



(12)发明专利申请

(10)申请公布号 CN 109583584 A

(43)申请公布日 2019.04.05

(21)申请号 201811353465.X

(22)申请日 2018.11.14

(71)申请人 中山大学

地址 510275 广东省广州市海珠区新港西路135号

(72)发明人 卢宇彤 瞿毅力 陈志广

(74)专利代理机构 湖南兆弘专利事务所(普通合伙) 43008

代理人 谭武艺

(51)Int.Cl.

G06N 3/08(2006.01)

G06N 3/04(2006.01)

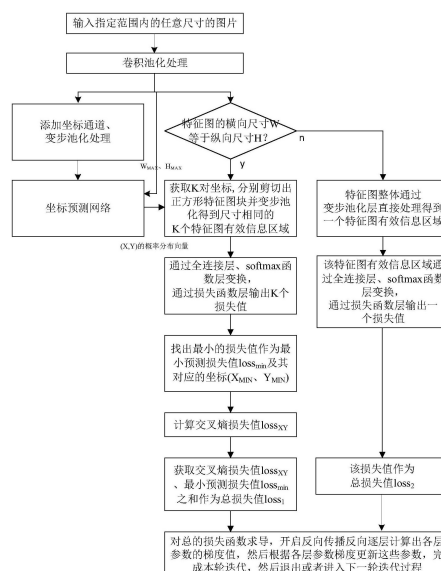
权利要求书3页 说明书10页 附图4页

(54)发明名称

可使具有全连接层的CNN接受不定形状输入的方法及系统

(57)摘要

本发明公开了一种可使具有全连接层的CNN接受不定形状输入的方法及系统,本发明通过在现有的具有全连接层的CNN网络的卷积池化后新增加一个包含在对特征图添加坐标通道后进行处理输出固定尺寸的新特征图的变步池化层以及一个由多个全连接层和一个softmax函数层构成的坐标预测网络,通过坐标预测网络确定包含有效像素区域的特征图块中心点坐标(X,Y)的概率分布向量,并根据特征图的横向尺寸W与纵向尺寸H的情况做不同处理,最终使网络能接受不定形状输入。本发明能够有效解决在具有全连接层的CNN中不能接受不定形状输入的缺陷,可使给定的只可接受固定形状输入的具有全连接层的CNN能接受不定形状输入,且其他要求不变,具有兼容性好的优点。



1. 一种可使具有全连接层的CNN接受不定形状输入的方法,其特征在于实施步骤包括:
 - 1) 输入指定范围内的任意尺寸的图片;
 - 2) 将图片通过卷积池化处理得到特征图;
 - 3) 针对步骤2)得到的特征图的横向尺寸W与纵向尺寸H,如果特征图的横向尺寸W与纵向尺寸H不相等则跳转执行下一步;否则,跳转执行步骤10);
 - 4) 对特征图添加坐标通道后进行变步池化处理输出固定尺寸的新特征图;
 - 5) 根据新特征图,通过多个全连接层和一个softmax函数层构成的坐标预测网络确定包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量,其中坐标预测网络的全连接层根据卷积池化处理得到特征图的最小的横向尺寸 W_{\min} 、最小的纵向尺寸 H_{\min} 和变步池化层输出的新特征图拉直后的长度进行设计;
 - 6) 根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量获取K对坐标,根据K对坐标分别剪切出正方形特征图块并通过变步池化变成尺寸相同的K个特征图有效信息区域,将K个特征图有效信息区域分别通过全连接层、softmax函数层变换得到K个概率分布向量,并通过损失函数层输出K个损失值,找出最小的损失值作为最小预测损失值 $loss_{\min}$ 及其对应的坐标(X_{\min} 、 Y_{\min});
 - 7) 将坐标(X_{\min} 、 Y_{\min})作为标签和包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 $loss_{xy}$;
 - 8) 获取交叉熵损失值 $loss_{xy}$ 、最小预测损失值 $loss_{\min}$ 之和作为总损失值 $loss_1$,且以总损失值 $loss_1$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代,然后退出或者进入下一轮迭代过程;
 - 9) 将特征图整体直接经过变步池化处理得到一个特征图有效信息区域,将该特征图有效信息区域通过全连接层、softmax函数层缩放为一个概率分布向量,并通过预设的损失函数计算输出一个损失值 $loss_2$;以该损失值 $loss_2$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代,然后退出或者进入下一轮迭代过程。
2. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法,其特征在于,步骤4)中对特征图添加坐标通道的具体步骤包括:在其通道方向添加特征图元素在纵横方向上的X、Y两个坐标通道信息,X、Y坐标均以特征图左上角为原点,从0开始取整数值,特征图当前所有通道中相同位置的元素坐标值相同,X、Y两个坐标通道依次堆叠到现有通道的后面。
3. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法,其特征在于,步骤4)中进行变步池化处理输出固定尺寸的新特征图的详细步骤包括:
 - 4.1) 设定对输入的特征图横向和纵向固定分区的数量,在通道方向不做处理,保留所有通道;
 - 4.2) 设定对每个分区进行池化的池化核的大小,核的尺寸可以大于、小于和等于分区的尺寸;
 - 4.3) 设定池化方式,可选的方式包括最大池化、平均池化和最小池化;
 - 4.4) 根据固定数量的分区,对每个分区按照设定池化方式进行池化;初始时池化核左上角与分区的左上角对齐,按先横向后纵向的顺序,逐行地对每个分区执行一次池化操作;

