

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 [6] 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的风格迁移工作.

尽管当前工作已经从效率 [9], 质量 [14], 速度 [37], 多模态 [1] 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 [5, 7, 10, 11, 14, 17, 25, 26, 29] 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 [?] 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 [23] 和 MSCOCO [18] 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 [23] 与 MSCOCO [18] 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 [28] 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 [6] 提出. Gatys 等人 [6] 将噪声图的像素值看作神经网络中的参数, 在以 VGG [27] 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 [6] 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 [30] 与 Johnson [11] 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: 参数化特定作品的风格信息, 并用反向传播算法固化在前馈神经网络中, 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是早期为了实现在实时风格迁移的尝试. Li 等人 [16] 用马尔科夫随机场进一步优化了他们之前的一项工作 [15], 他们通过对抗训练得到了一个马尔科夫前馈网络, 以解决效率问题. 在这之后, 还有以变分自编码器 [36]、生成式对抗网络 (Generative Adversarial Network, GAN) [12, 22, 38] 为基础的诸多方法为了提升风格迁移速度而努力. 尽管他们的方法大降低了风格迁移的耗时, 但只能迁移预先训练好的一种风格, 灵活性较差.

随着风格迁移的进一步发展, 研究者们不再满足于迁移风格数量与效率之间的权衡, 开始追求高效且可以迁移多种风格的方法. 据我们所知, Dumoulin 等人 [4] 第一个在数量-效率权衡之间做出了突破, 他们在保证风格迁移效率的同时, 将可迁移的风格数量从一个扩展至多个. 具体来说, 他们发现对归一化参数进行缩放或变换足以适应特定的风格, 因此提出了 Conditional Instance Normalization(CIN) 方法. 通过该方法调整网络参数, 可以完成多种类似但不同的风格迁移任务. Chen 等人 [2] 不同于 Dumoulin [4] 通过公式调整网络参数的方式, 提出了名为 StyleBank 的网络参数替换技术. 他们认为网络中处理内容信息的部分可以是相同的, 仅需要替换处理风格图像的参数部分即可. 所以他们将训练好的参数保存在一个"Bank" 中, 当模型被要求迁移某种风格时, 网络将从"Bank" 中读取对应参数层并加载入网络中. 通过这种参数替换方式, 在保证效率的前提下增加了可迁移风格种类. 这两种方法作为 Per-Model Multi-style 方法的典型方法, 分别具有不同的缺陷: Dumoulin 等人 [4] 方法以迁移质量为代价, 而 Chen 等人 [2] 模型的数量会随着可迁移种类的增加而增加, 造成存储与使用困难.

发展到现在, 大部分风格迁移工作均可满足实时且任意 (Arbitrary and Real-time) 的要求. 按实现方式的不同, 任意且实时风格迁移可以分成五类: 1. 基于全局参数调整; 2. 使用 GANs; 3. 使用注意力机制; 4. 使用预训练大模型; 5. 使用自搭建的网络. 第一个实现任意和同时风格迁移的工作由 Huang 等人在 2017 年提出 [9], 属于全局参数调整. 受 CIN 方法 [4] 的启发, Huang 等人以自适应实例归一化层 (Adaptive Instance Normalization layer, AdaIN) 作为风格迁移网络的核心, 该层可以将内容特征的方差和均值与风格特征的均值和方差进行匹配, 从而实现任意的风格迁移. Xu 等人的工作 [31] 是利用 GAN 实现任意且实时风格迁移的一个典范. 他们提出

