

基于VGG的图像风格迁移

引言

随着神经网络的发展，机器在某些视觉感知的关键领域，比如物体和人脸识别等有着接近于人类甚至超越人类的表现。

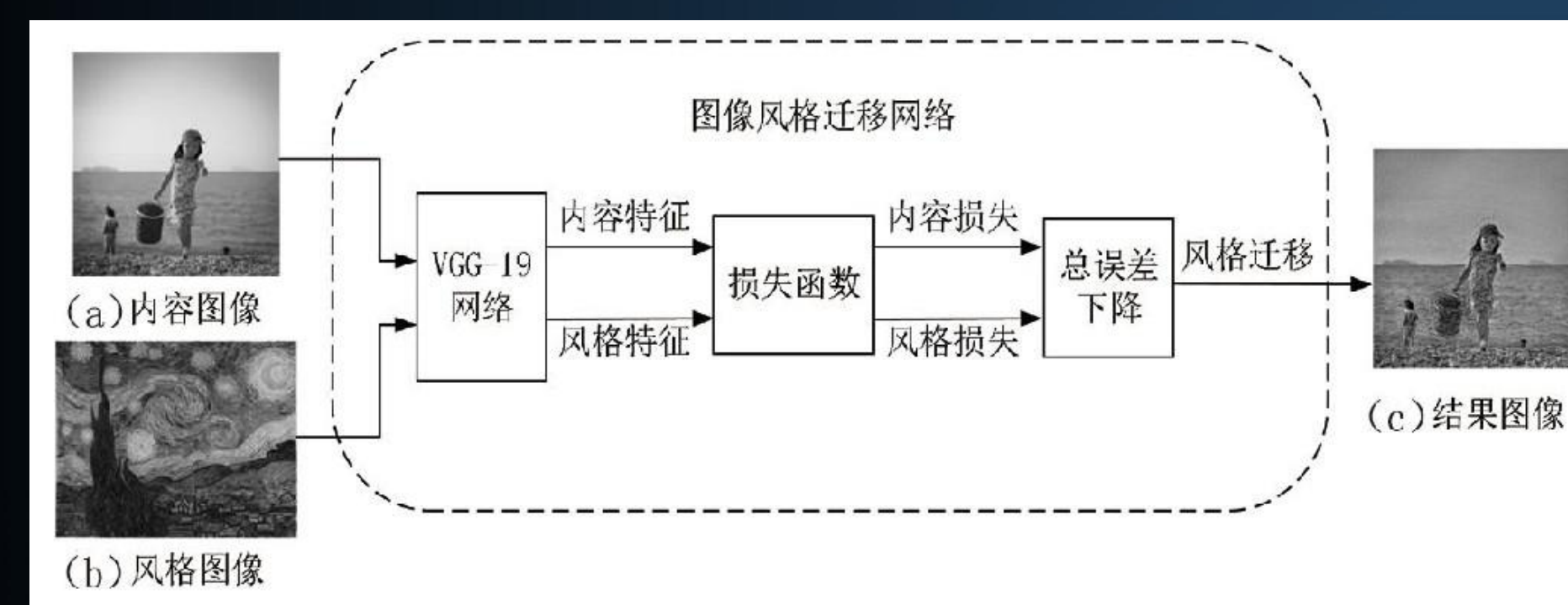
在神经网络之前，图像风格迁移的程序采用的思路是：分析一种风格的图像，为这种风格建立一个数学统计模型；再改变要做迁移的图像使它的风格符合建立的模型。该方法可以取得不同的效果，但是有一个较大的缺陷：一个模型只能实现一种图像风格的迁移。因此，基于传统方法风格迁移的模型应用十分有限。

本实验则用到了一种基于深度神经网络的机器学习模型——卷积神经网络，它可以分离并结合任意图片的风格和内容，生成具有高感知品质的艺术图片。相较于传统的风格迁移的方法，该方法具有更好的普适性。

实验简介

思路：图像风格迁移包括三部分：风格style特征提取（一般是纹理特征），目标图像内容content提取，style+content合成新的图片。

实验过程介绍：首先通过VGG-19神经网络模型分别提取出内容图像的内容特征和风格图像的风格特征，接着通过损失函数分别计算出内容损失和风格损失，并以总误差下降为目标，不断调整模型参数和迭代次数，最终实现图像风格迁移。

