

(19)中华人民共和国国家知识产权局



(12)发明专利申请

(10)申请公布号 CN 111127346 A

(43)申请公布日 2020.05.08

(21)申请号 201911246343.5

(22)申请日 2019.12.08

(71)申请人 复旦大学

地址 200433 上海市杨浦区邯郸路220号

(72)发明人 颜波 陈鹤丹

(74)专利代理机构 上海正旦专利代理有限公司

31200

代理人 陆飞 陆尤

(51)Int.Cl.

G06T 5/00(2006.01)

G06N 3/04(2006.01)

G06N 3/08(2006.01)

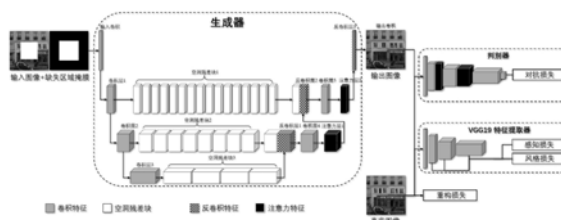
权利要求书2页 说明书5页 附图2页

(54)发明名称

基于部分到整体注意力机制的多层次图像修复方法

(57)摘要

本发明属于数字图像智能处理技术领域，具体为一种基于部分到整体注意力机制的多层次图像修复方法。图像修复是指利用算法来替换和生成丢失或缺损的图像数据。本发明方法包括：一种多层次的深度卷积神经网络生成器结构；一个部分到整体的注意力机制卷积层，分别集成到网络的生成器和判别器中。在网络的训练过程中引入了图像块判别器和重构损失、感知损失、风格损失、对抗损失四种损失函数来辅助生成器学习图像修复任务。实验结果表明，本发明可以生成细节逼真且整体结构合理的修复图像，有效解决图像修复问题。



CN 111127346 A

1. 一种基于部分到整体注意力机制的多层次图像修复方法,其特征在于,具体步骤为:

(1) 在U-Net结构的基础上构建多层次的深度卷积神经网络生成器,即去除U-Net结构的最后一个下采样层,并将其中原有的跳跃连接结构均替换为一系列空洞卷积组成的残差块网络结构;

(2) 将部分到整体的注意力机制卷积层分别集成到深度卷积神经网络生成器和判别器中;训练深度卷积神经网络生成器,得到图像修复模型;

(3) 利用图像修复模型,得到修复的图像。

2. 根据权利要求1所述的基于部分到整体注意力机制的多层次图像修复方法,其特征在于,所构建的多层次的深度卷积神经网络生成器中:

主干网络包括:1个输入特征提取层conv_in;3个下采样卷积层,分别记为conv_1, conv_2, conv_3;3个空洞残差块,分别记为drconv_1, drconv_2, drconv_3,分别连接在3个下采样卷积层conv_1, conv_2, conv_3之后;2个特征通道压缩层,记为conv_4, conv_5;2个注意力机制卷积特征层att_1和att_2,分别连接在2个特征通道压缩层conv_4, conv_5之后;1个输出卷积层conv_out;

所述判别器中,2个注意力层分别连接在第2、第3个卷积层之后;上述所有卷积层除输出卷积层conv_out以外均采用ReLU激活函数,输出卷积层采用Tanh激活函数。

3. 根据权利要求2所述的基于部分到整体注意力机制的多层次图像修复方法,其特征在于,步骤(2)中所述将部分到整体的注意力机制卷积层,分别集成到深度卷积神经网络生成器和判别器中,是将生成器或判别器网络某一层特征图M作为注意力机制卷积层的输入;具体是将整个特征图M和其中心占总大小25%左右的正方形区域m分别通过 1×1 卷积操作得到M_1与m_1,这里, 1×1 卷积输出通道数为输入通道数的八分之一;然后将其尺寸分别调整为 $N \times C/8$ 和 $C/8 \times n$,其中N、n为相应的高、宽,C/8为通道数,然后进行矩阵相乘,将结果按列进行softmax运算后得到注意力图A,注意力图A的尺寸为 $N \times n$;最后,将特征图M再次通过 1×1 卷积(输出通道数不变)的结果M_2的尺寸调整为 $C \times N$ 并与注意力图A进行矩阵相乘,得到特征图F,并特征图F的尺寸调整为 $C \times h \times w$,其中h、w分别为特征图的高和宽;最后,该卷积层的输出O为:

