格保持图像结构的任务中表现优异。然而，模型对于生成复杂纹理可能不如 ResNet 表现精细。

4.3 实验结果

4.3.1 模型评价指标

在 CycleGAN 模型的训练过程中，为了全面评估生成器与判别器的性能，使用以下评价指标对模型的训练过程进行监控。各项指标的具体含义和作用如下：

- **D_A**: 判别器 A 的损失 (Discriminator A Loss)。用于评估判别器 A 对生成图像 (从域 B 到域 A 的映射) 与真实图像的区分能力。较低的 D_A 值通常表明判别器 A 在真实与生成样本的分类上表现较好。
- **G_A**: 生成器 A 的损失 (Generator A Loss)。用于评估生成器 A (从域 B 到域 A 的映射) 的生成质量。G_A 值包括生成器生成逼真样本的能力以及生成样本骗过判别器 A 的能力。
- **cycle_A**: 循环一致性损失 (Cycle Consistency Loss) A。衡量域 B 的样本经生成器 A 和生成器 B 的双向映射后，是否能够接近原始域 B 的样本。较低的值表明循环一致性较强，生成的样本在域间映射中保留了更多的原始特征。
- **idt_A**: 恒等映射损失 (Identity Mapping Loss) A。用于衡量域 A 中的样本在通过生成器 A (理论上不应发生变化) 时，是否保持了原有的图像特征。这一指标通常用于保证生成器不会对不需要转换的样本进行无意义的修改。
- **D_B**: 判别器 B 的损失 (Discriminator B Loss)。用于评估判别器 B 对生成图像 (从域 A 到域

B 的映射)与真实图像的区分能力。较低的 D_B 值表明判别器 B 在真实与生成样本的分类上表现较好。

- **G_B**: 生成器 B 的损失 (Generator B Loss)。用于评估生成器 B (从域 A 到域 B 的映射) 的生成质量。 G_B 值包括生成器生成逼真样本的能力以及生成样本骗过判别器 B 的能力。
- **cycle_B**: 循环一致性损失 (Cycle Consistency Loss) B。衡量域 A 的样本经生成器 B 和生成器 A 的双向映射后, 是否能够接近原始域 A 的样本。较低的值表明循环一致性较强, 生成的样本在域间映射中保留了更多的原始特征。
- **idt_B**: 恒等映射损失 (Identity Mapping Loss) B。用于衡量域 B 中的样本在通过生成器 B (理论上不应发生变化) 时, 是否保持了原有的图像特征。该指标与 idt_A 类似, 但针对域 B 的样本进行计算。

判别器 (D_A, D_B) 和生成器 (G_A, G_B) 损失衡量生成对抗网络的基本性能; 循环一致性损失 ($cycle_A, cycle_B$) 和恒等映射损失 (idt_A, idt_B) 则用来保证域间映射的可逆性和原始特征的保留, 从而确保生成图像的高质量和一致性。

4.3.2 三风格结果对比

本节展示在三个数据集的训练过程中指标参数的变化, 并进行分析。

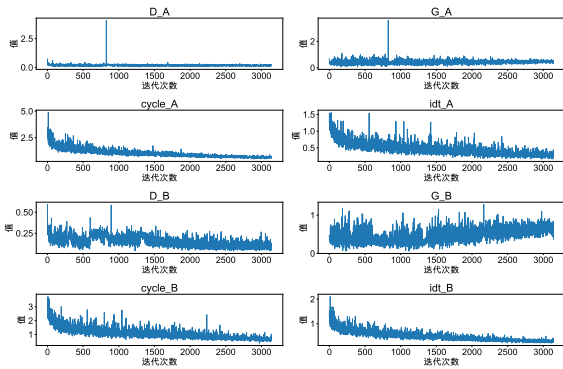


图 1 莫奈风格基础模型训练过程参数曲线

Fig. 1 Parameter curve of the Monet style basic model training process

图 1 展示了 CycleGAN 模型在莫奈风格迁移任务中的关键参数随训练迭代次数的变化曲线, 包括判别器损失 (D_A 和 D_B)、生成器损失 (G_A

和 G_B)、循环一致性损失 ($cycle_A$ 和 $cycle_B$) 以及恒等映射损失 (idt_A 和 idt_B)。从曲线变化可以看出, 判别器损失在训练初期较高, 随后逐渐下降并趋于平稳, 表明判别器逐步学会了区分真实与生成样本; 生成器损失则显示了类似的下降趋势, 说明生成器的生成质量不断提升, 能够较好地骗过判别器。同时, 循环一致性损失的显著下降表明域间映射的可逆性逐步得到保证, 生成器能够保留输入图像的特征。此外, 恒等映射损失的快速下降反映了模型对输入域样本保持恒等映射的能力逐步增强。总体来看, 各项损失的下降和趋于平稳表明模型训练达到了较好的收敛效果, 生成器和判别器在对抗学习中达成了动态平衡。