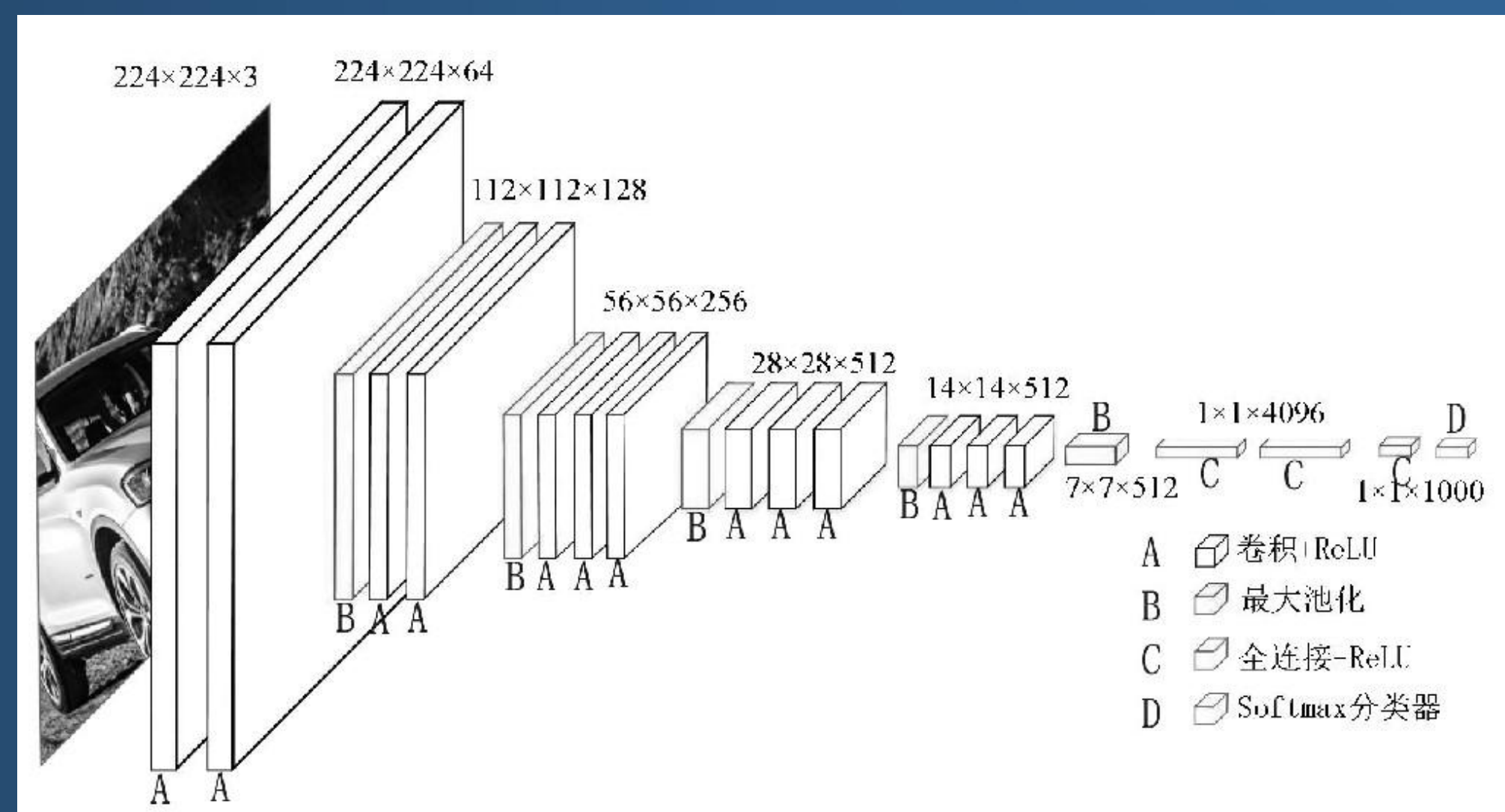


模型描述

•模型介绍：

•VGG模型是于2014年提出的图像分类模型，这一模型采用了简单粗暴的堆砌3*3卷积层的方式构建模型并花费了大量时间逐层训练，最终斩获了2014ImageNet图像分类比赛的亚军。

•这一模型的优点是结构简单，容易理解，便于利用到其他任务中，缺点是在计算方面会耗资更多的资源并且使用了更多的参数，导致更多的内存占用。其中绝大多数的参数都是来自于第一个全连接层。而且VGG是有3个全连接层的



损失函数

当卷积神经网络用于物体识别时，随着网络的层次越来越深，网络层产生的物体特征信息越来越清晰。即在CNN中，一般较低层描述了图像具体视觉特征（纹理颜色），较高层描述了图像的内容特征

通过根据内容和风格图来对一张白噪声图片进行训练，从而获得风格迁移后的图片。训练方法是和训练网络一样的梯度下降法，那么问题的关键就是如何构建损失函数了。

•内容损失公式：
$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

•风格损失：

Gram 矩阵：
$$G_{ij}^l = \sum_k F_{ik}^l F_{kj}^l$$

表示第L层中两个向量化特征矩阵的内积。

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

风格损失 $L_{style} = \sum_{l=0}^L w_l E_l$ 表示每一层的损失权重

总损失函数：

$$L_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{content}(\vec{p}, \vec{x}) + \beta L_{style}(\vec{a}, \vec{x})$$