$$O = M - m + \alpha m + \beta F$$

其中, α 、 β 为可学习的参数,初始值分别为1和0。

4. 根据权利要求3所述的基于部分到整体注意力机制的多层次图像修复方法,其特征在于,生成器主干网络中,输入特征提取层conv_in的卷积核大小为7,步长为1,边缘填充大小为3,输入通道数为4,输出通道数为32;下采样卷积层conv_1, conv_2, conv_3的卷积核大小为4,步长为2,边缘填充大小为1,输入通道数分别为32、64、128,输出通道数分别为64、128、256;3个空洞残差块drconv_1, drconv_2, drconv_3中子块个数分别为16个、8个和4个,每个子块由卷积核大小为3,步长为1,边缘填充大小为2,扩张率为2,输出通道数维持不变的空洞卷积核和卷积核大小为3,步长为1,边缘填充大小为1,输出通道数维持不变的普通卷积核构成,每个子块的输出为输入经过上述卷积核得到的结果与输入的求和;2个特征通道压缩层conv_4, conv_5中,conv_4的卷积核大小为3,步长为1,边缘填充大小为1,输入通道数为256,输出通道数为128;conv_5的卷积核大小为5,步长为1,边缘填充大小为2,输入通道数为128,输出通道数为64;上采样反卷积层deconv_1, deconv_2, deconv_3的卷积核大

小为4,步长为2,边缘填充大小为1,输入通道数分别为256、128、64,输出通道数分别为128、64、32;输出卷积层conv_out的卷积核大小为7,步长为1,边缘填充大小为3,输入通道数为32,输出通道数为3。

5.根据权利要求4所述的基于部分到整体注意力机制的多层次图像修复方法,其特征在于,训练网络所采用的损失函数共4个,分别为重构损失 \mathcal{L}_{ℓ_1} 、对抗损失 \mathcal{L}_G 、感知损失 \mathcal{L}_{perc} 和风格损失 \mathcal{L}_{style} ;设 I_{input} 为输入网络的待修复图像以及缺失区域掩膜, I_{gt} 为对应的真实图像, $G(I_{input})$ 为网络输出的图像, $D(I_{input})$ 为判别器的输出(输出大于1表示判别器认为输入为真,小于-1表示判别器认为输入为假),则:

重构损失 \mathcal{L}_{ℓ_1} 定义如下:

$$\mathcal{L}_{\ell_1} = \|I_{gt} - G(I_{input})\|_1$$

对抗损失 \mathcal{L}_G 定义如下:

$$\mathcal{L}_G = -\mathbb{E}_{I_{input} \sim P_z(I_{input})}[D(I_{input})]$$

感知损失 \mathcal{L}_{perc} 定义如下:

$$\mathcal{L}_{perc} = \mathbb{E}\left[\sum_i \frac{1}{N_i} \|\phi_i(I_{gt}) - \phi_i(G(I_{input}))\|_1\right]$$

其中, ϕ_i 代表预训练的VGG-19网络的第i层的特征图,具体来说,指的是在ImageNet数据集上预训练的VGG-19网络中的relu1_1、relu2_1、relu3_1、relu4_1和relu5_1五个层中的特征图;

风格损失 \mathcal{L}_{style} 定义如下:

$$\mathcal{L}_{style} = \mathbb{E}\left[\sum_j \|G_j^\phi(I_{gt}) - G_j^\phi(G(I_{input}))\|_1\right]$$

其中, G_j^ϕ 是从特征图 ϕ_j 中构造出的大小为 $C_j \times C_j$ 的格拉姆矩阵,此处的特征图选取同上述感知损失中选择的五个特征图。

6.根据权利要求1-5之一所述的基于部分到整体注意力机制的多层次图像修复方法,其特征在于,网络模型的训练过程如下:

