



河南大學

明德新民 止于至善

第12章 图像风格化处理

陈小潘

计算机与信息工程学院



1. 图像风格化技术基础

摄影爱好者一定接触过滤镜。

滤镜能改变照片的颜色样式，从而使风景照更加锐利或者令人像更加美白。但一个滤镜通常只能改变照片的某个方面。

如果要照片达到理想中的样式，经常需要尝试大量不同的组合，其复杂程度不亚于模型调参。

1. 图像风格化技术基础

■ 图像风格化，即风格迁移（Style Transfer），是指对图像进行处理以生成具有艺术效果图像的一种处理技术。图像风格迁移利用算法学习现有图像风格，再把将其移植到另一张目标图像上。

- ✓ 输入：内容图像和风格图像，内容图像按照风格图像生成定制的风格
- ✓ 输出：风格迁移后的结果图

1. 图像风格化技术基础

■ 什么是风格？

- ✓ 风格通常是指图像的**纹理、色彩、线条等视觉特征的统计属性**，它代表了图像的艺术风格或者视觉外观。
- ✓ 风格在图像中的表现形式非常多样化，可以是油画风格、水彩画风格、印象派风格等各种形式的艺术风格，也可以是具有特定结构和纹理的视觉外观。

1. 图像风格化技术基础

■ 不同的艺术风格



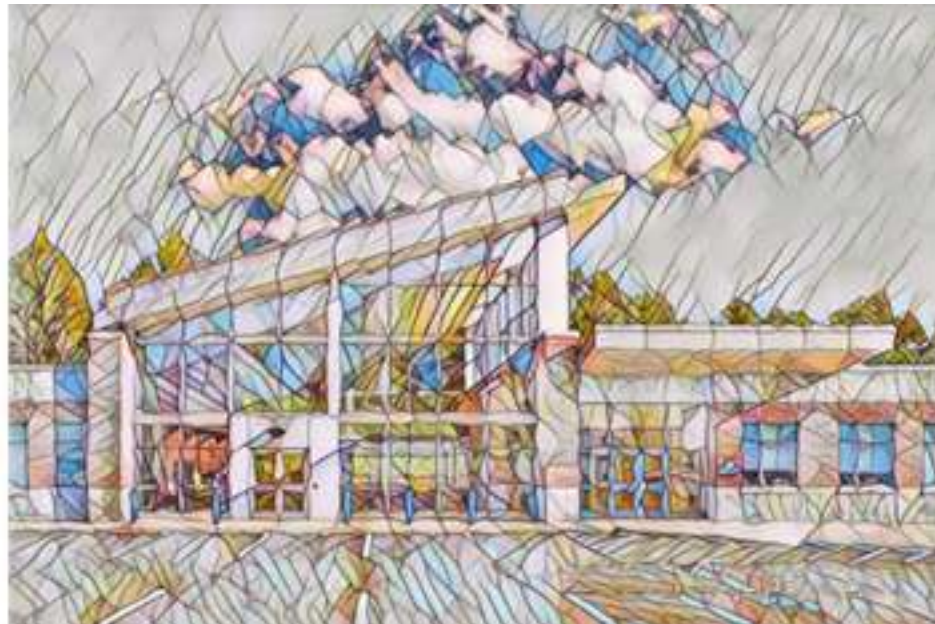
1. 图像风格化技术基础

风格迁移的目的

- ✓ 把提取的风格迁移并融合到内容图像上
- ✓ 生成结果既保留风格图像中色彩与线条组合排列的风格特征，又保留图像中内容
- ✓ 在目标结果中，风格与内容完美地融合

1. 图像风格化技术基础

风格迁移结果实例



1. 图像风格化技术基础

风格迁移结果实例



原始图像



梵高画作《星空》



风格迁移后的图像

1. 图像风格化技术基础

把风格图像 X_s 的绘画风格融入到内容图像 X_c ，得到一幅新的图像 X_n 。则新的图像 X_n 既要保持内容图像 X_c 的原始图像内容（内容画是一部汽车，融合后应仍是一部汽车，不能变成摩托车），又要保持风格图像 X_s 的特有风格（比如纹理、色调、笔触等）。

1. 图像风格化技术基础

■ 传统风格迁移的思路

- ✓ 分析某一种图像的风格，建立数学或者统计模型
- ✓ 修改目标图像的风格，使之更好的符合建立的模型
- ✓ 取得的效果基本能满足要求

■ 传统风格迁移存在主要的缺点

- ✓ 算法不具备通用性，一个程序基本只能做某一种风格或者某一个场景。
- ✓ 基于传统风格迁移技术的应用很有限

2.基于深度学习的风格迁移

- 2015年，Gatys等人发表在Arxiv上的文章《A Neural Algorithm of Artistic Style》，首次使用深度学习进行艺术画风格学习，是开山之作。
- 文章后来修改后，改名为Image Style Transfer Using Convolutional Neural Networks，发表在CVPR 2016上。
- 该研究基于深度神经网络的人工智能生成策略，可以生成高感知品质的艺术图片，使用神经网络重新生成内容和风格结合的图像。

