



華東師範大學

EAST CHINA NORMAL UNIVERSITY

## 《智能计算系统》课程期中项目

### 图像风格迁移

姓 名：李鹏达

学 号：10225101460

2025 年 5 月

## 1 实验背景

图像风格迁移是一项旨在将一幅图像的内容结构与另一幅图像的艺术风格相结合的计算机视觉任务。传统方法主要依赖于图像处理技术与纹理合成算法，例如基于非参数采样的纹理合成方法 [2]，但这类方法在内容与风格的有效分离方面存在较大局限，生成图像的表达能力有限。

随着深度学习技术的迅速发展，Gatys 等人于 2015 年首次提出了基于卷积神经网络的图像风格迁移方法 [3]，并取得了显著突破。这一方法首次实现了基于神经网络的、具有高质量视觉效果的图像风格迁移，并开启了后续多种风格迁移技术的发展，为本实验提供了理论基础和技术支撑。

## 2 实验内容

### 2.1 所选神经网络模型简介

Gatys 等人风格迁移方法 [3] 通过对输出图像进行迭代优化，在风格迁移质量上取得了显著成果，但其每处理一张图像需数百次迭代，计算效率极低，难以满足实时应用需求。为此，Johnson 等人于 2016 年提出了 *Fast Style Transfer* 模型 [4]，采用前馈卷积神经网络进行图像变换，并通过感知损失（Perceptual Loss）进行训练，实现了高质量且近实时的风格迁移。在实验中，我们将基于该模型进行实现。

#### 2.1.1 网络结构

该方法的核心是训练一个图像变换网络  $f_W(x)$ ，将输入的内容图像  $x$  映射为风格化输出图像  $\hat{y} = f_W(x)$ 。

图 1 展示了网络的基本结构，输入图像  $x$  首先通过图像变换网络  $f_W$  生成合成图像  $\hat{y}$ 。为了确保  $\hat{y}$  在内容上接近内容目标图像  $y_c$ ，并在风格上与风格目标图像  $y_s$  匹配，利用预训练的 VGG-16 网络  $\phi$  作为损失网络。内容损失通过比较  $\hat{y}$  和  $y_c$  在  $\phi$  的 relu3\_3 层的特征表示计算得到；而风格损失则通过比较  $\hat{y}$  和  $y_s$  在  $\phi$  的多个不同层的风格特征计算得到。最终，通过最小化内容损失和风格损失的加权和来优化  $f_W$  的参数  $W$ ，实现风格迁移。

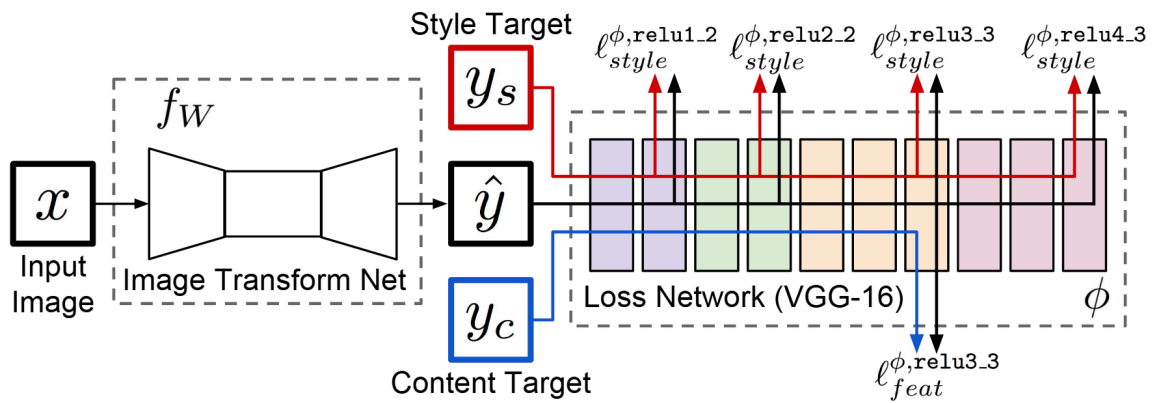


图 1: 网络的基本结构，摘自 [4]。

### 2.1.2 感知损失与 VGG 网络

为度量图像的内容与风格差异，该方法引入一个在 ImageNet[1] 上预训练的 VGG-16 网络 [5]  $\phi$  作为固定的损失网络，从中提取图像的中高层语义特征。损失函数由三部分组成：

#### 1) 内容损失 (Content Loss)

内容损失衡量输出图像  $\hat{y}$  与内容图像  $y_c$  在 VGG 网络中某一中间层  $j$  的特征表示之间的差异：

$$\mathcal{L}_{\text{content}}(\hat{y}, y_c) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(y_c)\|_2^2 \quad (1)$$

其中， $\phi_j(x)$  表示图像  $x$  在第  $j$  层的特征图， $C_j, H_j, W_j$  分别为通道数、高度和宽度。

#### 2) 风格损失 (Style Loss)

风格损失基于多个卷积层提取的特征图的 Gram 矩阵，可以衡量输出图像  $\hat{y}$  与风格图像  $y_s$  在纹理上的相似性：

$$G_j(x) = \frac{1}{C_j H_j W_j} \phi_j(x) \phi_j(x)^T \quad (2)$$

$$\mathcal{L}_{\text{style}}(\hat{y}, y_s) = \sum_{j \in \mathcal{S}} \|G_j(\hat{y}) - G_j(y_s)\|_F^2 \quad (3)$$