训练过程中总体损失函数为:

$$\mathcal{L} = \lambda_{\ell_1} \mathcal{L}_{\ell_1} + \lambda_G \mathcal{L}_G + \lambda_p \mathcal{L}_{perc} + \lambda_s \mathcal{L}_{style}$$

其中, λ_{ℓ_1} 、 λ_G 、 λ_p 、 λ_s 为损失函数的系数,分别设置为1、0.1、0.1、250;

对生成器和判别器均采用Adam优化器,其中生成器的学习率为0.0001,判别器的学习率为生成器学习率大小的十分之一;每次训练同步更新生成器和判别器的网络参数,直至损失函数收敛;然后,将生成器的学习率降低至0.00001,判别器的学习率仍为其十分之一,进一步进行训练直至网络输出达到理想的效果。

基于部分到整体注意力机制的多层次图像修复方法

技术领域

[0001] 本发明属于数字图像智能处理技术领域,涉及一种图像修复方法,更具体的说,涉及一种基于部分到整体注意力机制的多层次图像修复方法。

背景技术

[0002] 随着人们用来存储信息的介质的改变,图像修复这一任务的定义也在相应地发生变化,从文艺复兴时期对受损的壁画的修复,到早些年对老化的纸质照片的修复,再到目前对计算机上储存的数字图片的处理。需要强调的是,图像修复的应用已经不仅仅局限于“修复”这一功能,除此以外,现在的修复技术还可以应用在物体去除、去水印、去遮挡、人脸祛痘、磨皮等场景。

[0003] 早期的图像修复技术从其针对的待修复区域大小出发,主要分为两类:基于信息扩散的方法和基于图像块匹配的方法。基于信息扩散的方法采用迭代补全的思路,每一步迭代将图像的信息沿着等照度线方向向缺失区域内传播,从而将信息从缺失区域的边界逐步扩散到缺失区域内。这种方法只适用于处理细小的缺失区域,例如斑点、划痕等。基于图像块匹配的方法可以处理图像中规模较大的缺失区域,其核心思想是通过图像特征来进行图像块的配对和替换。然而,这种方法假设缺失区域的内容可以在背景区域内找到,但这并不总是成立的。除此以外,上述所有早期的传统修复技术都无法理解图像的结构等高层次语义信息,且无法生成新的、独特的图像内容。

[0004] 近年来,卷积神经网络(CNN)的发展为计算机视觉领域带来了新的突破,由此也产生了一系列基于深度学习技术的图像修复算法。这些算法克服了传统算法的不足,可以生成视觉上较为真实且整体结构一致的修复结果。特别地,由于图像修复是一种生成任务,所以可以很好地运用生成对抗网络(GenerativeAdversarialNetworks, GANs)^[1]。GAN是一种特殊的深度卷积神经网络,它包括一个生成器用于生成图像、一个可学习的判别器用于判断生成的图像是否是真实的。

[0005] 对目前已有的网络来说,其生成器多采用类似U-Net^[2]的网络结构,这种结构最早被用在医疗图像分割中,且取得了较好的效果。其中,跳跃连接是U-Net的关键部件,可以整合不同层次的图像信息。然而,Yu等人指出^[3],将跳跃连接用于中心大规模缺失区域的修复无法得到理想的效果,原因是中心缺失区域在跳跃连接中无法传递有价值的纹理、颜色等信息。

[0006] 此外,注意力机制的应用也使得图像修复网络可以对图像中的长距离相关性进行建模。Yu等人提出了一种上下文注意力模块(Contextual Attention, 简称CA)^[4],可以较好地利用背景区域中的已有信息进行修复。然而,这种方法过度关注对已有信息的利用,忽略了缺失区域内部的语义连续性,因此可能产生视觉上不合理的修复结果。