a) 当核的尺寸小于分区的尺寸时, 分区内未被池化核覆盖的区域则不作处理而被丢弃; b) 当核的尺寸等于分区时, 特征图的每个元素刚好均被池化核覆盖一遍; c) 当核的尺寸大于分区的尺寸时, 核超出的部分将覆盖到其他分区, 则池化范围扩大为池化核覆盖的范围; 如果池化核覆盖区域超出下边界, 则将池化核的下边界自动被调整为特征图的下边界; 如果池化核覆盖区域超出右边界, 则将池化核的右边界自动被调整为特征图的右边界;

4.5) 每个分区池化出的元素根据原来的分区位置进行拼接组装, 得到一个尺寸为设定的纵横方向上分区的数量的特征图。

4. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法, 其特征在于, 步骤5) 特征图的最的横向尺寸 W_{MAX} 、最大的纵向尺寸 H_{MAX} 的获取步骤包括: 针对图片数据集分别迭代输入任意形状的图片, 将图片通过卷积池化处理得到特征图, 并记录特征图的横向尺寸 W 与纵向尺寸 H , 最终计算或预估出特征图的所有横向尺寸 W 与纵向尺寸 H 中的最大的横向尺寸 W_{MAX} 和最大的纵向尺寸 H_{MAX} 。

5. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法, 其特征在于, 步骤5) 中多个全连接层和一个softmax函数层构成的坐标预测网络中, 第一层全连接层的输入长度均设定为变步池化层输出特征图拉直后的长度; 第二层全连接层与基础模型一致且以第一层全连接层的输出作为输入; 最后一层全连接层全连接层为两个并列的全连接层, 均以第二层全连接层输出作为输入, 两者输出长度分别设定为特征图的最的横向尺寸 W_{MAX} 、最大的纵向尺寸 H_{MAX} , 分别用来预测X坐标和Y坐标; softmax函数层位于最后一层全连接层之后, 用于将最后一层全连接的输出缩放为概率分布向量, 从而输出包含有效像素区域的特征图块的中心点坐标 (X, Y) 的概率分布向量。

6. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法, 其特征在于, 步骤6) 中根据包含有效像素区域的特征图块的中心点坐标 (X, Y) 的概率分布向量获取K对坐标的详细步骤包括: 根据包含有效像素区域的特征图块的中心点坐标 (X, Y) 的概率分布向量得到概率前K的K个X坐标和概率前K的K个Y坐标, 根据获取步骤2) 得到的特征图的横向尺寸 W 与纵向尺寸 H 以判断特征图的形状, 当特征图的横向尺寸 W 大于纵向尺寸 H 时, 将得到的概率前K的K个X坐标与概率最高的Y坐标组成K对坐标; 当特征图的横向尺寸 W 小于纵向尺寸 H 时, 得到的概率最高的X坐标与概率前K的K个Y坐标组成K对坐标。

7. 根据权利要求1所述的可使具有全连接层的CNN接受不定形状输入的方法, 其特征在于, 步骤6) 中剪切出正方形特征图块的详细步骤包括:

6.1) 针对每一对坐标, 计算其对应特征图的横向尺寸 W 与纵向尺寸 H ;