的 DRB-GAN 通过动态残差块 (Dynamic ResBlocks) 统一整合风格编码与迁移网络, 利用风格类感知注意力机制生成动态风格代码, 从而完成任意且实时的风格迁移. Liu 等人于近期提出了一份使用注意力机制辅助风格迁移的有趣工作 [20]. 该工作允许用户从风格图像中交互式地选择特定区域, 并将其对应的风格应用于内容图像中的相同区域, 从而实现个性化的风格迁移效果. 随着如 CLIP, Diffusion 等模型的发展, 使用这些模型作为辅助的风格迁移工作也逐渐增多. Liu 等人 [21] 从用户体验的角度出发, 认为使用文字描述代替风格图像是提升用户体验的重要因素. 因此他们提出了基于 CLIP 的 TxST 模型用于提取文字描述中的风格信息, 以此作为风格图像特征的代替品, 这使得用户在没有风格图像时, 也能进行风格迁移. Zhang 等人 [34] 基于 Diffusion 模型提出了一种新的风格迁移方法, 其核心思想是将艺术风格视为可学习的文本标签, 并以此引导扩散模型进行图像生成. 为实现这一目标, 他们提出了名为 InST(Inversion-Based Style Transfer) 的反演式风格迁移方法, 能够高效且精确地学习图像相关信息, 从而实现风格迁移. 研究者根据自己的想法设计有效的自搭建风格迁移网络. Li 等人 [17] 不同于以往专注于空间域的方法, 提出从频率域角度解耦图像的内容与风格特征. 基于图像的”频率可分特性”, 他们设计了 FreMixer 模块, 通过傅里叶变换将特征图映射至频率域, 完成风格与内容的分离与重构. 该方法不仅有效缓解了其他方法中的特征纠缠问题, 也为风格迁移提供了新的实现思路.

当前风格迁移领域正向着提升迁移质量、提高迁移速度、提升人机交互体验、由单模态转向多模态的方向发展, 但依旧没有考虑到 alpha 通道. 我们的模型, PartAlphaNet, 考虑到现实生活中被设计师广泛使用的带有不透明度通道的图像 (以 png 格式为主要研究对象), 不再简单舍弃 alpha 通道信息, 而是将 alpha 通道纳入风格迁移考量, 很好的弥补了这一方面的研究空缺.

3 Method

3.1 Background: Alpha 通道在风格迁移任务中的缺失

RGB vs. RGBA: 不透明度信息的重要性 在计算机视觉领域, 图像往往以 R、G、B 三个通道的格式存取与处理, 因为这已足以适应绝大多数任务场景. 然而, 当涉及到对图像中不同区域进行精细的显示与隐藏控制, 或者需要调节多张图像间的遮挡与融合程度时, 传统的 RGB 图像便显得力不从

心. RGBA 图像在原有的 RGB 图像基础上增加了一个 A 通道, 即 alpha 通道, 该通道用以表示图像每个像素的不透明度. alpha 通道的概念最早可追溯至 Porter 和 Duff 在 1984 年提出的数字图像合成理论, 该理论系统阐述了如何利用 alpha 通道实现图层间的混合与透明度控制 [24]. 在实际应用中, alpha 值通常用来指示像素的不透明程度, 其取值范围可以设定为 $[0, 1]$ 或 $[0, 255]$, 通过直接指定或计算生成, 以反映图像中各部分的显示或隐藏程度.

在实际应用中, alpha 通道对于实现图像中的局部区域控制具有显著优势. 以图形设计与合成为例, alpha 通道允许设计师指定图像的前景与背景, 并精确调控多图像之间的遮挡关系. 相比之下, 传统的 RGB 图像由于仅包含颜色信息, 在需要实现精细区域控制时往往显得不足. 设想有两幅图像: 一幅前景图像为人物肖像, 以 PNG 格式 (一种常见的 RGBA 格式) 存储, 其中人物区域完全不透明 (alpha 值接近 1 或 255), 而背景部分则完全透明 (alpha 值接近 0); 另一幅图像为背景图像, 例如一幅风景图. 在传统的 RGB 图像处理中, 仅保留颜色信息的前景图像无法明确区分出真正的前景和需要由背景填充的区域, 因此直接将 RGB 前景叠加到背景上往往会产生不自然的矩形边框和锐利的边缘. 通过提取 PNG 图像中的 alpha 通道, 并将其作为蒙版来使用, 可以仅保留前景图像中不透明区域, 将透明区域交由背景图显示, 从而达到前景与背景自然过渡的效果. 这种方法不仅保留了前景对象的完整形态, 还使背景能够无缝融入, 为图像合成提供了更高的精确性与美观度.