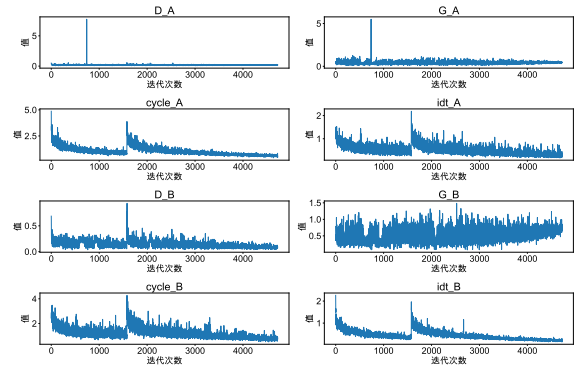


图 2 浮世绘风格基础模型训练过程参数曲线

Fig. 2 Parameter curve of the Ukiyoe style basic model training process

从图 2 中可以看出, 各曲线在训练初期波动较大, 随后逐渐趋于平稳, 表明模型逐步收敛。在约 900 次迭代时, 部分曲线出现明显的损失值暴增现象。这种中途剧烈波动可能由模型内部动态变化引起, 例如生成器与判别器之间的短期不平衡或学习率调整导致的优化剧烈波动。尽管如此, 训练后期损失值趋于平稳, 说明模型逐渐达到了动态平衡, 训练过程总体符合预期。

图 3 中, 同图 2 类似, 多条曲线出现了损失值暴增现象, 随着训练的继续, 模型再次恢复平衡并稳定。

最终结果如图 4:

4.4 改进实验

改进实验中, 修改了生成器结构, 下面展示训练过程中指标的变化。

观察两图发现, D_A 与 G_A 曲线的波动明显大于使用 9ResNet 块的模型, 证明减少参数量、

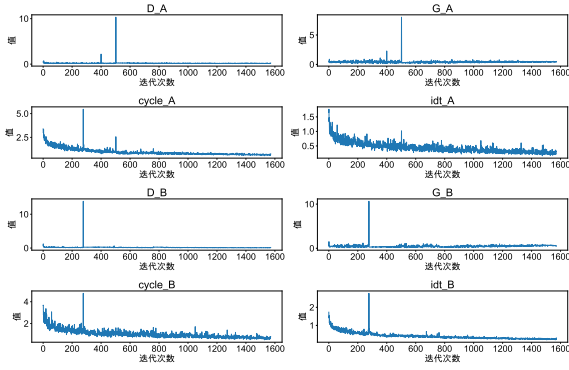


图 3 梵高风格基础模型训练过程参数曲线

Fig. 3 Parameter curve of the Van Gogh style basic model training process

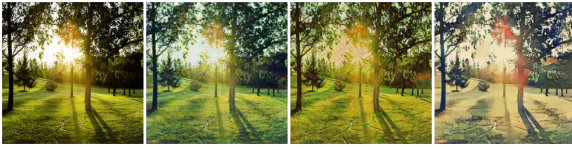


图 4 三风格迁移结果图，从左至右依次是：原图，莫奈，梵高，浮世绘

Fig. 4 Three style transfer results, from left to right: original image, Monet, Van Gogh, Ukiyoe

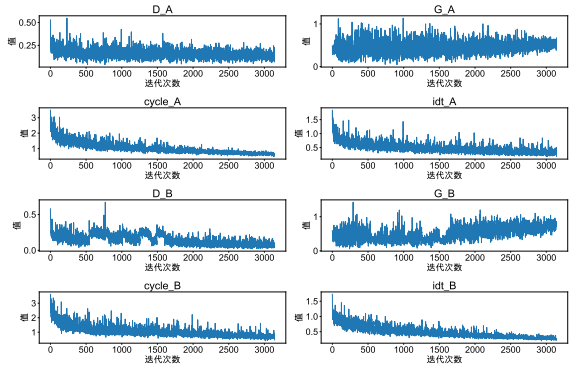


图 5 Resnet_6_block 参数曲线

Fig. 5 Parameter curve of Resnet_6_block

降低网络深度后，模型的训练稳定性出现了下降。而 UNet128 生成器的 cycle_A 与 idt_A 的收敛速度更快，值更低，证明该模型不会篡改原图像的特征，保留原图像信息的能力较强。