6.2) 将横向尺寸 W 和纵向尺寸 H 进行比较, 如果横向尺寸 W 大于纵向尺寸 H , 则以坐标 X_c 对应的纵线为裁剪中线截取 X_c 对应的纵线两边各 $H/2$ 宽的范围作为正方形特征图块; 否则如果横向尺寸 W 小于纵向尺寸 H , 则以坐标 Y_c 对应的横线为裁剪中线截取 Y_c 对应的横线两边各 $W/2$ 宽的范围作为正方形特征图块; 其中, X_c 、 Y_c 为预测的特征图块的中心点的坐标。

8. 根据权利要求7所述的可使具有全连接层的CNN接受不定形状输入的方法, 其特征在于, 步骤6.2) 中以坐标 X_c 对应的纵线为裁剪中线截取 X_c 对应的纵线两边各 $H/2$ 宽的范围具体是指: 当 $H/2 \geq X_c$ 时, 直接截取 $(0, 0)$ 、 $(0, H)$ 、 (H, H) 、 $(H, 0)$ 四个坐标点依次连线围成的正方形特征图块; 当 $H/2 \geq W - X_c$ 时, 我们直接截取 $(W - H, 0)$ 、 $(W - H, H)$ 、 (W, H) 、 $(W, 0)$ 四个坐标点依次连线围成的正方形特征图块; 当 $H/2 < X_c < W - H/2$ 时, 截取 $(X_c - H/2, 0)$ 、 $(X_c - H/2, H)$ 、 $(X_c + H/2,$

H)、 $(X_c+H/2, 0)$ 四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块的中心点的坐标, W 为特征图的横向尺寸, H 为特征图的纵向尺寸。

9. 根据权利要求7所述的可使具有全连接层的CNN接受不定形状输入的方法,其特征在于,步骤6.2)中以坐标 Y_c 对应的横线为裁剪中线截取 Y_c 对应的横线两边各 $W/2$ 宽的范围具体是指:当 $W/2 \geq Y_c$ 时,直接截取 $(0, 0)$ 、 $(0, W)$ 、 (W, W) 、 $(W, 0)$ 四个坐标点依次连线围成的正方形特征图块;当 $W/2 \geq H - Y_c$ 时,直接截取 $(H - W, 0)$ 、 $(H - W, W)$ 、 (H, W) 、 $(H, 0)$ 四个坐标点依次连线围成的正方形特征图块;当 $W/2 < Y_c < H - W/2$ 时,截取 $(Y_c - W/2, 0)$ 、 $(Y_c - W/2, W)$ 、 $(Y_c + W/2, W)$ 、 $(Y_c + W/2, 0)$ 四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块的中心点的坐标, W 为特征图的横向尺寸, H 为特征图的纵向尺寸。

10. 一种可使具有全连接层的CNN接受不定形状输入的系统,包括计算机设备,其特征在于:所述计算机程序被编程以执行权利要求1~9中任意一项所述可使具有全连接层的CNN接受不定形状输入的方法的步骤。

可使具有全连接层的CNN接受不定形状输入的方法及系统

技术领域

[0001] 本发明属于深度学习的模型设计领域,具体而言涉及一种可使给定的只可接受固定形状输入的具有全连接层的CNN能接受不定形状输入的方法及系统。

背景技术

[0002] 在实际生活中,我们所能接触的图片是千姿百态多种多样的,有的是长条形,有的是正方形,有的分辨率高,有的分辨率很低。这些数量庞大、种类繁多的图片促进了以卷积神经网络(convolution neural network,简称CNN)为基础的计算机视觉技术的发展。基于CNN的计算机视觉技术在图像分类、目标检测和许多其他识别任务,甚至是非识别任务中展现出了巨大的潜能,因此对CNN中的基本算子的发展就显得更为必要和有意义。