发明内容

[0007] 为了克服现有技术的不足,本发明的目的在于提供一种产生真实、自然的修复结

果的基于部分到整体注意力机制的多层次图像修复方法。

[0008] 本发明提供的基于部分到整体注意力机制的多层次图像修复方法,主要针对图像中心正方形缺失场景进行修复,其中缺失区域占整体图像大小的25%左右;本发明首先构建一个多层次的深度卷积神经网络生成器,然后将所提出的部分到整体注意力层分别集成到生成器和判别器中,同时引入重构损失、感知损失、风格损失、对抗损失四种损失函数来约束网络的训练;最后利用训练的多层次的深度卷积神经网络生成器,得到修复的图像。本发明提出的修复方法可以产生真实、自然的修复结果。

[0009] 本发明提供的基于部分到整体注意力机制的多层次图像修复方法,具体步骤包括:

[0010] (1) 在U-Net^[2]结构的基础上构建多层次的深度卷积神经网络生成器;

[0011] (2) 将部分到整体的注意力机制卷积层分别集成到深度卷积神经网络生成器和判别器中;训练深度卷积神经网络生成器,得到图像修复模型;

[0012] (3) 利用训练的深度卷积神经网络生成器(图像修复模型),得到修复的图像。

[0013] 本发明步骤(1)中,所述多层次的深度卷积神经网络生成器,是在U-Net结构的基础上构建,去除U-Net结构的最后一个下采样层后,将其中原有的跳跃连接结构均替换为一系列空洞卷积组成的残差块网络结构。

[0014] 本发明中,生成器主干网络包括:

[0015] 1个输入特征提取层conv_in,其卷积核大小为7,步长为1,边缘填充大小为3,输入通道数为4,输出通道数为32;

[0016] 3个下采样卷积层,分别记为conv_1,conv_2,conv_3,其卷积核大小为4,步长为2,边缘填充大小为1,输入通道数分别为32、64、128,输出通道数分别为64、128、256;

[0017] 3个空洞残差块,分别记为drconv_1,drconv_2,drconv_3,分别连接在3个下采样卷积层conv_1,conv_2,conv_3之后;其中子块个数分别为16个、8个和4个,每个子块由卷积核大小为3,步长为1,边缘填充大小为2,扩张率为2,输出通道数维持不变的空洞卷积核和卷积核大小为3,步长为1,边缘填充大小为1,输出通道数维持不变的普通卷积核构成,每个子块的输出为输入经过上述卷积核得到的结果与输入的求和;

[0018] 3个上采样反卷积层,分别记为deconv_1,deconv_2,deconv_3,其卷积核大小为4,步长为2,边缘填充大小为1,输入通道数分别为256、128、64,输出通道数分别为128、64、32;

[0019] 2个特征通道压缩层记为conv_4,conv_5,其中conv_4的卷积核大小为3,步长为1,边缘填充大小为1,输入通道数为256,输出通道数为128;conv_5的卷积核大小为5,步长为1,边缘填充大小为2,输入通道数为128,输出通道数为64;

[0020] 2个注意力机制卷积特征层att_1和att_2,分别连接在2个特征通道压缩层conv_4,conv_5之后;

[0021] 1个输出卷积层conv_out,其卷积核大小为7,步长为1,边缘填充大小为3,输入通道数为32,输出通道数为3;

[0022] 所述判别器中,2个注意力层分别连接在第2、第3个卷积层之后;

[0023] 上述所有卷积层除输出卷积层conv_out以外均采用ReLU激活函数,输出卷积层采用Tanh激活函数。

[0024] 本发明步骤(2)中,所述将部分到整体的注意力机制卷积层,分别集成到深度卷积

神经网络生成器和判别器中,是将生成器或判别器网络某一层特征图M作为注意力机制卷积层的输入;具体是将整个特征图M和其中心占总大小25%左右(一般为22%-28%)的正方形区域m分别通过 1×1 卷积(输出通道数为输入通道数的八分之一)操作得到M₁与m₁,然后将其尺寸分别调整为 $N \times C/8$ 和 $C/8 \times n$,其中N、n为相应的高、宽,C/8为通道数,然后进行矩阵相乘,将结果按列进行softmax运算后得到注意力图A,注意力图A的尺寸为 $N \times n$;最后,将特征图M再次通过 1×1 卷积(输出通道数不变)的结果M₂的尺寸调整为 $C \times N$ 并与注意力图A进行矩阵相乘,得到特征图F,并特征图将F的尺寸调整为 $C \times h \times w$,其中h、w分别为特征图的高和宽。最后,该卷积层的输出O为:

$$[0025] \quad O = M - m + \alpha m + \beta F$$

[0026] 其中, α 、 β 为可学习的参数,初始值分别为1和0。

[0027] 本发明中,网络训练所采用的损失函数共4个,分别为重构损失 \mathcal{L}_{ℓ_1} 、对抗损失 \mathcal{L}_G 、感知损失 \mathcal{L}_{perc} 和风格损失 \mathcal{L}_{style} 。设 I_{input} 为输入网络的待修复图像以及缺失区域掩膜, I_{gt} 为对应的真实图像, $G(I_{input})$ 为网络输出的图像, $D(I_{input})$ 为判别器的输出(输出大于1表示判别器认为输入为真,小于-1表示判别器认为输入为假),则:

[0028] 重构损失 \mathcal{L}_{ℓ_1} 定义如下:

$$[0029] \quad \mathcal{L}_{\ell_1} = \|I_{gt} - G(I_{input})\|_1$$

[0030] 对抗损失 \mathcal{L}_G 定义如下:

$$[0031] \quad \mathcal{L}_G = -\mathbb{E}_{I_{input} \sim P_Z(I_{input})}[D(I_{input})]$$

[0032] 感知损失 \mathcal{L}_{perc} 定义如下:

$$[0033] \quad \mathcal{L}_{perc} = \mathbb{E} \left[\sum_i \frac{1}{N_i} \|\phi_i(I_{gt}) - \phi_i(G(I_{input}))\|_1 \right]$$

[0034] 其中, ϕ_i 代表预训练的VGG-19网络的第i层的特征图,具体来说,指的是在ImageNet数据集上预训练的VGG-19网络中的relu1_1、relu2_1、relu3_1、relu4_1和relu5_1五个层中的特征图。

[0035] 风格损失 \mathcal{L}_{style} 定义如下:

$$[0036] \quad \mathcal{L}_{style} = \mathbb{E} \left[\sum_j \|G_j^\phi(I_{gt}) - G_j^\phi(G(I_{input}))\|_1 \right]$$

[0037] 其中, G_j^ϕ 是从特征图 ϕ_j 中构造出的大小为 $C_j \times C_j$ 的格拉姆矩阵,此处的特征图选取同上述感知损失中选择的五个特征图。

[0038] 训练过程中总体损失函数为:

$$[0039] \quad \mathcal{L} = \lambda_{\ell_1} \mathcal{L}_{\ell_1} + \lambda_G \mathcal{L}_G + \lambda_p \mathcal{L}_{perc} + \lambda_s \mathcal{L}_{style}$$

[0040] 其中, λ_{ℓ_1} 、 λ_G 、 λ_p 、 λ_s 为损失函数的系数,分别设置为1、0.1、0.1、250。

[0041] 进一步的,网络模型的训练方法如下:

[0042] 对生成器和判别器均采用Adam优化器,其中生成器的学习率为0.0001,判别器的学习率为生成器学习率大小的十分之一。每次训练同步更新生成器和判别器的网络参数,

直至损失函数收敛。然后,将生成器的学习率降低至0.00001,判别器的学习率仍为其十分之一,进一步进行训练直至网络输出达到理想的效果。

[0043] 本发明步骤(3)中,修复图像时,网络只需要使用生成器。

[0044] 本发明的有益效果在于:本发明设计了一个基于部分到整体注意力机制的多层次图像修复方法,通过端到端的生成器网络,对中心缺失的图像进行修复。该方法合理利用了输入图像不同层次的特征,能够生成比以往修复方法更加精细的修复结果;同时,由于采用了部分到整体的注意力机制,该方法可以从背景区域中合理选择所需要的信息用来进行修复,且不会产生修复结果不连贯不合理的问题。

