

## 1 Introduction

基于神经网络的风格迁移 (也即神经风格迁移) 是深度学习与神经网络在艺术领域做出的尝试. 将具有艺术信息的图像与具有真实场景或人像的内容图像同时输入到网络中, 最终得到具有艺术图像艺术特征以及内容图像内容特征的风格化图像, 是风格迁移最主要的任务. 与大多数计算机视觉任务一样, 风格迁移随着神经网络的兴起而兴起. 神经风格迁移真正走入大众视野可以追溯到由 Gatys 等人 [5] 于 2016 年使用 VGG 中特定层作为损失函数的核心, 以图像像素作为优化对象进行风格迁移工作.

尽管当前工作已经从效率 [7], 质量 [], 速度 [20], 多模态 [3] 等不同方向有了广泛的发展, 但是依旧无法处理具有不透明度信息的图像. 对于一般用户而言, 不透明度通道可能可有可无, 得到一张完整的风格化图像可以满足他们的需求, 但是对于艺术工作者来说, 一副完整作品需要使用多种背景透明的基本元素作为点缀. 这种基本元素往往具有不规则的外形, 而图像往往是以矩形的方式保存的, 所以透明背景在基本元素中也具有重要的作用—防止遮挡其他图像. 至今大部分风格迁移工作 [4, 6, 8–10, 12, 14, 15, 18] 在读入具有不透明度信息的图像时, 大多舍弃不透明度通道, 而是以三通道张量 (R,G,B) 的方式处理他们. 在这中情况下, 原本透明部分由于不透明度通道的丢失, 导致其变成白色或黑色色块. 这些纯色色块会被风格迁移网络误认为是图像风格的一部分, 从而导致风格化图像中存在大量不属于风格信息的色块, 最终生成难以令人满意的迁移结果.

在这份工作中, 我们提出了 **PartAlphaNet**, 一种新颖的端到端风格迁移模型, 在能正常处理无 alpha 通道图像的同时, 专门考虑了具有 alpha 通道图像作为输入的情况. 为了克服传统神经风格迁移方法无法处理 alpha 通道的情况, 本文以部分卷积 [13] 为基础, 因为部分卷积可以很好的忽略图像中不被关心的像素. 但是传统的部分卷积依旧没有处理不透明度信息的方法, 本文为了克服这个困难, 我们使用了一种将 alpha 通道转化为数值在 0 1 之间的浮点掩膜的方法, 成功的将部分卷积扩展到 alpha 通道上. 基于这个方法, 我们还提出了一种浮点掩膜的自动生成手段, 以弥补部分图像不具有的 alpha 通道信息. 同时, 由于当前风格迁移工作数据集往往以 WikiArt [2] 和 MSCOCO [1] 为基础. 这两个数据集中的图像格式基本均为 jpeg, 不包含 alpha 通道信息. 为此, 本工作以 WikiArt 与 MSCOCO 为基础, 从完整图像中拆分外形不规则且具有不透明信息的场景与风格元素, 并以 png 的格式保存, 形成了名为 AlphaImg 的数据集. 实验结果表明, 我们的方法能

够很好的将 alpha 通道的信息纳入考量, 且表现优于当前的 state-of-the-art 的神经风格迁移方法. 总结来说, 我们的贡献可以被罗列如下:

- 我们提出了一个名为“PartAlphaNet”的新颖网络, 可以将图像 alpha 通道信息纳入风格考量, 生成不受背景色块干扰的风格化图像. 这是通过名为 PartAlpha 的部分卷积模块实现的, 该模块以输入图像 alpha 通道信息为基础, 生成浮点掩膜. 在训练时, PartAlpha 以浮点掩膜作为权重, 参与网络训练, 以防止图像格式转换时出现的背景颜色干扰现象.
- 我们提出了 AlphaImg 数据集. 以 WikiArt [2] 与 MSCOCO [1] 为基础, 从完整图像中裁剪并挑选了部分外形不规则的风格元素 (如太阳、花朵、人物等), 并以 png 格式保存. 该数据集弥补了当前风格迁移领域较少具有不透明度信息数据集的缺陷.
- 我们在多种艺术风格上进行了广泛的实验, 结果表明我们的网络可以在处理多种风格的同时, 保留原图像的不透明度信息. 与当前其他方法相比, 我们的方法不会受到图片格式转换造成的风格差异现象, 取得了 state-of-art 的成绩.