2.基于深度学习的风格迁移

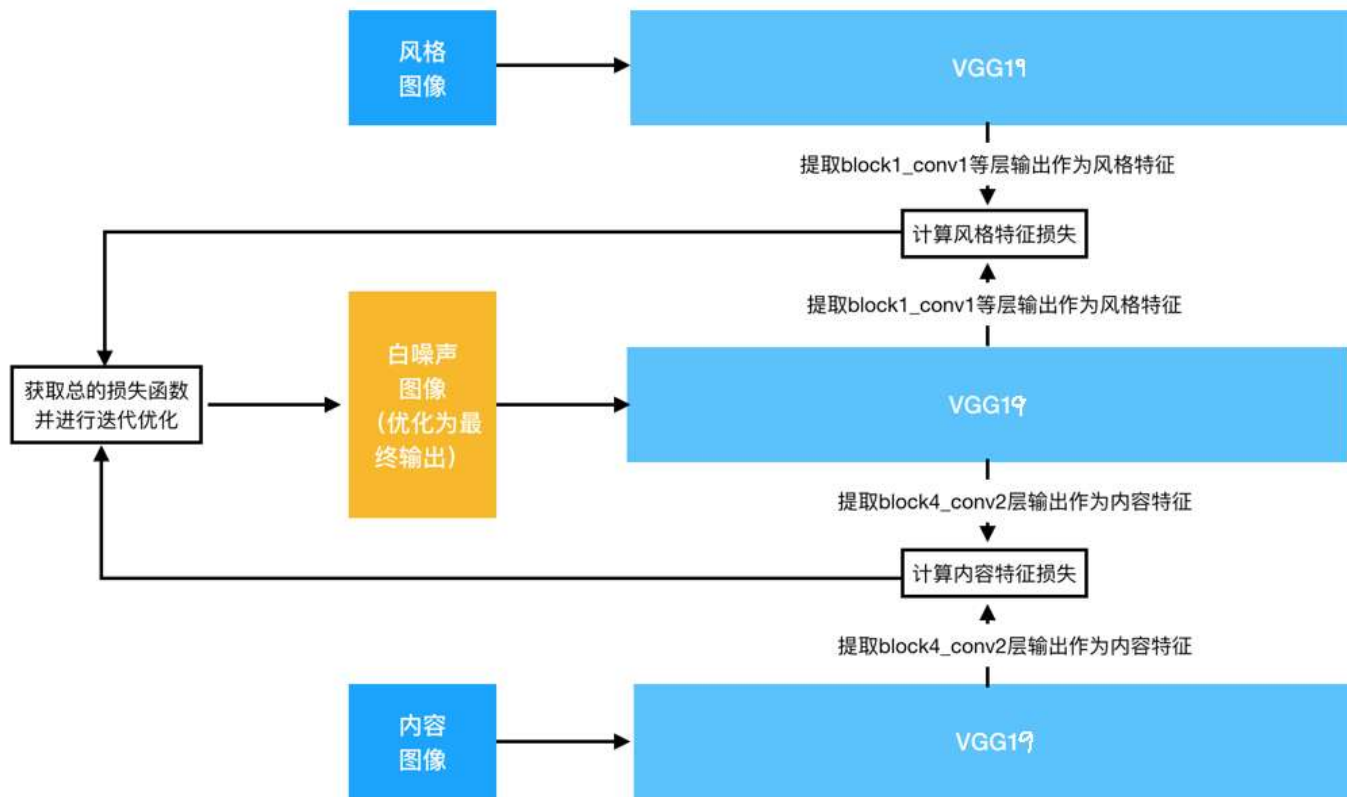
- 论文中定义了内容图像和风格图像这两个概念，并通过Gram矩阵作为风格的度量标准。给定一张内容图像和一张风格图像，该方法能保持内容图像的内容不变，但是将其风格转换为目标的风格图像。
- 算法从一张白噪声图像出发，利用生成算法不断生成目标图像。在优化期间，不断对生成的特征进行约束，使得学习得到的内容特征与输入图像的内容接近，使学习得到的风格特征与输入的风格图像特征接近，从而获得风格迁移后的图像。

2.基于深度学习的风格迁移

该算法的本质是利用深度卷积网络对图像输入的抽象，主要是三部分：

- 将风格图像输入卷积神经网络，将某些层输出作为风格特征（做一次）；
- 将内容图像输入卷积神经网络，将某些层输出作为内容特征（做一次）；
- 不断优化一个随机图像，使得它在该卷积神经网络的对应层输出不断接近上述两个图像的风格和内容特征（迭代）；

2.基于深度学习的风格迁移



2.基于深度学习的风格迁移

(1) 算法主要利用VGG19网络，使用了16个Conv层和5个Pooling层。将VGG19的Max Pooling替换为Mean Pooling（实验中发现均值池化效果更好），没有使用全连接。