附图说明

[0045] 图1为本发明的网络框架图。

[0046] 图2为本发明的部分到整体注意力机制层示意图。

[0047] 图3为本发明的修复效果图。

具体实施方式

[0048] 下面对本发明实施方案进行详细说明,但是本发明的保护范围不局限于所述实施例。

[0049] 采用图1所示的网络结构,用数据集Paris StreetView^[5](巴黎街景数据集,共包括14900张训练图片,100张测试图片)训练该网络,获得图像修复模型。

[0050] 具体步骤是:

[0051] (1) 训练前,随机初始化网络参数,并将训练集中的图像大小统一调整到610×350;

[0052] (2) 训练时,随机裁剪图像尺寸到256×256,并归一化到[0,1]区间内。使用Adam优化器进行模型参数更新,生成器的初始学习率为0.0001,判别器的学习率为生成器的十分之一。用小批量随机梯度下降的方法,最小化损失函数。批大小设为3。待网络收敛后,将生成器的学习率更新为0.00001,判别器的学习率更新为当前生成器学习率的十分之一,继续训练到网络重新收敛;

[0053] (3) 测试时,生成一个与图像大小相同(均为256×256)的单通道缺失区域掩膜,其中用1表示背景区域,用0表示缺失区域,将其与待修复的三通道RGB格式图像按通道拼接后送入生成器,得到输出结果。将输出结果的背景区域替换成原图的背景区域,即得到最终修复结果。

[0054] 本发明的图像修复定量评测结果,见表1。第一行为Contextual Attention^[4](简称CA)的评测结果,第二行为本发明的评测结果,其中L1值越小越好,PSNR、SSIM值越大越好。

[0055] 图3为本发明图像修复效果图,第1行为输入待修复图像,第2行为CA的修复结果,第3行为本发明的修复结果,第4行为真实图像。可以看出,本发明的修复效果更加真实,且修复区域的连续性更好。此外,从第5列图片可以看出,本发明的方法不会滥用背景区域的已知信息,生成结果更加合理。

[0056] 表1本发明的修复图像定量评测结果

[0057]		L1	PSNR	SSIM
	CA	3.231%	23.961	0.811
	Ours	2.988%	24.728	0.818

[0058]

[0059] 参考文献

[0060] [1]Goodfellow I,Pouget-Abadie J,Mirza M,et al.Generative adversarial nets[C]//Advances in Neural Information Processing Systems.2014:2672-2680.

[0061] [2]Ronneberger O,Fischer P,Brox T.U-net:Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention.Springer,Cham,2015:234-241.

[0062] [3]Yu J,Lin Z,Yang J,et al.Free-form image inpainting with gated convolution[C]//Proceedings of the IEEE International Conference on Computer Vision.2019:4471-4480.

[0063] [4]Yu J,Lin Z,Yang J,et al.Generative image inpainting with contextual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018:5505-5514.

[0064] [5]DoerschC,Singh S,Gupta A,et al.What makes Paris look like Paris?[J].ACM Transactions on Graphics,2012..