## 2 Related Work

**神经风格迁移**使用神经网络作为核心, 以风格图像与内容图像作为输入, 获得具有内容图像内容信息与风格图像风格信息的风格化图像的过程, 被称作神经风格迁移. 神经风格迁移借助如 CNN [17] 等神经网络具有提取全局或局部风格信息的特点, 弥补了传统风格迁移需要人工建模模拟风格纹理的缺陷. To our best Knowledge, 神经风格迁移工作最早由 Gatys 等人于 2016 年 [5] 提出. Gatys 等人 [5] 将噪声图的像素值看作神经网络中的参数, 在以 VGG [16] 为核心的损失函数指导下, 优化噪声图最终得到了风格化图像, 开辟了神经风格迁移的先河.

但由于优化对象是噪声图像中的像素值, 所以 Gatys 等人 [5] 的方法在处理大批量风格迁移任务时需要分别一次每一个风格-内容图像对, 导致迁移速度缓慢, 效率难以令人满意. 同时, 由于每次生成风格化图像需要等待较长的时间, 所以无法做到实时风格迁移. Ulyanov [19] 与 Johnson [9] 几乎同时且独立提出了实时风格迁移成果, 且二人的实现思路类似. 他们的主要

思想是训练一个在网络的参数中编码风格信息的前馈神经网络, 允许风格化图像的生成, 而不需要在噪声图像上进行多次迭代优化. 使用马尔科夫随机场也可以提高风格迁移的速度. Li 和 Wand 等 [11] 改进了他们之前的工作 [33], 通过对抗训练得到马尔科夫前馈网络 (图 5),

## 参考文献

- [1] COCO - Common Objects in Context — cocodataset.org. <https://cocodataset.org/#home>. [Accessed 25-03-2025].
- [2] WikiArt.org - Visual Art Encyclopedia — wikiart.org. <https://www.wikiart.org/en/App/Home>. [Accessed 25-03-2025].
- [3] Namhyuk Ahn, Junsoo Lee, Chunggi Lee, Kunhee Kim, Daesik Kim, Seung-Hun Nam, and Kibeom Hong. Dreamstyler: Paint by style inversion with text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 674–681, 2024.
- [4] Bin Fu, Junjun He, Jianjun Wang, and Yu Qiao. Neural transformation fields for arbitrary-styled font generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22438–22447, 2023.
- [5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [7] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.
- [8] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019.
- [9] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands*,

- October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.
- [10] Zhanghan Ke, Yuhao Liu, Lei Zhu, Nanxuan Zhao, and Rynson WH Lau. Neural preset for color style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14173–14182, 2023.
- [11] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 702–716. Springer, 2016.
- [12] Dongyang Li, Hao Luo, Pichao Wang, Zhibin Wang, Shang Liu, and Fan Wang. Frequency domain disentanglement for arbitrary neural style transfer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 1287–1295, 2023.
- [13] Guilin Liu, Aysegul Dundar, Kevin J Shih, Ting-Chun Wang, Fitsum A Reda, Karan Sapra, Zhiding Yu, Xiaodong Yang, Andrew Tao, and Bryan Catanzaro. Partial convolution for padding, inpainting, and image synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6096–6110, 2022.
- [14] Eric Risser, Pierre Wilmot, and Connelly Barnes. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv preprint arXiv:1701.08893*, 2017.
- [15] Artsiom Sanakoyeu, Dmytro Kotovenko, Sabine Lang, and Bjorn Ommer. A style-aware content loss for real-time hd style transfer. In *proceedings of the European conference on computer vision (ECCV)*, pages 698–714, 2018.
- [16] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

- [17] Srikanth Tammina. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10):143–150, 2019.
- [18] Licheng Tang, Yiyang Cai, Jiaming Liu, Zhibin Hong, Mingming Gong, Minhu Fan, Junyu Han, Jingtuo Liu, Errui Ding, and Jingdong Wang. Few-shot font generation by learning fine-grained local styles. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7895–7904, 2022.
- [19] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. *arXiv preprint arXiv:1603.03417*, 2016.
- [20] Zhanjie Zhang, Quanwei Zhang, Wei Xing, Guangyuan Li, Lei Zhao, Jiakai Sun, Zehua Lan, Junsheng Luan, Yiling Huang, and Huaizhong Lin. Artbank: Artistic style transfer with pre-trained diffusion model and implicit style prompt bank. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7396–7404, 2024.