(2) 利用梯度下降算法，找到与原图特征（内容或者风格）相匹配的目标结果，它们会使损失较小，因此得以保留下来。

2. 基于深度学习的风格迁移

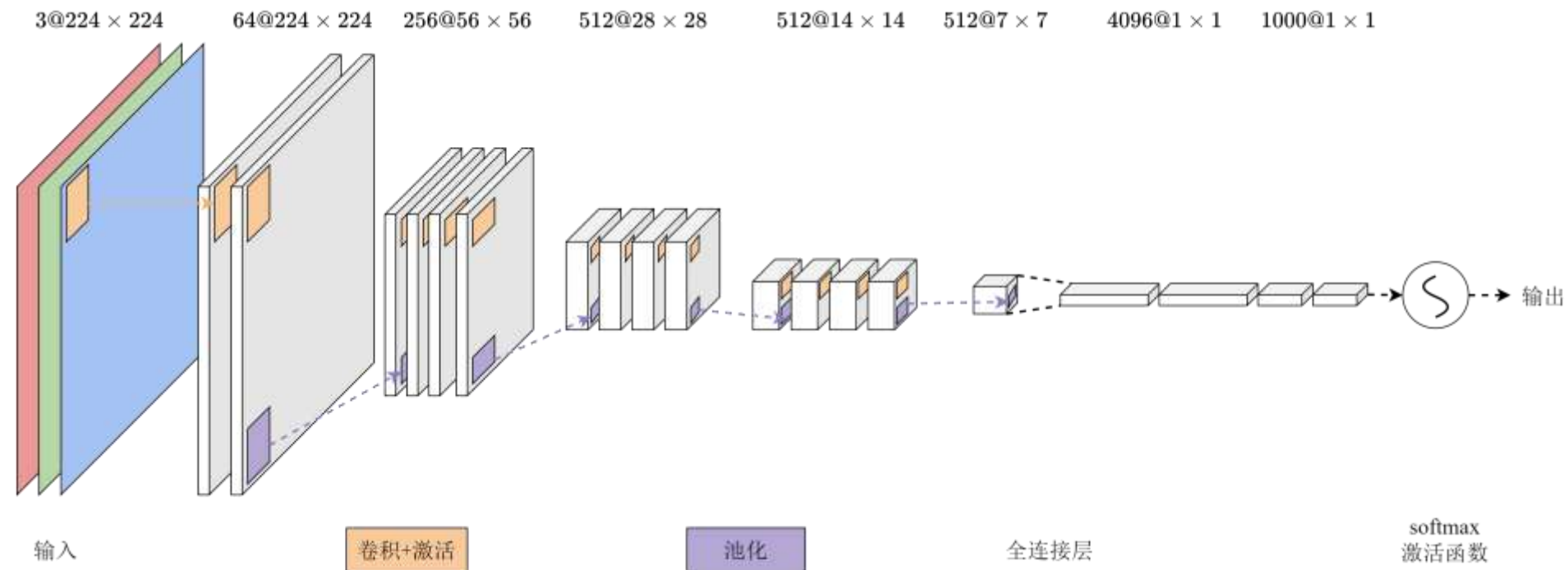


图 VGG网络结构示意图

2.基于深度学习的风格迁移

■ 风格和内容在卷积神经网络中表达是独立的

- ✓ 可以独立控制两种表达，产生不同的合成结果
- ✓ 生成（内容和风格）不同混合方案的结果
- ✓ 例如，描绘不同时期著名风格的艺术作品

- 在CNN网络中，一般认为**较低层**的特征描述了图像的具体视觉特征（即纹理、颜色等），**较高层**的特征则是较为**抽象**的图像内容描述。

低层次响应描述图像的风格，高层次响应描述图像的内容。

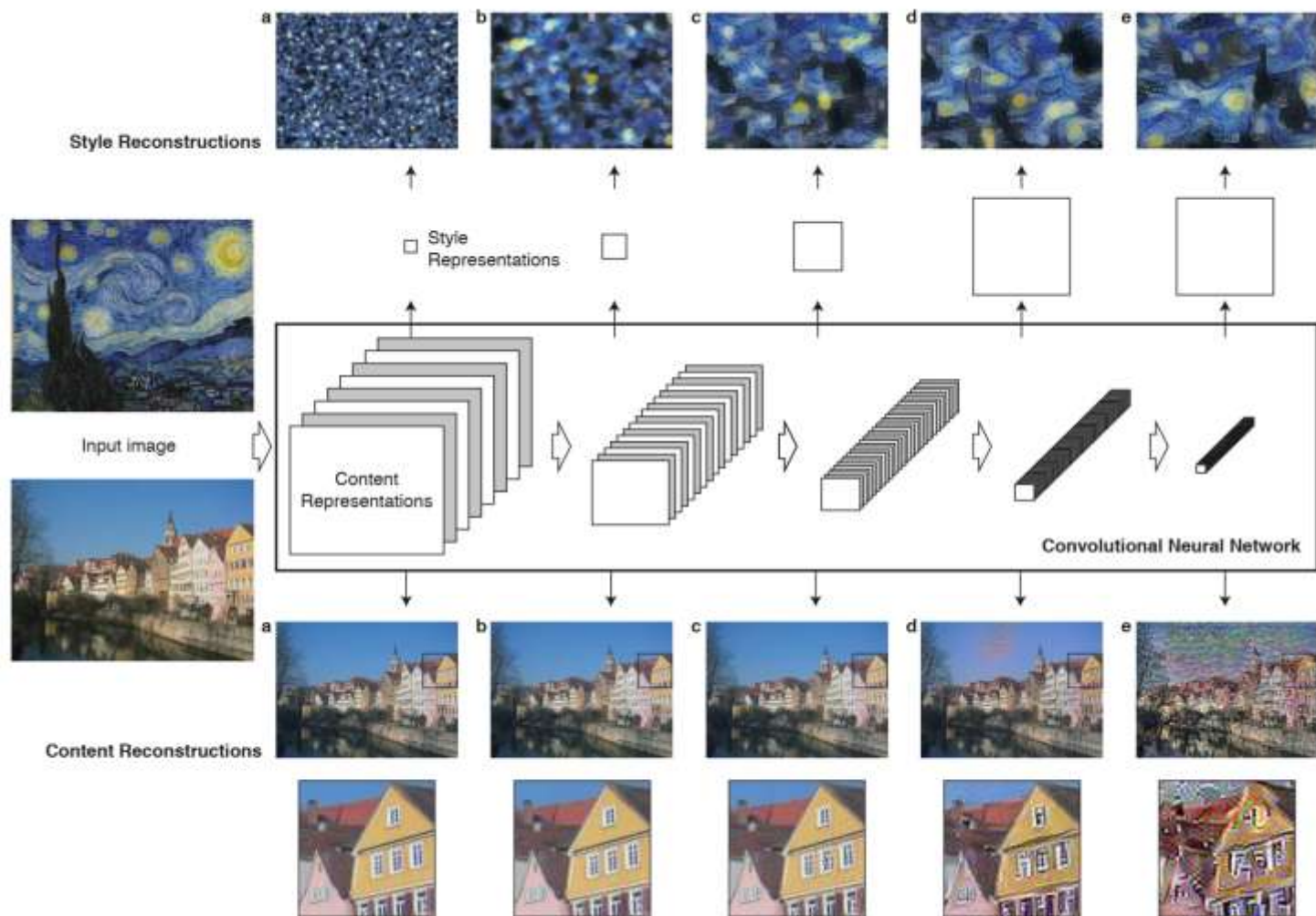


Figure 1. Image representations in a Convolutional Neural Network (CNN). A given input image is represented as a set of filtered images

2.基于深度学习的风格迁移

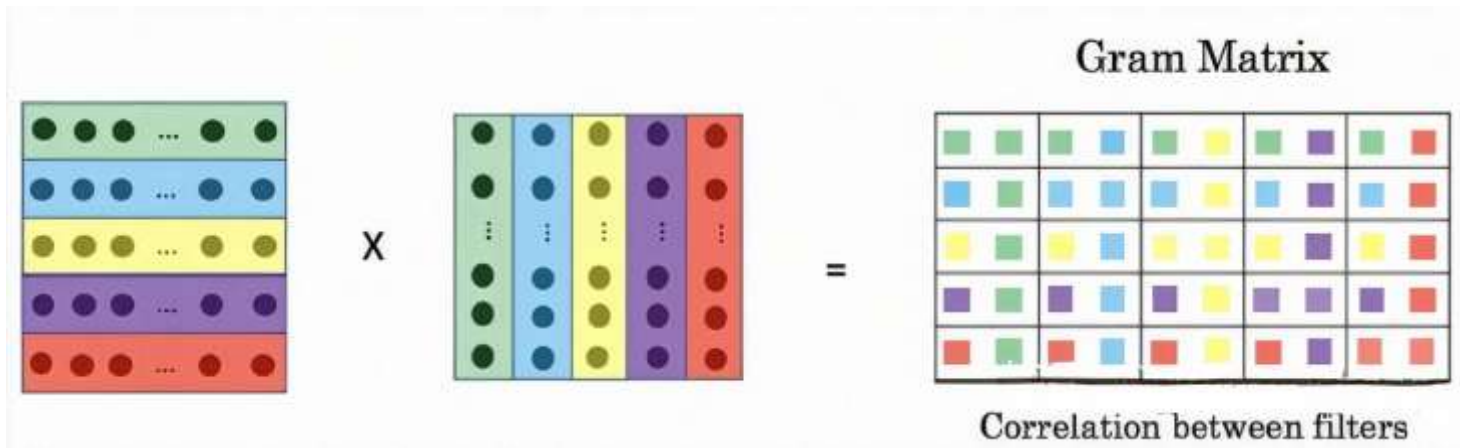
内容损失：要比较两幅图像的内容相似性，可以通过比较两幅图像在CNN网络中**高层特征**的相似性（欧氏距离）。

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)$$

第 l 层中的响应（activation）存储在矩阵 $F^l \in \mathcal{R}^{N_l \times M_l}$ 中，其中 F_{ij}^l 是层 l 中位置 j 处第 i 个滤波器的激活。 P_{ij}^l 表示第 l 层网络对内容图像学习得到的编码特征。其中， p 为内容来源图像， x 为生成图像（最初输入为白噪声图像）。

2.基于深度学习的风格迁移

Gram矩阵（格拉姆矩阵）是线性代数中的一个重要概念，指的是n维欧式空间中任意 k 个向量之间两两的内积所组成的矩阵。Gram矩阵是向量间内积组成的矩阵。



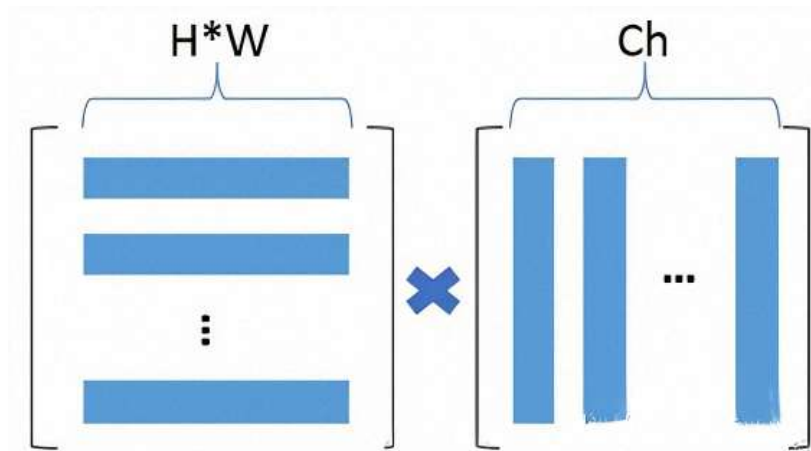
2.基于深度学习的风格迁移

Gram矩阵在多个领域有广泛应用，包括：

- 1.机器学习：表示向量之间的相似性和关联。
- 2.信号处理：分析信号的协方差结构。
- 3.统计学：与PCA等降维技术密切相关。
- 4.控制论：分析系统的可观性和可控性。