同时，本文对三种网络结构的生成图像质量进行了评估。结果如表 1 所示。评估指标包括 FID、PSNR 和 SSIM 三个方面。FID (Fréchet Inception Distance) 用于衡量生成图像与真实图像分布的相似度，数值越小表示生成效果越好；PSNR (Peak Signal-to-Noise Ratio) 反映生成图像的细节还原质量，数值越大越好；SSIM (Structural Similarity

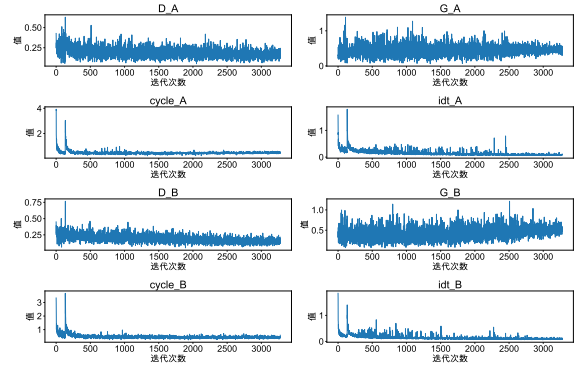


图 6 Unet_128 参数曲线

Fig. 6 Parameter curve of Unet_128

Index Measure) 则衡量生成图像与真实图像的结构相似性，取值范围为 0 到 1，越接近 1 表示相似性越高。

表 1 生成质量评估

Table 1 Generating Quality Assessments			
方法	FID	PSNR	SSIM
Resnet_9	158.9928	28.0167	0.7046
Resnet_6	147.7259	28.1109	0.7201
Unet_128	169.6322	28.2989	0.6851

从实验结果来看，Resnet_6 在 FID 和 SSIM 两项指标上均表现优异，分别达到 147.7259 和 0.7201，表明其生成图像在分布和结构相似性上最接近真实图像。而 Unet_128 的 PSNR 最高 (28.2989)，显示其在细节还原方面具有更好的表现。综上，Resnet_6 在综合生成质量上最优，而 Unet_128 在图像细节还原方面更具优势。但考虑到本文任务是图像风格迁移，故 FID 值最大的 Unet_128 (169.6322) 迁移风格最明显。

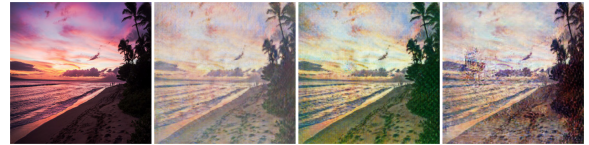


图 7 改进实验结果图，从左至右依次是：原图，Resnet_9, Resnet_6, Unet_128

Fig. 7 Improved experimental results, from left to right: original image, Resnet_9, Resnet_6, Unet_128

5 结论

本文通过在图像风格迁移任务中改进 CycleGAN 模型的训练参数和生成器网络结构，并在

monet2photo 数据集上进行重新训练, 验证了改进方法的有效性。实验结果表明, 与原始模型相比, 优化后的模型在生成质量 (FID 和 SSIM) 及图像细节还原 (PSNR) 方面均有所提升, 尤其是 Resnet_6 在综合性能上表现最优, 而 Unet_128 在细节还原上具有一定优势。这表明, 合理调整模型结构和训练策略能够显著提高图像风格迁移任务的效果。由于硬件资源有限, 本文训练的模型效果并非最好, 如有条件, 将尝试更多数据集与更多风格迁移任务。

参考文献 (References)

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [2] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [4] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning*, pages 1857–1865. PMLR, 2017.
- [5] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017.
- [6] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.

CycleGAN-based image style migration task implementation

Abstract: In this study, CycleGAN is used to implement the image style migration task, and the monet2photo, ukiyoe2photo and vangogh2photo datasets are used to complete the conversion of three styles of images. In order to further optimise the performance, this paper modifies the training parameters of the model based on CycleGAN, replaces the structure of the generator network, and re-trains on the monet2photo dataset. The results of the original model and the improved model are compared through experiments to verify the effectiveness of the improved method.

Key words: CycleGAN; Image style migration; generator network optimisation