[0003] 当前许多的CNN的结构都由两部分组成,卷积部分和其后的全连接部分。卷积部分最基本的操作是卷积和池化。卷积通过一个参数可学习的卷积核采用窗口滑动方式对图片计算加权,得到一个变换后的特征图。特征图是一个三维矩阵,除了横向和纵向两个维度的尺寸外,还有第三个通道方向的维度,这个维度上是相互独立的特征矩阵,这些矩阵在该维度上堆叠成完整的特征图。池化是通过一个参数固定的池化核,采用窗口滑动方式对图片进行采样,得到一个较原图更为抽象的缩略图。卷积和池化组合可以使得网络能对不同尺度级别的特征进行学习,设置多个卷积核实现了多重特征的学习。卷积部分学习之后,历经了多次变换的神经元将被传送到全连接层,全连接层采用all-to-all方式将传进来的神经元经过多次变换得到一个固定长度的向量,再通过一个softmax函数将向量变换为一个概率分布向量。在图片分类任务中,概率分布向量中每个维度的值表示一个类别的概率,向量长度与类标个数一致;在图片定位任务中,概率分布向量中每个维度的值表示一个坐标的概率,向量长度与坐标个数一致。

[0004] 从前述CNN的作用原理可以看出,卷积部分网络参数数目与输入图片的形状无关,并不需要固定的图像尺寸,他可以接收任意形状的图片输入产生对应尺寸的特征图。而另一方面,全连接层的参数数目直接取决于输入神经元的个数和输出个数。神经元输出个数和类标个数一致,是定值。一个网络的参数数目是固定的才能采用迭代更新的方式学习。因此,CNN的全连接部分需要固定尺寸的输入。

[0005] 当前绝大部分CNN在训练和预测时都有一个使用限定,它们都需要输入的图像尺寸是固定的(比如 224×224)。从前面的分析可知,固定尺寸输入的问题来源于全连接层,也是网络的最后阶段。而全连接层在图片分类任务、检测任务、定位任务等领域依然有着广泛的应用和前景。因此,当前通用的解决方案是对图片进行裁剪或缩放变形。裁剪的过程中会损失掉许多像素,缩放变形会破坏图片内容的角度、比例等结构信息,造成几何失真。

[0006] 在当前计算机视觉领域,使用CNN对图片分类、物体检测、语义分割等,都使用ImageNet、cifar10等基准数据集,这些数据集中的图片基本都是生活中常见的物品,这些物品在这些图片在进行缩放形变后,基本不影响我们识别。比如一只猫,在图片被纵向压缩一半后,我们依然能识别出这是一只猫。绝大多数应用场景,就和识别猫一样,对形变不敏

感。

[0007] 然而近年来,随着计算机视觉领域的发展,越来越多的科学大数据采用CNN进行处理。计算机视觉领域应用场景对图片处理的要求也越来越细粒度,例如人脸识别。这些都说明,很多应用场景将是对形变敏感的,过去通过缩放形变方式对图片的预处理将不能满足需求。而对图片的裁剪在很多场景下都不适用。因此,在这些重叠的应用场景下,我们需要一种能接收并学习不定形状图片的CNN改进方法。

[0008] 2014年,SPPNet被提出,它的核心方法是一个叫做“空间金字塔池化(spatial pyramid pooling,简称SPP)”的池化策略,可以消除形状固定的限制,不管输入图像的尺寸或比例如何,SPPNet的网络结构能够产生固定大小的输出。由于这些优点,SPP可以帮助改进各类基于CNN的图像分类方法。在ImageNet大规模视觉识别任务挑战(ILSVRC)2014上的38个参赛组中,SPP方法在物体检测上排名第2,在物体分类上排名第3。但SPP在对具有几何标签的不定形状的几何形状图片数据集的分类任务中表现较差,这说明基于空间位置的采样本质上依然是一种压缩形变,并没有解决几何失真问题,同时也说明经过多层卷积和池化之后的特征图依然保留了输入图片中的几何相关性。

[0009] 在固定尺寸输入时,全连接层将多层卷积和池化后高度抽象的特征图拉直成一个神经元向量,经过多次变换得到概率分布向量。2012年dropout被提出来用于防止在CNN训练中产生过拟合。dropout指随机将全连接层一定比例的神经元置零,使其失效。dropout的有效说明了全连接层的分类不需要全部的特征图输入也能有效。因此,在不定形状输入时,在这些不同形状特征图中提取出一个有效信息区域,通过对SPP的改进可以实现所有有效信息区域的固定形状尺寸的输出,再用这些特征图的有效信息区域进行分类是可行的。2018年7月,优步公司提出通过添加坐标通道可以更好的处理图片定位任务。这与我们对特征图有效信息区域的提取任务十分契合。因此,如何实现使具有全连接层的CNN接受不定形状输入,已经成为一项亟待解决的关键技术问题。