当前风格迁移工作 [5, 7, 10, 11, 14, 17, 25, 26, 29] 大多忽视了图形设计与合成的应用场景. 在网络输入端, 这些方法通常仅接受 3 通道的 RGB 图像作为输入, 而当输入为 RGBA 图像时, 系统只处理 RGB 部分, 忽略了其中的 alpha 通道. 在网络的处理中, 由于随意抛弃 alpha 信息, 导致生成过程中处理的图像与原始 RGBA 图像之间产生较大差距. 网络可能将这些差距视为图像风格的一部分, 进而在输出图像中引入大量不属于原输入图像的风格信息, 不仅使风格迁移的质量大幅下降, 也无法保留原图的不透明度信息. 这对于实际应用场景——如需要精细控制前景与背景, 以实现无缝图层合成的图形设计工作来说, 是致命的缺陷.

alpha 通道与 vanilla convolution 的不适配性 RGBA 图像中的 alpha 通道与 RGB 三个通道在信息表达、语义含义及处理需求上存在本质差异. RGB 通道主要承载图像的颜色与亮度信息, 直接参与图像的视觉重建, 而

alpha 通道则用于编码像素的不透明度, 决定图像在合成、遮挡与可视化中的显隐关系, 更多体现为一种控制性的“元信息”. 与 RGB 通道的结构性视觉特征不同, alpha 通道的数值变化通常呈现出遮挡块状、边界过渡或局部渐变等规律, 具有独立的数据分布与语义特征.

然而, 当前工作常使用 vanilla convolution 统一处理输入通道. 作为一个通用模块, vanilla convolution 无法区分 RGB 通道与 alpha 通道之间的本质差异. 因此在训练与推理过程中, vanilla convolution 会错误的将 alpha 通道当作颜色通道处理, 导致语义污染, 最终导致风格迁移质量下降. 因此, 在处理 RGBA 图像时, 有必要引入对 alpha 通道敏感的建模机制, 以避免信息在特征提取过程中的语义丢失或功能混淆.

3.2 Soft Partial Convolution for Countinuous Alpha Transparency

引入 alpha 通道敏感的建模机制有两种实现方式. 其一, 为 alpha 通道与 RGB 通道分别设计 Encoder, 并在二者之间实现信息交流; 其二, alpha 信息看作 RGB 的补充信息, 用融合策略将 alpha 通道融入主编码网络中, 在统一的特征空间中捕捉颜色与透明度的细粒度特征. 考虑到 alpha 通道用于控制像素显隐程度, 在风格迁移任务中可以表示某区域风格化程度, 可以作为主网络的补充信息; 同时考虑到计算成本, 故本文以第二种实现方式作为主要研究对象.

为了将代表风格化程度的 alpha 信息融入主网络, 我们参考了 Image Inpainting 任务的 Partial Convolution [19]. Partial Convolution 的核心思想是利用一个二值掩膜来区分已知和缺失区域, 从而仅对已知区域进行卷积运算. 然而, 对与具有连续 alpha 通道信息的 RGBA 图像而言, alpha 通道的像素值通常在 $[0, 255]$ 之间变化, 用于反应像素的不透明度与风格化程度. 在这种情况下, 简单地将 alpha 通道二值化容易丢失细腻的透明度信息, 从而影响网络对图像透明边缘与分层合成的精细建模. 因此, 为了让部分卷积更好适配风格迁移任务, 我们有必要对部分卷积进行修改, 使其能够直接使用连续值作为掩膜, 从而更好地保留 alpha 通道的语义信息.

我们提出了 Soft Partial Convolution for image style transfer network. 具体来说, 我们将图像中 alpha 通道归一化为 $[0, 1]$ 之间的 soft mask, 并用该 soft mask 作为风格化程度信息, 辅助风格迁移. 假设输入特征图为 $X \in \mathbb{R}^{C \times H \times W}$, 其对应的 soft mask 由 alpha 通道归一化得到, 记做 $M_s \in$

$[0, 1]$; 给定卷积核 $W \in \mathbb{R}^{C \times K \times K}$ 以及偏置 b , 则在 (i, j) 处的 Soft Partial Convolution 输出 $Y_{i,j}$ 定义为:

$$Y_{i,j} = \begin{cases} \frac{1}{\sum_{k,l} M_s^{(i+k,j+l)}} \sum_{k,l} W \cdot X^{(i+k,j+l)} \cdot M_s^{(i+k,j+l)} + b, & \text{if } \sum_{k,l} M_s^{(i+k,j+l)} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