2.基于深度学习的风格迁移

输入图像的feature map为 $[ch, h, w]$ ，经过flatten（即是将 $h*w$ 进行平铺成一维向量）和矩阵转置操作，可以变形为 $[ch, h*w]$ 和 $[h*w, ch]$ 的矩阵。再对两个作内积得到Gram Matrices。（蓝色条表示每个通道flatten后特征点，最后得到 $[ch * ch]$ 的G矩阵）



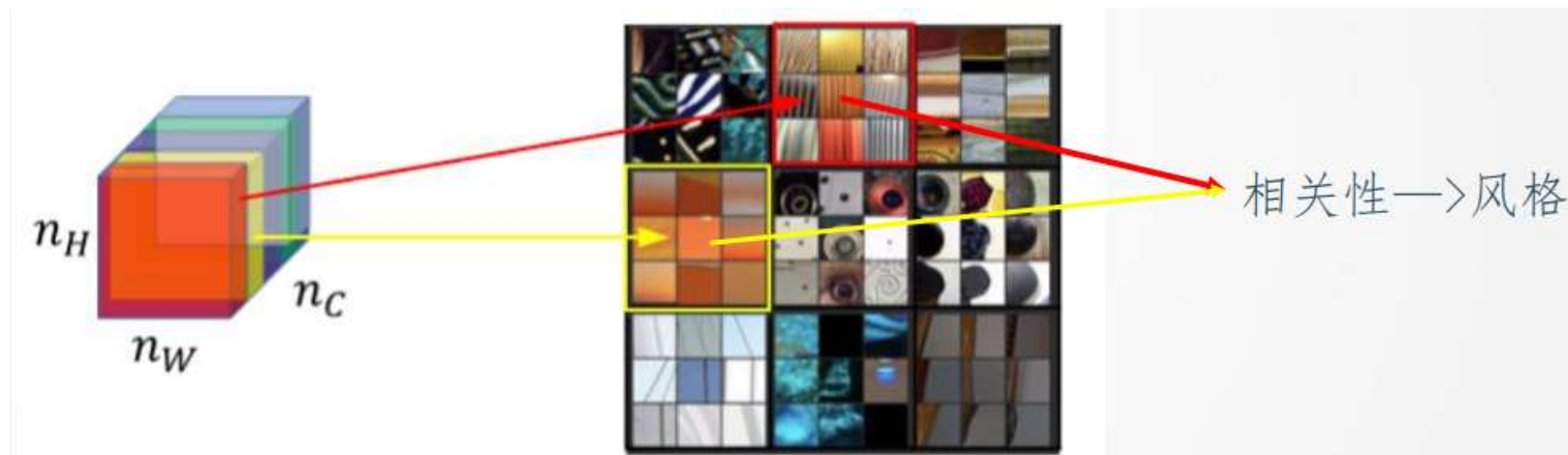
2.基于深度学习的风格迁移

基于Gram的风格损失

风格的数学表达通常采用基于卷积神经网络的特征表示。可以通过在训练的卷积神经网络中的不同层次提取特征，并计算这些特征之间的统计属性来表示风格。一种常用的方法是计算特征图之间的Gram矩阵，它反映了特征图之间的相关性和纹理信息。在计算乘积的过程中，把与特征在图像中的位置有关的维度消掉了，只留下了和卷积核有关的维度。直观上来说，这体现了图像的风格特征与其相对位置无关，是图像的全局属性。

2.基于深度学习的风格迁移

基于Gram的风格损失



2.基于深度学习的风格迁移

基于Gram的风格损失

采用了Gram Matrix来表示图像的风格，Gram Matrix定义为：

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$$

Gram矩阵可以看作体现了不同filter特征的相互关系，同时忽略了内容上的信息。具有 N_l 个不同过滤器的卷积层有 N_l 个特征图，每个特征图的大小为 M_l ，其中 M_l 是特征图的高度乘以宽度。

定义 l 层的风格损失为： $E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$

2.基于深度学习的风格迁移

由于不同卷积层的参数量可能不同，我们额外除以 $4N_l^2M_l^2$ 来进行层间的归一化。最后，再用权重对不同卷积层的损失做加权平均，得到总的风格损失为：

$$\mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^L w_l E_l$$

\vec{a} 和 \vec{x} 分别表示原始图像和生成的图像， A^l 和 G^l 代表了原始图像和生成图像在 l 层的风格表示。

总损失

$$\mathcal{L}_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{\text{style}}(\vec{a}, \vec{x})$$