发明内容

[0010] 本发明要解决的技术问题:针对现有技术的上述问题,提供一种可使具有全连接层的CNN接受不定形状输入的方法及系统,本发明能够有效解决在具有全连接层的CNN中不能接受不定形状输入的缺陷,可使给定的只可接受固定形状输入的具有全连接层的CNN能接受不定形状输入,且其他要求不变,具有兼容性好的优点。

[0011] 为了解决上述技术问题,本发明采用的技术方案为:

[0012] 一种可使具有全连接层的CNN接受不定形状输入的方法,实施步骤包括:

[0013] 1) 输入指定范围内的任意尺寸的图片;

[0014] 2) 将图片通过卷积池化处理得到特征图;

[0015] 3) 针对步骤2)得到的特征图的横向尺寸W与纵向尺寸H,如果特征图的横向尺寸W与纵向尺寸H不相等则跳转执行下一步;否则,跳转执行步骤10);

[0016] 4) 对特征图添加坐标通道后进行变步池化处理输出固定尺寸的新特征图;

[0017] 5) 根据新特征图,通过多个全连接层和一个softmax函数层构成的坐标预测网络确定包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量,其中坐标预测网络的全连接层根据卷积池化处理得到特征图的最小的横向尺寸 W_{\min} 、最小的纵向尺寸 H_{\min} 和

变步池化层输出的新特征图拉直后的长度进行设计;

[0018] 6) 根据包含有效像素区域的特征图块的中心点坐标 (X, Y) 的概率分布向量获取 K 对坐标, 根据 K 对坐标分别剪切出正方形特征图块并通过变步池化变成尺寸相同的 K 个特征图有效信息区域, 将 K 个特征图有效信息区域分别通过全连接层、softmax 函数层变换得到 K 个概率分布向量, 并通过损失函数层输出 K 个损失值, 找出最小的损失值作为最小预测损失值 $loss_{min}$ 及其对应的坐标 (X_{min}, Y_{min}) ;

[0019] 8) 将坐标 (X_{min}, Y_{min}) 作为标签和包含有效像素区域的特征图块的中心点坐标 (X, Y) 的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 $loss_{xy}$;

[0020] 9) 获取交叉熵损失值 $loss_{xy}$ 、最小预测损失值 $loss_{min}$ 之和作为总损失值 $loss_1$, 且以总损失值 $loss_1$ 对总的损失函数求导, 开启反向传播反向逐层计算出各层参数的梯度值, 然后根据各层参数梯度更新这些参数, 完成本轮迭代, 然后退出或者进入下一轮迭代过程;

[0021] 10) 中心点坐标特征图直接经过变步池化处理得到一个特征图有效信息区域, 将该特征图有效信息区域通过全连接层、softmax 函数层缩放为一个概率分布向量, 并通过预设的损失函数计算输出一个损失值 $loss_2$; 以该损失值 $loss_2$ 对总的损失函数求导, 开启反向传播反向逐层计算出各层参数的梯度值, 然后根据各层参数梯度更新这些参数, 完成本轮迭代, 然后退出或者进入下一轮迭代过程。

[0022] 步骤4) 中对特征图添加坐标通道的具体步骤包括: 在其通道方向添加特征图元素在横纵方向上的 X, Y 两个坐标通道信息, X, Y 坐标均以特征图左上角为原点, 从0开始取整数值, 特征图当前所有通道中相同位置的元素坐标值相同, X, Y 两个坐标通道依次堆叠到现有通道的后面。

[0023] 步骤4) 中进行变步池化处理输出固定尺寸的新特征图的详细步骤包括:

[0024] 4.1) 设定对输入的特征图横向和纵向固定分区数量, 在通道方向不做处理, 保留所有通道;

[0025] 4.2) 设定对每个分区进行池化的池化核的大小, 核的尺寸可以大于、小于和等于分区的尺寸;

[0026] 4.3) 设定池化方式, 可选的方式包括最大池化、平均池化和最小池化;