同时, 我们采样局部平均值来更新掩膜, 得到新的连续掩膜 M_{out}

$$M_{\text{out}}^{(i,j)} = \frac{1}{K^2} \sum_{k,l} M_s^{(i+k,j+l)} \quad (2)$$

值得一提的是, Gated Convolution [32] 通常被认为是对 Partial Convolution 的改进, 其核心创新在于将掩膜的更新策略从基于规则的更新机制替换为可学习的门控机制. 该机制能够在训练中自动学习图像中哪些区域应当被保留, 从而在某些任务中 (如图像修复) 取得了更灵活的表现. 然而, 在本研究聚焦的风格迁移任务中, 我们发现 Partial Convolution 具备更强的结构适配性. 首先, Gated Convolution 更适合目标区域不确定的情况, 而在风格迁移中, 掩膜 (即 alpha 通道) 来源于输入图像本身, 具有明确且一致的语义信息, 无须使用额外的网络进行门控预测. 其次, 风格图像与内容图像的 alpha 通道提供了明确的不透明度信息, 我们更喜欢模型严格遵循这一先验知识建模, 而非通过门控机制学习透明度信息, 以避免不必要的偏差. 虽然 Gated Convolution 作为 Partial Convolution 的改进在某些任务中取得了优越表现, 但在本研究面向的 alpha 感知风格迁移任务中, Partial Convolution 由于其简洁与精确的特性, 表现出了更高的适用性.

3.3 Style Transfer Network Architecture

我们使用所提出的 Soft Partial Convolution 与 AesFA 网络 [14] 定制了一个能够处理 RGBA 图像的风格迁移网络. 具体来说, 我们采用并修改了 [14] 中的主干网络结构, 在保持原始网络优势的基础上, 额外引入了预处理模块与后处理模块, 以更好地适应带有透明度信息的图像建模需求.

在输入 RGBA 图像时, 原始网络的 Encoder 会对所有通道进行统一的特征提取, 未对 alpha 通道与 RGB 通道加以区分. 为解决这一缺陷, 我们在 Encoder 之前增加了预处理模块, 使得 RGB 通道与 alpha 通道被明确分离, 并分别建模. 其中, RGB 图像经过 Soft Partial Convolution 处理, 而 alpha 通道被视为 soft mask, 在整个建模过程中作为引导信息持续参与. 若输入图

像本身不含 alpha 通道（例如 JPG 格式图像），该模块会自动补充一个值均为 255 的 alpha 通道，以增强网络在多图像格式下的鲁棒性与适应性。

在 Decoder 输出图像后，我们进一步引入了后处理模块。该模块以原始输入图像的 alpha 通道为参考，对生成图像的有效区域进行裁剪，使网络输出不仅保持风格一致性，还能严格遵守原图像的透明度约束。这一机制对于图形合成与界面设计等应用场景具有重要意义，确保风格迁移结果能无缝集成于复杂图形层结构中。

此外，我们将原始网络 [14] 中的所有 vanilla convolution 替换为 Soft Partial Convolution，以更好地融合透明度信息带来的结构引导。在此基础上，我们适当加深了网络结构的整体深度，并将特征提取阶段原本使用的双尺度特征图处理机制拓展为三尺度设计，从而增强网络的风格表达能力与多层次语义建模能力。值得一提的是，我们的网络仍保持全卷积结构，能够灵活适配任意分辨率的输入图像，进一步提升其实用性与泛化性。整体网络结果如图所示 (picture needed)。

4 Experiment

此部分中，我们首先介绍提出的 AlphaImg 数据集。随后，我们验证了所提出的 Soft Partial Convolution 在进行具有不透明度信息的风格迁移时的有效性，并说明其相对于其他先进方法的优越性。