2. 基于深度学习的风格迁移

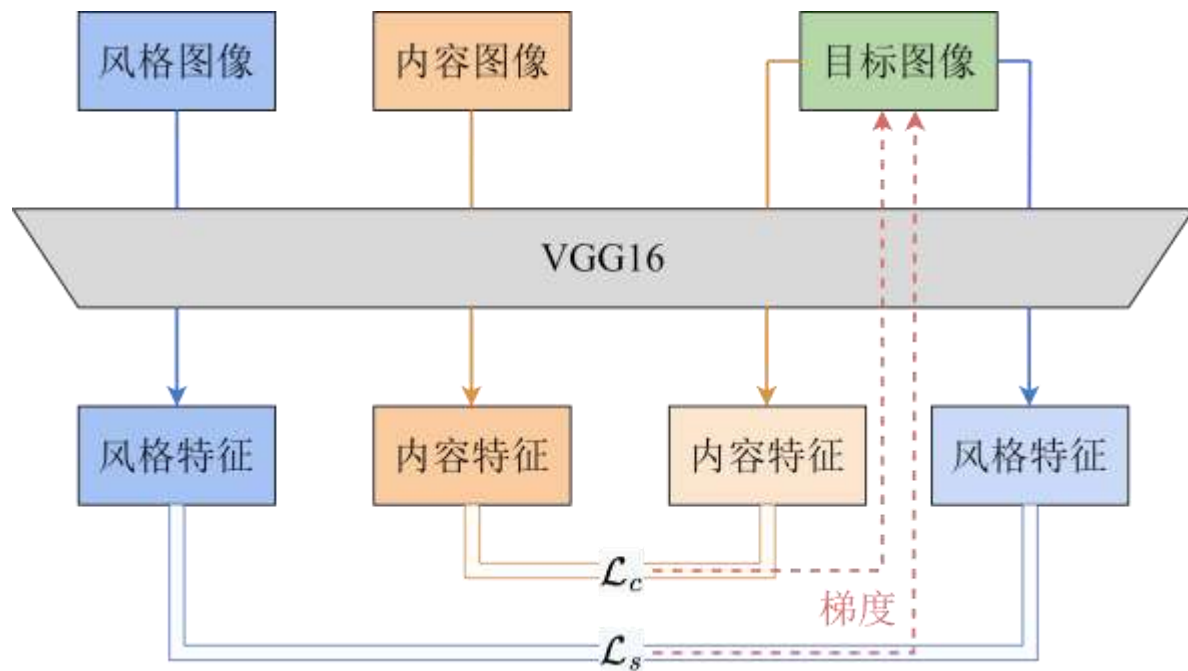


图 风格迁移的训练流程

A



B



C



D



E



F



3. 离线处理的风格化技术

- Johnson J, Alahi A, Li F F (李飞飞). Perceptual Losses for **Real-Time Style Transfer** and **Super-Resolution** [C]// European Conference on Computer Vision. 2016
- 对现有问题（实例1）的分析：

2015年，Gatys发表两篇风格迁移的研究论文，第一个基于神经网络的图像风格迁移算法，生成时间：5-20 分钟

3. 离线处理的风格化技术

■ 特点

- ✓ 通过预训练好的网络提取高级特征、定义并优化感知损失函数来产生高质量的风格化结果。采用感知损失函数训练前馈网络实现迁移后的结果
- ✓ 将低分辨率图像输入，可得到高分辨率图像，只进行一次网络前向传播的计算，速度非常快，可以达到实时效果
- ✓ 训练一个前馈网络解决实时优化问题，达到快速生成的性能

3. 离线处理的风格化技术

Style
Transfer

Style



Content



Gatys *et al.* [11]



Ours



3. 离线处理的风格化技术

■ 克服问题

对于实例1（现有方法），对于每一张要生成图片，都需要初始化，然后保持CNN的参数不变，反向传播更新图像，得到最后的结果，性能问题较差。

■ 吸取以下方法的优点：

- ✓ 训练前馈卷积神经网络，输出图像与原图像逐像素级的损失函数
- ✓ 通过用预训练好的网络提取高级特征、定义并优化感知损失函数

3. 离线处理的风格化技术

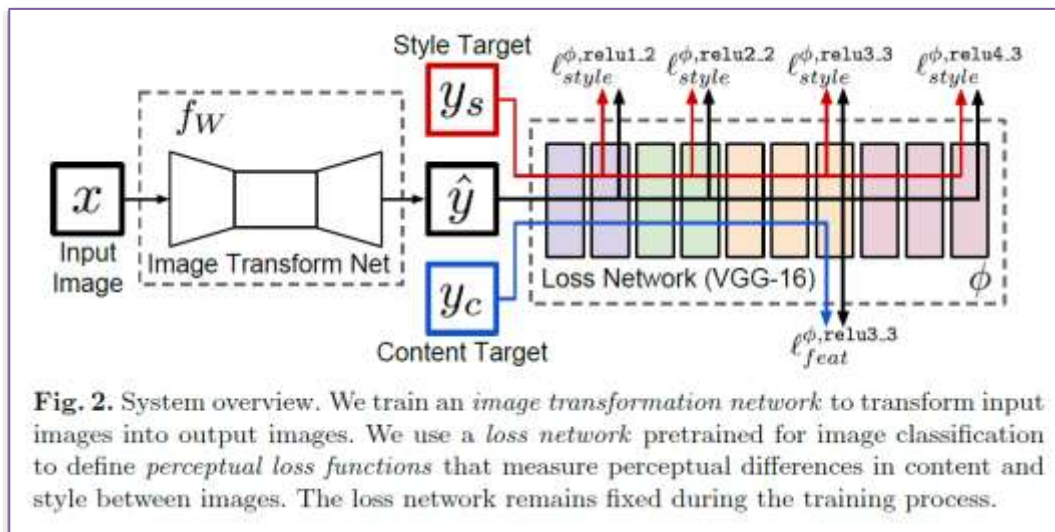
- 不是基于像素间的相似性，而是利用抽取的高维特征之间比对

high-quality images can be generated using differences between high-level image feature representations extracted from pretrained convolutional neural networks. Images are generated by minimizing a loss function.
This strategy has been applied to feature inversion

3. 离线处理风格化技术

■ 系统由两部分组成

- ✓ 图像转换网络
- ✓ 损失网络 ϕ （用来定义一系列损失函数 l_1, l_2, l_3 ）



3. 离线处理的风格化技术

图片转换网络

- 网络结构：深度残差网络 f_W
- 功能：将输入的图像 x 通过映射 $\hat{y} = f_W(x)$ ，转换成输出图像
- 损失函数利用 \hat{y} 和目标图像之间 y_i 的差异计算 $\ell_i(\hat{y}, y_i)$
- 用随机梯度下降SGD（stochastic gradient descent）训练，使得 W^* 保持下降

$$W^* = \arg \min_W \mathbf{E}_{x, \{y_i\}} \left[\sum_{i=1} \lambda_i \ell_i(f_W(x), y_i) \right]$$

3. 离线处理的风格化技术

损失控制网络 ϕ

- 针对图像分类预训练的卷积神经网络已经学会了编码想要在损失函数中测量的感知和语义信息。因此，使用预训练的图像分类网络 ϕ 作为固定的损失网络，进而定义损失函数
- 网络 ϕ 损失定义由特征（内容）损失 ℓ_{feat}^{ϕ} 和风格损失 ℓ_{style}^{ϕ}

3. 离线处理的风格化技术

- 对于每一张输入的图像 x ，有一个内容目标 y_c 和一个风格目标 y_s
- 风格转换：内容目标 y_c 就是输入图像 x
 - ✓ 输入： y_c
 - ✓ 输出： \hat{y} ，应将风格 y_s 结合到内容 $x = y_c$ 上

3. 离线处理的风格化技术

(1) 图片转换网络 f_W 的结构

- 网络设计遵循DCGAN(Deep Convolutional Generative Adversarial Network)的思路
- 不使用任何池化层，使用跨步和分段跨步卷积进行上采样和下采样
- 网络由5个残差模块组成，非残差模块后面都连接有空间批归一化处理和线性的激活函数
- 输出层则使用缩放的tanh来确保输出图像的像素在[0, 255]范围
- 除第一层和最后一层使用 9×9 的卷积核外，其它卷积核是 3×3

3. 离线处理的风格化技术

■ 输入输出

- ✓ 对于风格迁移来说，输入和输出的大小都是 $256 \times 256 \times 3$ 。
- ✓ 对于图片超分辨率重建来说，输出是 $288 \times 288 \times 3$ ，而输入是 $288/f \times 288/f \times 3$ ， f 是比例因子
- ✓ Transform Net是全卷积的，可以支持任意的尺寸

3. 离线处理的风格化技术

对于风格转换来说，使用2个 $\text{stride}=2$ 的卷积层来做下采样，每个卷积层后面跟着若干个残差块。然后跟着两个 $\text{stride}=1/2$ 的卷积层来做上采样。虽然输入和输出相同，但有两个优点：