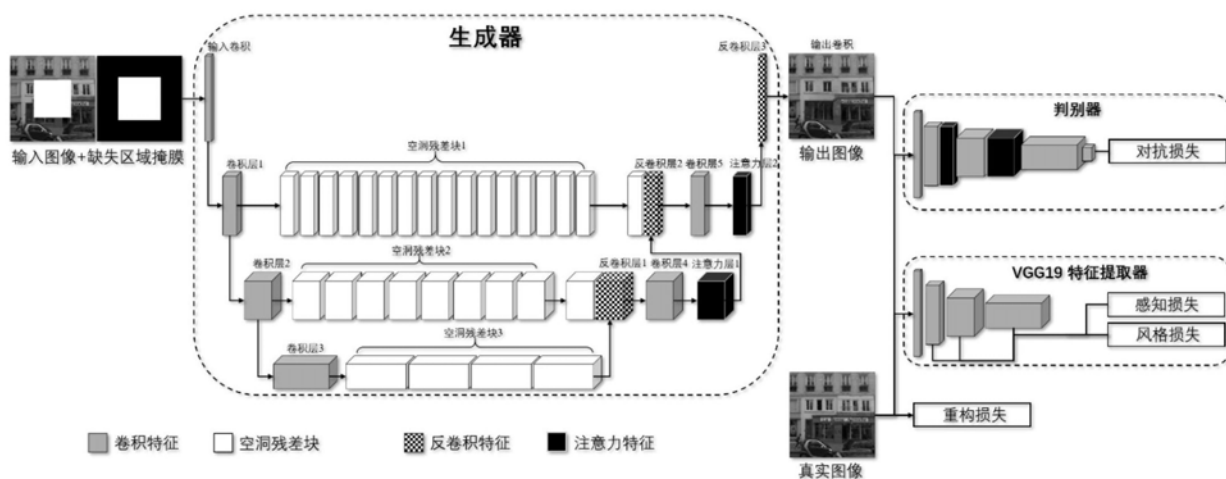


图1

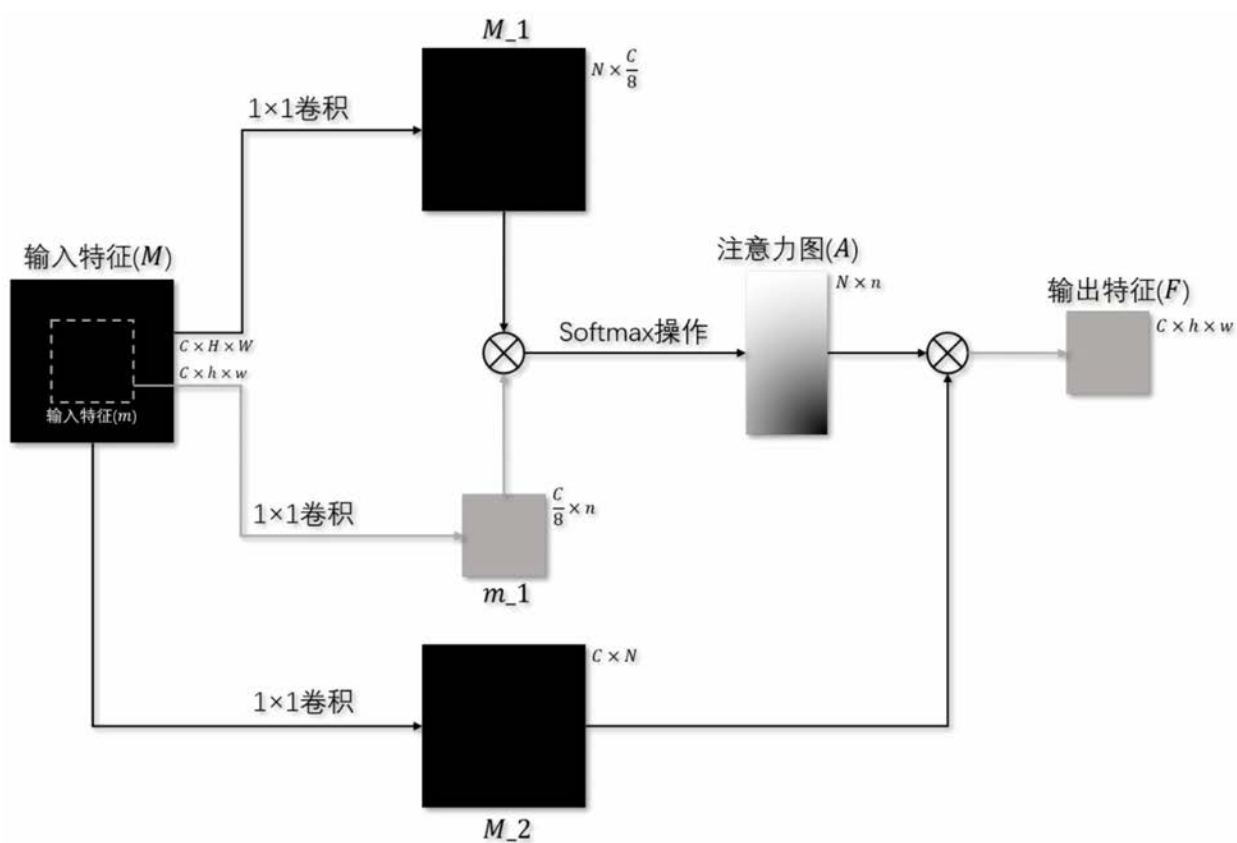