4.1 Dataset

为了验证所提出方法在透明图像风格迁移中的有效性，我们构建了一个名为 AlphaImg 的合成数据集。该数据集基于 MSCOCO [18] 与 WikiArt [23] 数据集，通过 Segment Anything Model(SAM) [13] 将整副图像分割并裁剪为多个形状不规则的图像，并为分割后图像添加 alpha 通道信息，以 png 格式存储。训练集与测试集均由内容图像集与风格图像集组成，内容图像集以处理后的 MSCOCO 数据为主，风格图像集以处理后的 WikiArt 数据为主。训练集包含 50,000 张来自 MSCOCO 且背景透明的内容图像，以及 10,000 张来自 WikiArt 的风格图像。测试集中包含 10,000 张来自 MSCOCO 且背景透明的图像，以及 5,000 张来自 WikiArt 的风格图像。其中内容图像不透明度仅为 0 或 255，而风格图像的不透明度则会在 0 255 之间变化，以模拟真实情况。

具体来说, 对一张来自 MSCOCO 的内容图像 C^i , 我们首先使用 SAM 模型为其创建包含 n 个掩膜的掩膜组 $M_c^i = \{m_c^{i,1}, m_c^{i,2}, \dots, m_c^{i,n}\}$. 随后将掩膜组 M_c^i 中的每一张掩膜 $m_c^{i,j}$ 当作内容图像 C^i 的 alpha 通道, 以获取 n 张具有透明背景的内容图像. 对于一张来自 WikiArt 的风格图像 S^i , 我们同样使用 SAM 为其生成风格掩膜组. 所有掩膜根据其白色区域的像素数量进行升序排序, 得到排序后的风格掩膜集合:

$$M_s^i = \{m_s^{i,1}, m_s^{i,2}, \dots, m_s^{i,n}\}, \sum m_s^{i,1} \leq \sum m_s^{i,2} \leq \dots \leq \sum m_s^{i,n} \quad (3)$$

为模拟风格图像中可能存在的透明度分布差异, 我们将每个掩膜 $m_s^{i,j}$ 转换为带有不同灰度级别的 soft mask $m_{s,(x,y)}^{'i,j}$. 具体地, 对于每个像素位置 (x, y) , 其值按如下方式赋值:

$$m_{s,(x,y)}^{'i,j} = \begin{cases} \frac{j+1}{n} \cdot 255, & \text{if } m_{s,(x,y)}^{i,j} = 255 \\ 0, & \text{if } m_{s,(x,y)}^{i,j} = 0 \end{cases} \quad (4)$$

该设计使得编号越大的掩膜 (即区域面积越大的部分) 对应的透明度值越高, 从而体现出一种从局部细节到整体结构的风格强度递进. 为进一步构建一张连续的 alpha 蒙版图, 我们将所有 soft mask 按照如下规则进行叠加合成:

$$\tilde{A}^i(x, y) = \begin{cases} m_{s,(x,y)}^{'i,j}, & \text{if } m_{s,(x,y)}^{i,j} = 255 \text{ 且 } (x, y) \text{ 尚未被赋值} \\ \tilde{A}^i(x, y), & \text{otherwise} \end{cases} \quad (5)$$

其中 \tilde{A}^i 表示图像 S^i 的最终透明度图. 我们初始化 $\tilde{A}^i(x, y) = 255$, 然后依序扫描每一个 soft mask. 一旦某个位置 (x, y) 被某个掩膜 $m_s^{i,j}$ 覆盖且未被之前的掩膜赋值, 则用 $m_{s,(x,y)}^{'i,j}$ 中的值更新之. 该策略可确保不同区域以“后掩膜优先”的方式分配透明度, 从而在一定程度上模拟层叠结构.

4.2 Style Transfer Results

在本节中, we compare the results between the proposed Soft Partial Convolution and previous SOTAs, including AdaIN [9], CycleGAN [38], CAST [35], StyTr2 [3], QuantArt [8], S2WAT [33] 以及 baseline 工作 AesFA [14].