- ✓ 提高性能，减少了参数
- ✓ 大视野，风格转换会导致物体变形，因而，结果图像中的每个像素对应着的初始图像中的视野越大越好。

3. 离线处理的风格化技术

■ Residual Connections

- ✓ 现有工作用残差连接训练非常深的网络用来做图像分类，残差连接能让网络更容易得到确定的函数
- ✓ 网络由几个残差块组成，每个包含两个 3×3 的卷积层，设计的残差块，有助于学习得到identity function（恒等函数）

3. 离线处理的风格化技术

- 定义了两个感知损失函数，用来衡量两张图片之间高级的感知及语义差别。
- 用一个预训练好的VGG-16模型进行设计，使用Imagenet的数据集进行预训练
- 定义了两种损失函数：特征（内容）损失和风格损失。使用 ϕ 来表示损失网络。

3. 离线处理的风格化技术

- **特征（内容）重构损失：**不做逐像素对比，而是用VGG计算得到高级特征（内容）表示，损失定义如下：

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j} \| \phi_j(\hat{y}) - \phi_j(y) \|_2^2$$

- ✓ j 表示网络的第 j 层。
- ✓ C_j 、 H_j 、 W_j 表示第 j 层的feature map的尺度（如果 j 是卷积层的话）
- ✓ $\phi_j(x)$ 是网络 ϕ 在处理图像 x 时第 j 层的激活值

3. 离线处理风格化技术

- 使用不同层的重建效果如下：

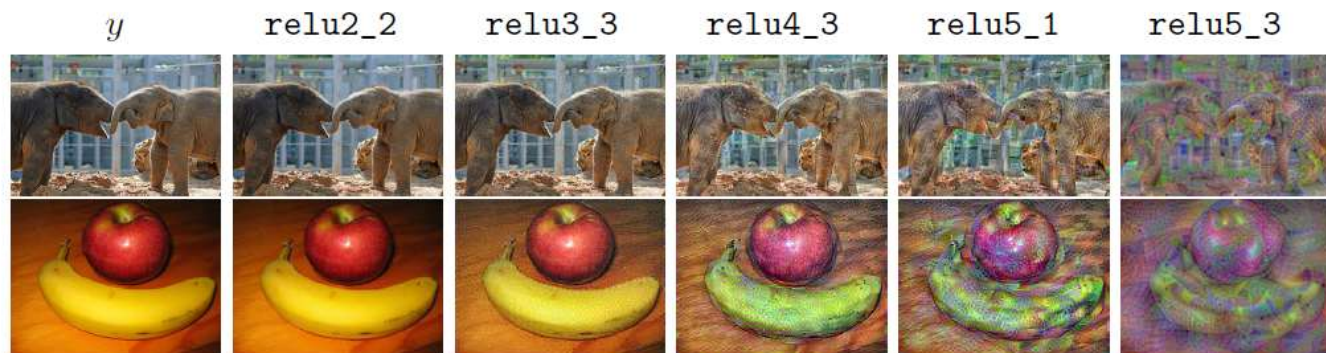


Fig. 3. Similar to [6], we use optimization to find an image \hat{y} that minimizes the feature reconstruction loss $\ell_{feat}^{\phi,j}(\hat{y}, y)$ for several layers j from the pretrained VGG-16 loss network ϕ . As we reconstruct from higher layers, image content and overall spatial structure are preserved, but color, texture, and exact shape are not.

3. 离线处理的风格化技术

- 对于风格重建的损失函数，首先要先计算Gram矩阵

$$G_j^\phi(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'}$$

第j层feature map尺度 $C_j H_j W_j$ ，可以看成是 C_j 个特征，这些特征两两之间内积

- 在loss网络的每一层都求出Gram矩阵，然后该层的损失计算为：

$$\ell_{style}^{\phi,j}(\hat{y}, y) = \| G_j^\phi(\hat{y}) - G_j^\phi(y) \|_F^2$$

- 最后将不同层的欧氏距离相加，得到最后的风格损失。

3. 离线处理的风格化技术

使用不同层的风格重建效果如下

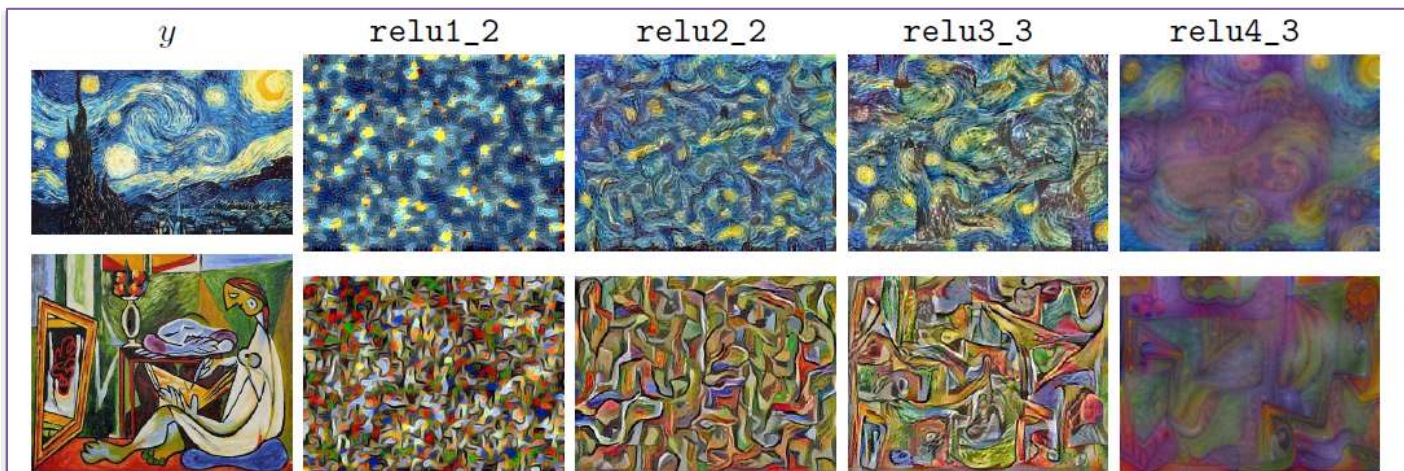


Fig. 4. Similar to [10], we use optimization to find an image \hat{y} that minimizes the style reconstruction loss $\ell_{style}^{\phi,j}(\hat{y}, y)$ for several layers j from the pretrained VGG-16 loss network ϕ . The images \hat{y} preserve stylistic features but not spatial structure.