[0027] 4.4) 根据固定数量的分区, 对每个分区按照设定池化方式进行池化; 初始时池化核左上角与分区的左上角对齐, 按先横向后纵向的顺序, 逐行地对每个分区执行一次池化操作: a) 当核的尺寸小于分区的尺寸时, 分区内未被池化核覆盖的区域则不作处理而被丢弃; b) 当核的尺寸等于分区时, 特征图的每个元素刚好均被池化核覆盖一遍; c) 当核的尺寸大于分区的尺寸时, 核超出的部分将覆盖到其他分区, 则池化范围扩大为池化核覆盖的范围; 如果池化核覆盖区域超出下边界, 则将池化核的下边界自动被调整为特征图的下边界; 如果池化核覆盖区域超出右边界, 则将池化核的右边界自动被调整为特征图的右边界;

[0028] 4.5) 每个分区池化出的元素根据原来的分区位置进行拼接组装, 得到一个尺寸为设定的横纵方向上分区数量的特征图。

[0029] 步骤5) 特征图的最大横向尺寸 W_{max} 、最大纵向尺寸 H_{max} 的获取步骤包括: 针对图片数据集分别迭代输入任意形状的图片, 将图片通过卷积池化处理得到特征图, 并记录特征图的横向尺寸 W 与纵向尺寸 H , 最终计算或预估出特征图的所有横向尺寸 W 与纵向尺寸 H 中的最大的横向尺寸 W_{max} 和最大的纵向尺寸 H_{max} 。

[0030] 步骤5)的坐标预测网络的多个全连接层和一个softmax函数层中,第一层全连接层的输入长度均设定为变步池化层输出特征图拉直后的长度;第二层全连接层与基础模型一致且以第一层全连接层的输出作为输入;最后一层全连接层全连接层为两个并列的全连接层,均以第二层全连接层输出作为输入,两者输出长度分别设定为特征图的最长的横向尺寸 W_{MAX} 、最大的纵向尺寸 H_{MAX} ,分别用来预测X坐标和Y坐标;softmax函数层位于最后一层全连接层之后,用于将最后一层全连接的输出缩放为概率分布向量,从而输出包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量。

[0031] 步骤6)中根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量获取K对坐标的详细步骤包括:根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量得到概率前K的K个X坐标和概率前K的K个Y坐标,根据获取步骤2)得到的特征图的横向尺寸W与纵向尺寸H以判断特征图的形状,当特征图的横向尺寸W大于纵向尺寸H时,将得到的概率前K的K个X坐标与概率最高的Y坐标组成K对坐标;当特征图的横向尺寸W小于纵向尺寸H时,得到的概率最高的X坐标与概率前K的K个Y坐标组成K对坐标。

[0032] 步骤6)中剪切出正方形特征图块的详细步骤包括:

[0033] 6.1)针对每一对坐标,计算其对应特征图的横向尺寸W与纵向尺寸H;

[0034] 6.2)将横向尺寸W和纵向尺寸H进行比较,如果横向尺寸W大于纵向尺寸H,则以坐标 X_c 对应的纵线为裁剪中线截取 X_c 对应的纵线两边各 $H/2$ 宽的范围作为正方形特征图块;否则如果横向尺寸W小于纵向尺寸H,则以坐标 Y_c 对应的横线为裁剪中线截取 Y_c 对应的横线两边各 $W/2$ 宽的范围作为正方形特征图块;其中, X_c 、 Y_c 为有效信息区域中心点的坐标。

[0035] 步骤6.2)中以坐标 X_c 对应的纵线为裁剪中线截取 X_c 对应的纵线两边各 $H/2$ 宽的范围具体是指:当 $H/2 \geq X_c$ 时,直接截取(0,0)、(0,H)、(H,H)、(H,0)四个坐标点依次连线围成的正方形特征图块;当 $H/2 \geq W - X_c$ 时,我们直接截取($W-H$,0)、($W-H$,H)、(W ,H)、(W ,0)四个坐标点依次连线围成的正方形特征图块;当 $H/2 < X_c < W-H/2$ 时,截取($X_c-H/2$,0)、($X_c-H/2$,H)、($X_c+H/2$,H)、($X_c+H/2$,0)四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块中心点的坐标,W为特征图的横向尺寸,H为特征图的纵向尺寸。