设置权重和训练

由损失函数知，分配在内容和风格的权重不同时，合成产生的图片效果也完全不一样。因此需要适当地调整内容表示和风格表示的权重比来产生具有视觉感染力的图片。



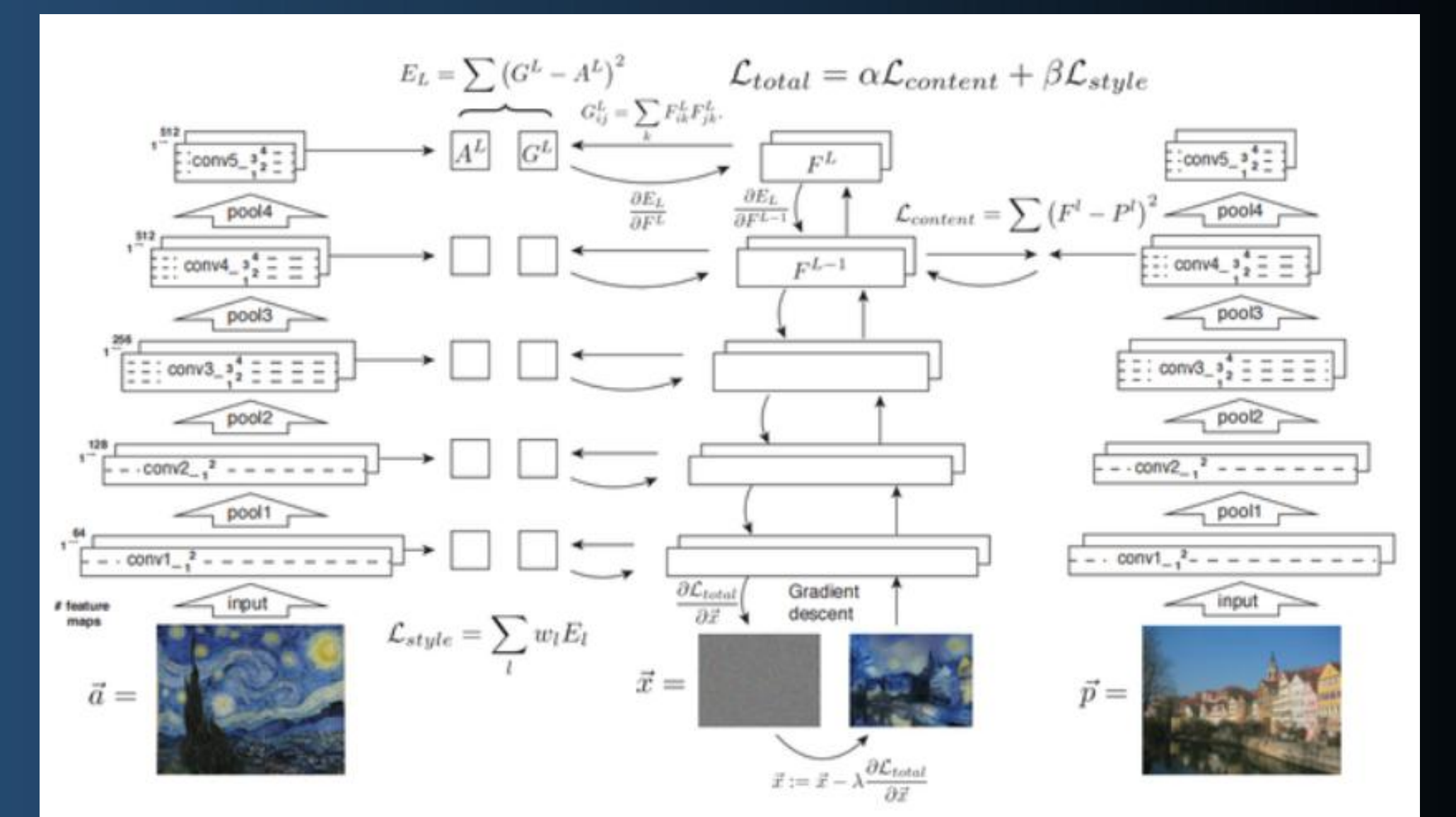
可以看出：

当 α 很小时，如左上角图，内容损失相对大，风格损失相对较小，所以生成的图片更凸显风格。

当 α 较大时，如右下角图，内容损失相对小，风格损失相对较大，生成的图片更凸显内容。

经实验测试采用的内容损失权重和风格损失权重为 0.025 和 5.0，能够得到一个较为令人满意的结果。

在损失函数公式中， α 和 β 为内容和风格的权重，而p, a和x分别是内容，风格和生成图片，其训练过程如下所示：



优化及结果

在本实验中，我们使用 L-BFGS 算法来优化总损失函数。由于我们使用的是梯度下降算法，我们引入一个Evaluator 类——通过两个独立的函数 loss 和 grads，来计算损失和梯度

对于合成图片，我们将其初始化为一个随机有效的像素的集合。最后通过 L-BFGS 算法来最小化损失函数。从实验结果可以看出，当迭代进行至 10 次后，损失就不再显著减少。

合成效果如下所示：