图2

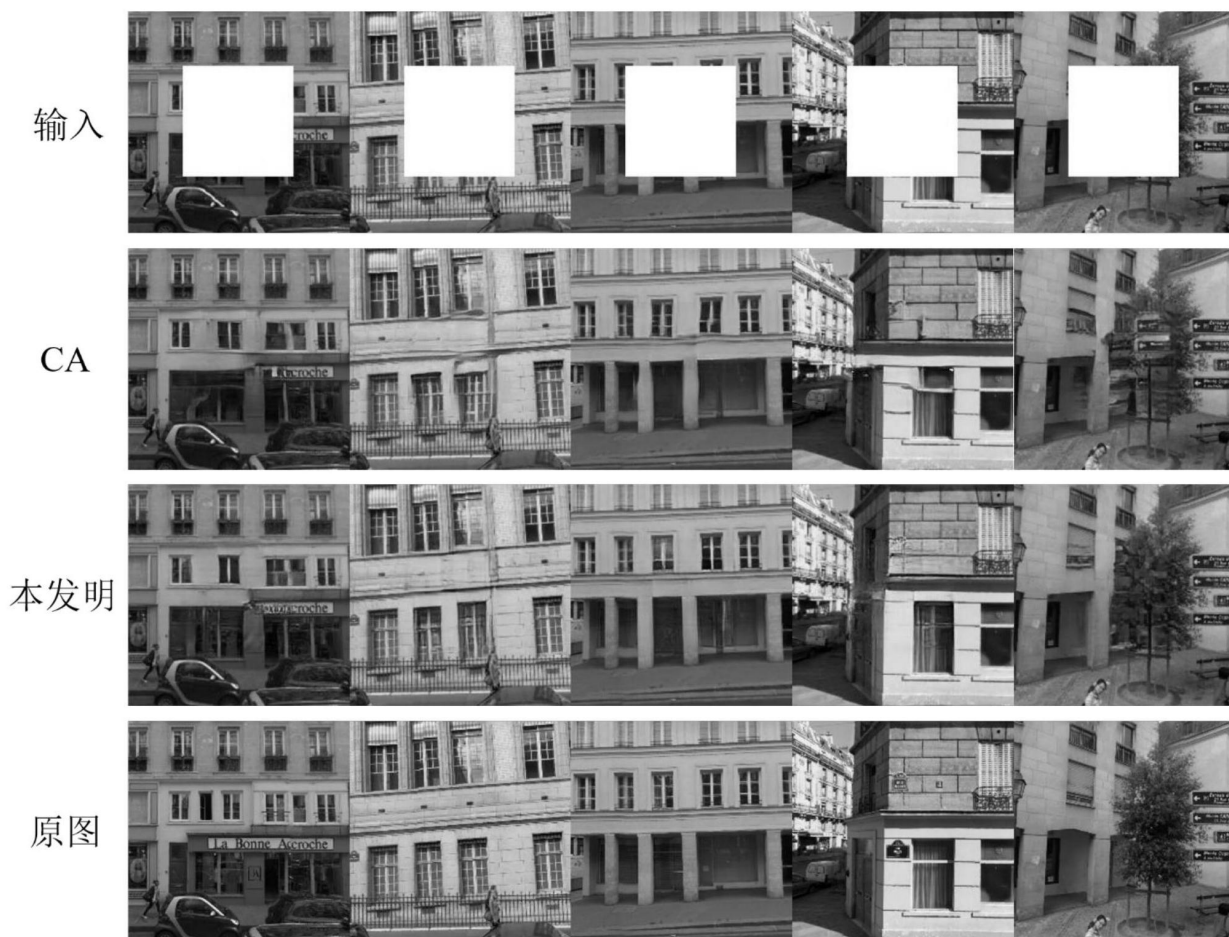


图3