其中  $\mathcal{S}$  为选定的风格层集合（如 `relu1_2`, `relu2_2`, `relu3_3`, `relu4_3`）。

#### 3) 总变差正则项 (Total Variation Regularization)

总变差正则项用于鼓励输出图像在空间上具有平滑性，它可以抑制噪声和伪影。

$$\mathcal{L}_{\text{TV}}(\hat{y}) = \sum_{i,j} ((\hat{y}_{i,j+1} - \hat{y}_{i,j})^2 + (\hat{y}_{i+1,j} - \hat{y}_{i,j})^2) \quad (4)$$

### 2.1.3 训练目标

综合上述损失项，图像变换网络的训练目标函数为：

$$\min_W \mathbb{E}_{x \sim \mathcal{D}} [\lambda_c \mathcal{L}_{\text{content}}(f_W(x), y_c) + \lambda_s \mathcal{L}_{\text{style}}(f_W(x), y_s) + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}}(f_W(x))] \quad (5)$$

其中  $\lambda_c, \lambda_s, \lambda_{\text{TV}}$  为各损失项的权重系数， $\mathcal{D}$  表示训练图像集。

## 2.2 所选神经网络模型亮点

- 1) 该模型使用基于深度卷积网络的特征空间差异作为损失函数，而不是简单的像素级 MSE。这样能更好捕捉图像的高层语义和纹理信息，使得风格迁移结果更自然、更艺术化；
- 2) 该模型设计了一个轻量级的前馈卷积网络，训练时用感知损失约束风格和内容，推理时只需一次前向传播即可生成风格化图像，速度远快于传统基于优化的风格迁移方法，相比 [3]，速度提升了数百倍甚至上千倍；
- 3) 该模型训练时间短，通常只需数小时即可完成训练，便于完成实验。

## 3 实验过程

### 3.1 实验环境

本次实验在以下环境下进行：

- 操作系统：Windows 11 家庭中文版 26100.4061
- 显卡：NVIDIA RTX 3070 Laptop
- 深度学习框架：PyTorch 2.7.0+cu126

### 3.2 实验步骤

1. 下载数据集：本实验使用了 ImageNet [1] 数据集的一个子集，共包含 50000 张图像；
2. 下载预训练模型：使用 VGG-16 网络作为损失网络，下载预训练模型并加载；
3. 训练图像变换网络：使用训练集对图像变换网络进行训练，设置超参数；
4. 生成风格化图像：使用训练好的图像变换网络对测试图像进行风格迁移，生成风格化图像；

## 4 实验结果与分析

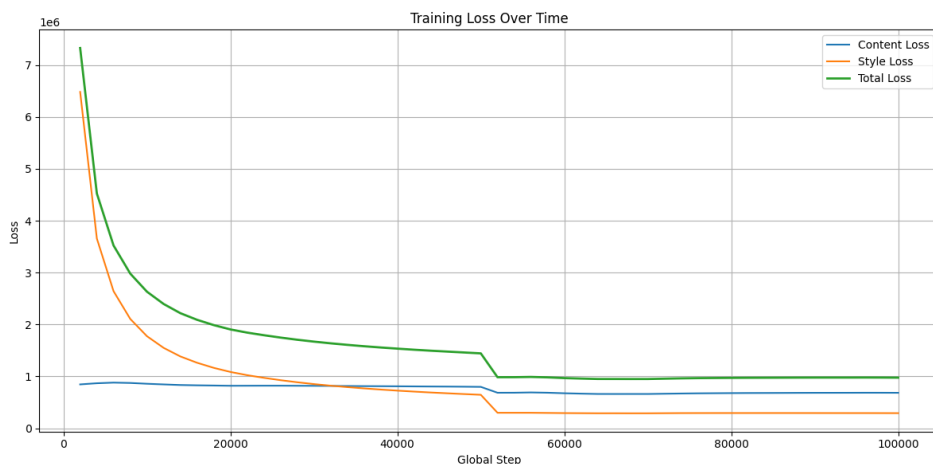


图 2: 训练过程中损失函数的变化曲线

图 2 展示了训练过程中损失函数的变化曲线。可以看到，随着训练的进行，内容损失和风格损失逐渐减小，总损失也趋于平稳，表明模型在不断学习和优化。最终，模型的总 Loss 收敛在  $0.9 \times 10^6$  左右。

当前 20000 次迭代时，内容损失和风格损失的变化较大，说明模型在学习初期对内容和风格特征的提取能力较弱。随着迭代次数的增加，损失函数逐渐收敛，表明模型对内容和风格特征的提取能力逐渐增强。而最后 60000 次迭代时，损失函数的变化幅度较小，说明模型已经趋于收敛，内容和风格特征的提取能力达到了较高水平。



图 3: 不同迭代次数下的风格迁移结果，风格分别为《星月夜》和《撑阳伞的女人》

图 3 展示了不同迭代次数下的风格迁移结果。可以看到，随着迭代次数的增加，生成的风格化图像质量逐渐提高，细节更加丰富，纹理更加自然。

风格迁移所用时间在 0.05 秒左右。

## 5 实验总结

本次实验实现了基于深度学习的图像风格迁移方法，使用了 Fast Style Transfer 模型。通过对模型的训练和测试，验证了该方法在风格迁移任务中的有效性和高效性。

## 参考文献

- [1] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [2] Alexei A Efros and Thomas K Leung. “Texture synthesis by non-parametric sampling”. In: *Proceedings of the seventh IEEE international conference on computer vision*. Vol. 2. IEEE. 1999, pp. 1033–1038.
- [3] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. “A neural algorithm of artistic style”. In: *arXiv preprint arXiv:1508.06576* (2015).
- [4] 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*. Springer. 2016, pp. 694–711.
- [5] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).