3. 离线处理的风格化技术

- 除了感知损失，还基于低维的像素信息定义了两种简单损失函数：**像素损失**和**总方差规范化损失**。

✓ **像素损失**：像素损失是计算输出图像和目标图像之间标准化的差距。

如果两者的形状都是 $C \times H \times W$ ，那么像素损失就是 $\ell_{pixel}(\hat{y}, y) = \|\hat{y} - y\|_2^2 / CHW$ 。这只能用于有完全确定的目标，让这个网络去做完全匹配

✓ **总方差规范化**（Total Variation Regularization）：为使得输出图像比较平滑，超分辨率重建研究使用了全变差正则化 $\ell_TV(\hat{y})$ ，用于控制模型参数的幅度。

3. 离线处理的风格化技术



《星夜》文森特·梵高



缪斯，巴勃罗·毕加索

3. 离线处理的风格化技术

瓦西里·康定斯基的抽象画

Style
Composition VII,
Wassily
Kandinsky, 1913



Style
The Great Wave off
Kanagawa, Hokusai,
1829-1832



Content

[10]

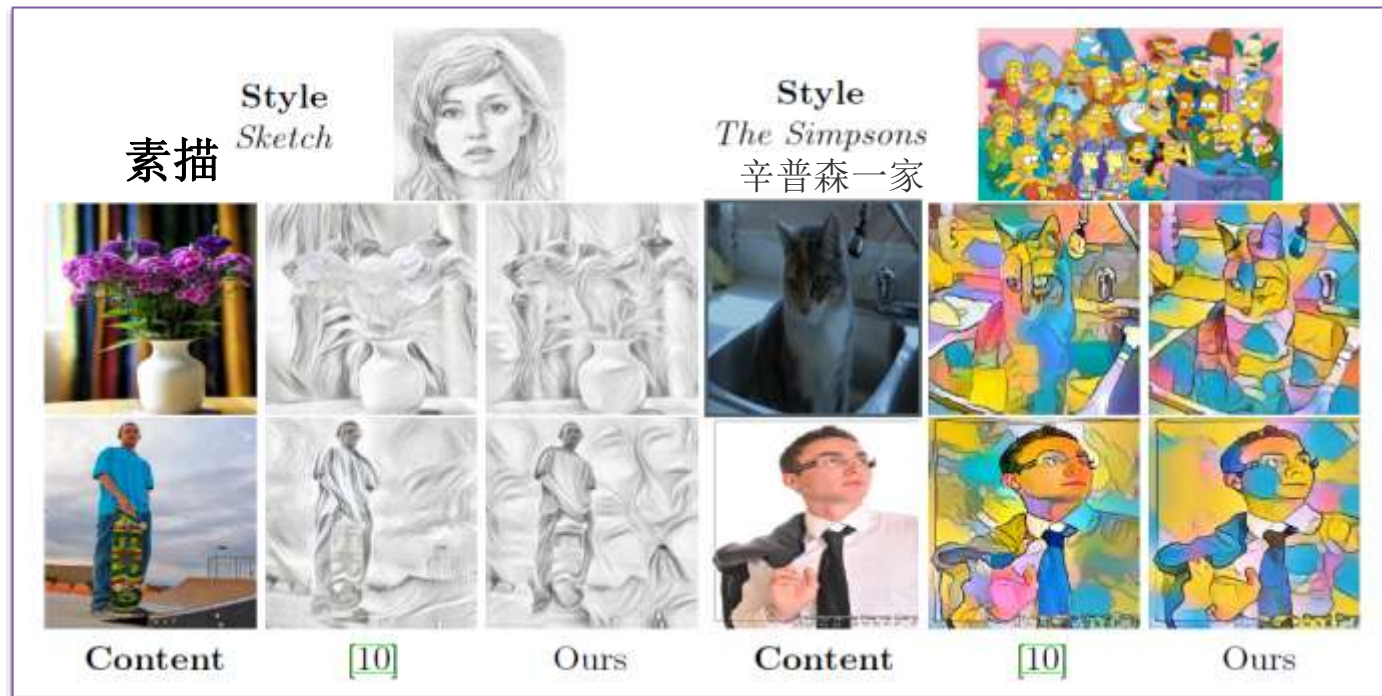
Ours

Content

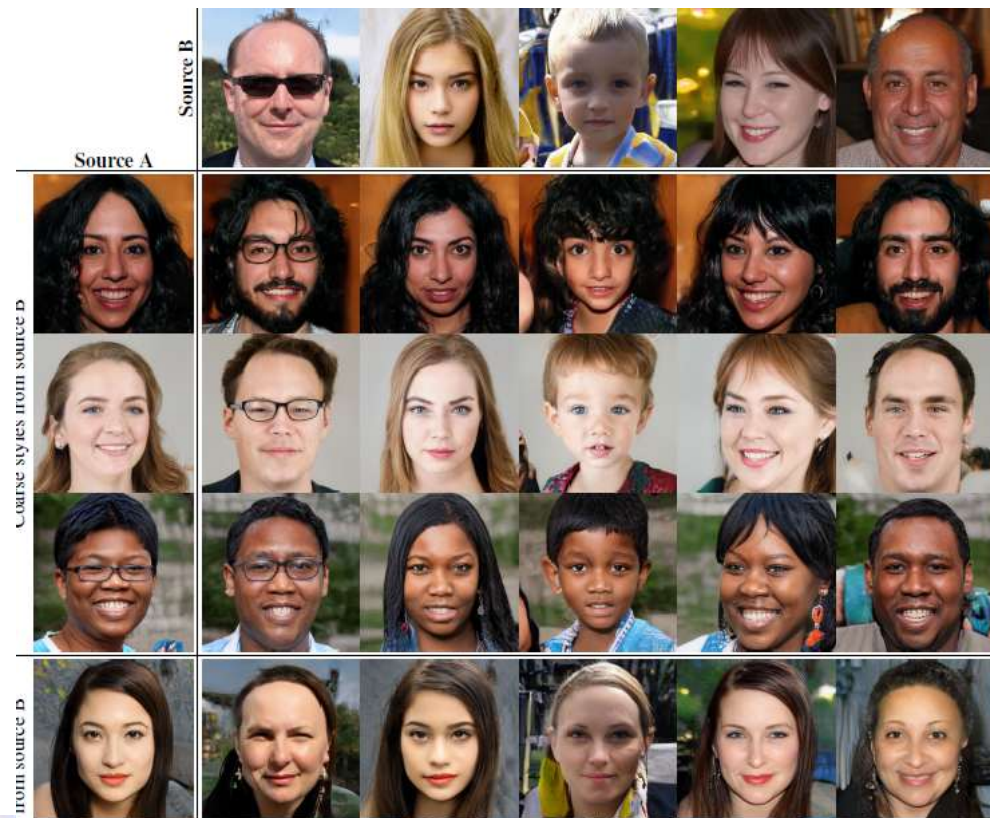
[10]

Ours

3. 离线处理的风格化技术



4. StyleGAN系列



4. StyleGAN系列

A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019

StyleGAN中的“**Style**”是指数据集中人脸的主要属性，比如人物的姿态等信息，而不是风格转换中的图像风格，这里Style是指人脸的风格，包括了脸型上面的表情、人脸朝向、发型等等，还包括纹理细节上的人脸肤色、人脸光照等方方面面。

4. StyleGAN系列

在StyleGAN中，大多数的生成的图像容易产生一个类似水滴状的伪影，并且这些伪影在64 x 64 后就便存在在特征图中。



Figure 1. Instance normalization causes water droplet -like artifacts in StyleGAN images. These are not always obvious in the generated images, but if we look at the activations inside the generator network, the problem is always there, in all feature maps starting from the 64x64 resolution. It is a systemic problem that plagues all StyleGAN images.