4.3 Qualitative Comparasion

尽管上述方法均基于 RGB 图像训练, 没有针对 RGBA 图像进行专门训练, 但我们认为使用 RGBA 图像为他们重新训练是不必要的. 上述方法在训练或推理时, 会将图像转换为 RGB 的格式, 即使原始图像具有 Alpha 通道信息. 在不修改部分现有成果官方代码的前提下, 它们甚至无法以 RGBA 格式图像作为输入 (如 S2WAT 等). 因此这些方法本身在模型设计和训练阶段并不具备处理不透明度信息的能力. 基于这一事实, 我们采样它们的官方预训练模型直接在 AlphaImg 数据集上进行推理测试, 从而分析其在面对透明信息时的泛化能力和表现差异. 这一设置虽然不是相同训练条件下的严格对比 (fair comparison), 但能真实反映现有方法在处理不透明度图像中的表现瓶颈与适用便捷, 也从实验层面佐证了本文引入不透明度建模机制的必要性与有效性. The qualitative comparisons presented in Figure (picture needed) provide a visual assessment of the outcomes achieved by different style transfer methods.

与上述方法相比, 我们提出的 Soft Partial Convolution 不仅可以将 alpha 通道纳入风格迁移网络, 也能保留原始输入图像的不透明度关系.

4.4 User Study

4.5 Ablation Study

参考文献

- [1] 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.
- [2] Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. Stylebank: An explicit representation for neural image style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1897–1906, 2017.
- [3] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11326–11336, 2022.
- [4] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Siyu Huang, Jie An, Donglai Wei, Jiebo Luo, and Hanspeter Pfister. Quantart: Quantizing image style transfer towards high visual fidelity.

- In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5947–5956, 2023.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [13] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
- [14] Joonwoo Kwon, Sooyoung Kim, Yuewei Lin, Shinjae Yoo, and Jiook Cha. Aesfa: an aesthetic feature-aware arbitrary neural style transfer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 13310–13319, 2024.
- [15] Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2479–2486, 2016.

- [16] 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.
- [17] 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.
- [18] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6–12, 2014, proceedings, part v 13*, pages 740–755. Springer, 2014.
- [19] 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 *Proceedings of the European conference on computer vision (ECCV)*, pages 85–100, 2018.
- [20] S Liu, J Ye, and X Wang. Any-to-any style transfer: Making picasso and da vinci collaborate. arxiv 2023. *arXiv preprint arXiv:2304.09728*.
- [21] Zhi-Song Liu, Li-Wen Wang, Wan-Chi Siu, and Vicky Kalogeiton. Name your style: text-guided artistic style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3530–3534, 2023.
- [22] Yifang Men, Yuan Yao, Miaomiao Cui, Zhouhui Lian, Xuansong Xie, and Xian-Sheng Hua. Unpaired cartoon image synthesis via gated cycle mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3501–3510, June 2022.
- [23] Fred Phillips and Brandy Mackintosh. Wiki art gallery, inc.: A case for critical thinking. *Issues in Accounting Education*, 26(3):593–608, 2011.

- [24] Thomas Porter and Tom Duff. Compositing digital images. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 253–259, 1984.
- [25] 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.
- [26] 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.
- [27] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] Wenju Xu, Chengjiang Long, Ruisheng Wang, and Guanghui Wang. Drb-gan: A dynamic resblock generative adversarial network for artistic style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6383–6392, 2021.
- [32] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Pro-*

- ceedings of the IEEE/CVF international conference on computer vision*, pages 4471–4480, 2019.
- [33] Chiyu Zhang, Xiaogang Xu, Lei Wang, Zaiyan Dai, and Jun Yang. S2wat: Image style transfer via hierarchical vision transformer using strips window attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 7024–7032, 2024.
- [34] Yuxin Zhang, Nisha Huang, Fan Tang, Haibin Huang, Chongyang Ma, Weiming Dong, and Changsheng Xu. Inversion-based style transfer with diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10146–10156, 2023.
- [35] Yuxin Zhang, Fan Tang, Weiming Dong, Haibin Huang, Chongyang Ma, Tong-Yee Lee, and Changsheng Xu. Domain enhanced arbitrary image style transfer via contrastive learning. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–8, 2022.
- [36] Zhanjie Zhang, Jiakai Sun, Jiafu Chen, Lei Zhao, Boyan Ji, Zehua Lan, Guangyuan Li, Wei Xing, and Duanqing Xu. Caster: Cartoon style transfer via dynamic cartoon style casting. *Neurocomputing*, 556:126654, 2023.
- [37] 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.
- [38] 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.