4. StyleGAN系列

StyleGAN2: **Analyzing and Improving the Image Quality of StyleGAN**

作者经过实验发现，AdaIN的归一化操作，是造成水滴伪影出现了根本原因，于是，作者把AdaIN层里面的归一化去掉。又经过实验发现，将噪声 \mathbf{B} 和偏置项 \mathbf{b} 移动出style模块之外，取得的效果更好。经过此操作后，作者发现归一化和 \mathbf{A} 映射的时候，只需要标准差就可以了（舍弃均值），如下图的图（c）

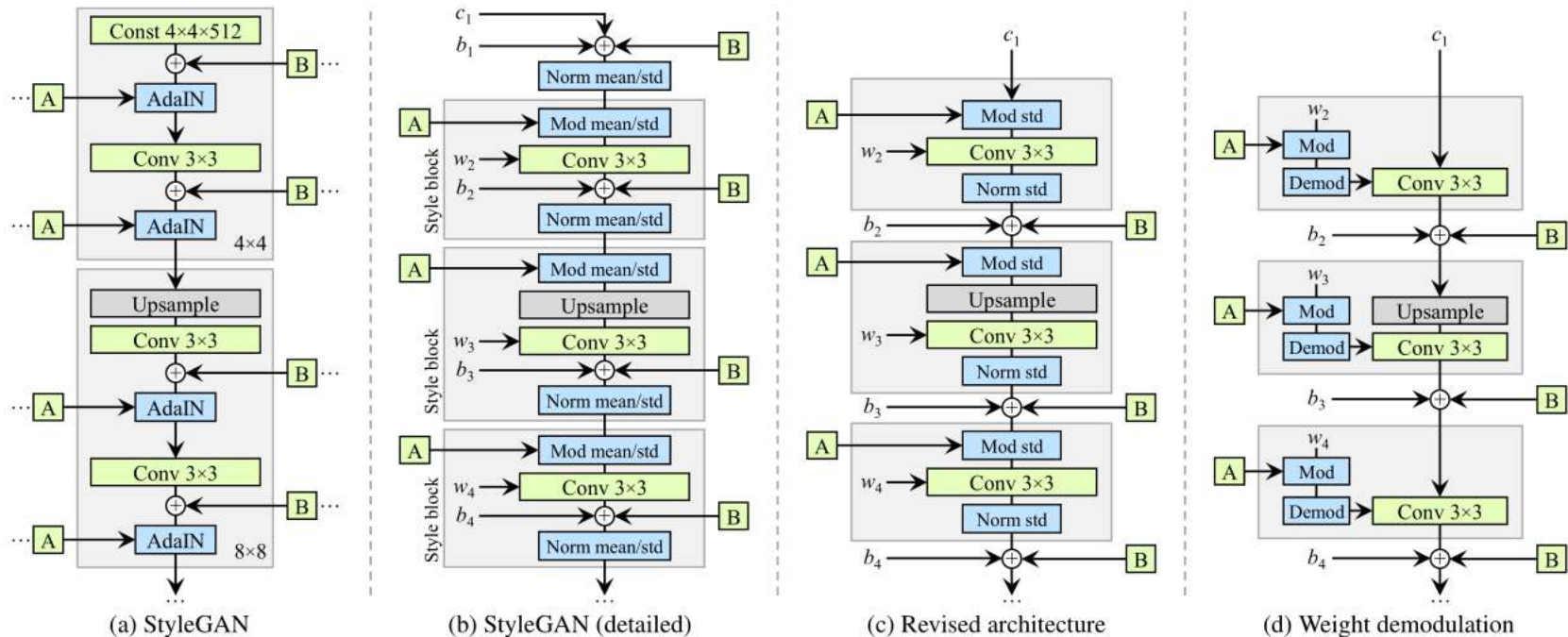


Figure 2. We redesign the architecture of the StyleGAN synthesis network. (a) The original StyleGAN, where \boxed{A} denotes a learned affine transform from \mathcal{W} that produces a style and \boxed{B} is a noise broadcast operation. (b) The same diagram with full detail. Here we have broken the AdaIN to explicit normalization followed by modulation, both operating on the mean and standard deviation per feature map. We have also annotated the learned weights (w), biases (b), and constant input (c), and redrawn the gray boxes so that one style is active per box. The activation function (leaky ReLU) is always applied right after adding the bias. (c) We make several changes to the original architecture that are justified in the main text. We remove some redundant operations at the beginning, move the addition of b and \boxed{B} to be outside active area of a style, and adjust only the standard deviation per feature map. (d) The revised architecture enables us to replace instance normalization with a “demodulation” operation, which we apply to the weights associated with each convolution layer.

4. StyleGAN系列

Alias-Free Generative Adversarial Networks, 不失真(无锯齿)的生成对抗网络。

一般称之为StyleGAN3, 是对StyleGAN2的改进, 主要是针对在StyleGAN2中平移或旋转时, 存在“粘滞”现象, 比如视频中, 头部的移动, 正常来说脸部上面的特征是要跟着一起移动的, 而在StyleGAN2中出现胡子没有很好贴着皮肤一起移动, 出现一种胡子像是粘在嘴巴上, 分离的感觉, 这种情况就是特征依赖绝对坐标, 而没有很好地做到跟随周边的特征一起移动。

4. StyleGAN系列

StyleGAN3的改进，使得即使在亚像素尺度上，它们也能够做到平移和旋转不变性，这样的结果为视频和动画的生成模型铺平了道路。图像是由像素组成的，其实像素之间存在空隙，而空隙之间存在的空间，叫作**亚像素**，比如做上采样时，常见的双线性插值就是在X，Y方向相邻像素之间分别采样，然后在采样点之间插入数值。

关于考核

1、**课程作业**：严格按照模版，格式规范、图表清晰、文字通顺，无错别字，参考文献规范，不能篇幅太短，首先要有量，再说质。**作业提交纸质打印版（A4双面打印，左侧装订）和电子版，保持二者一致。**

可以提交代码（得高分）。文件命名：学号-姓名-专业-数字图像处理课程论文。

截止时间：12月31日24时

2、**考试**：书写工整，回答问题（1）（2）.....，条理清楚，让阅卷老师很快找到答案；每个题认真分析，要多写，写满，不要一两句话的答案。**带铅笔、尺子。**

3、无论做什么，体现出你认真对待课程，对待作业，对待考试、严谨的态度

关于考核

■ 考试时间：12月24日 上午9点——11点

8:40到201机房准备排座位。

■ 考试时答案写在对应题号后的空白处（不得写在其他题号后，画个箭头）。

■ 年级：2025（不要写成研一，年级指的是入学年份；哪一届毕业生指的是毕业年份）

■ 认真书写自己的姓名（尤其是考试等非常正式的场所），让他人能辨认出来。

关于考核

姓名：于博珍 年级：研一

姓名：于博珍

姓名：于博珍

THE END
Thank You.