[0036] 步骤6.2)中以坐标 Y_c 对应的横线为裁剪中线截取 Y_c 对应的横线两边各 $W/2$ 宽的范围具体是指:当 $W/2 \geq Y_c$ 时,直接截取(0,0)、(0,W)、(W,W)、(W,0)四个坐标点依次连线围成的正方形特征图块;当 $W/2 \geq H - Y_c$ 时,直接截取($H-W$,0)、($H-W$,W)、(H ,W)、(H ,0)四个坐标点依次连线围成的正方形特征图块;当 $W/2 < Y_c < H-W/2$ 时,截取($Y_c-W/2$,0)、($Y_c-W/2$,W)、($Y_c+W/2$,W)、($Y_c+W/2$,0)四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块中心点的坐标,W为特征图的横向尺寸,H为特征图的纵向尺寸。

[0037] 本发明还提供一种可使具有全连接层的CNN接受不定形状输入的系统,包括计算机设备,所述计算机程序被编程以执行本发明所述可使具有全连接层的CNN接受不定形状输入的方法的步骤。

[0038] 和现有技术相比,本发明具有下述优点:本发明通过在现有CNN网络的卷积池化后新增加一个包含在对特征图添加坐标通道后进行处理输出固定尺寸的新特征图的变步池化层以及一个由多个全连接层和一个softmax函数层构成的坐标预测网络,通过坐标预测网络确定包含有效像素区域的特征图块中心点坐标(X,Y)的概率分布向量,并根据特征图的横向尺寸W与纵向尺寸H的情况做不同处理:

[0039] 在特征图的横向尺寸 W 与纵向尺寸 H 不同时,根据包含有效像素区域的特征图块的中心点坐标 (X,Y) 的概率分布向量获取 K 对坐标,根据 K 对坐标分别剪切出正方形特征图块并通过变步池化变成尺寸相同的 K 个特征图有效信息区域,将 K 个特征图有效信息区域分别通过全连接层、softmax函数层变换得到 K 个概率分布向量,并通过损失函数层输出 K 个损失值,找出最小的损失值作为最小预测损失值 $loss_{min}$ 及其对应的坐标 (X_{MIN},Y_{MIN}) ;将坐标 (X_{MIN},Y_{MIN}) 作为标签和有效像素区域的中心点坐标 (X,Y) 的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 $loss_{XY}$;获取交叉熵损失值 $loss_{XY}$ 、最小预测损失值 $loss_{min}$ 之和作为总损失值 $loss_1$,且以总损失值 $loss_1$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代;

[0040] 在特征图的横向尺寸 W 与纵向尺寸 H 相同时,中心点坐标特征图整体直接经过变步池化处理得到一个特征图有效信息区域,将该特征图有效信息区域通过变步池化层、全连接层、softmax函数层缩放为一个概率分布向量,并通过预设的损失函数计算输出一个损失值 $loss_2$;以该损失值 $loss_2$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代。

[0041] 通过上述方式,本发明能够有效解决在具有全连接层的CNN中不能接受不定形状输入的缺陷,可使给定的只可接受固定形状输入的具有全连接层的CNN能接受不定形状输入,且其他要求不变,具有兼容性好的优点。

附图说明

[0042] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

[0043] 图1为本发明实施例方法的基本流程示意图。

[0044] 图2为本发明实施例中CNN的分层结构示意图。

[0045] 图3为本发明实施例中变步池化的原理示意图。

[0046] 图4为本发明实施例中特征图添加坐标通道再进行变步池化处理的示意图

[0047] 图5为本发明实施例中 $W>H$ 时截取正方形特征图块的示意图。

[0048] 图6为本发明实施例中 $W<H$ 时截取正方形特征图块的示意图。

具体实施方式

[0049] 下文将以AlexNet为例,对本发明可使具有全连接层的CNN接受不定形状输入的方法及系统进行进一步的详细说明,AlexNet是一个具有全连接层的经典CNN模型,在图片分类等任务中表现优秀,但它不能接受不同形状的输入,本发明可使具有全连接层的CNN接受不定形状输入的方法及系统将使得AlexNet接受不定形状输入。

[0050] 如图1所示,本实施例可使具有全连接层的CNN接受不定形状输入的方法实施步骤包括:

[0051] 1) 输入指定范围内的任意尺寸的图片;

[0052] 2) 将图片通过卷积池化处理得到特征图;

[0053] 3) 针对步骤2)得到的特征图的横向尺寸 W 与纵向尺寸 H ,如果特征图的横向尺寸 W

与纵向尺寸H不相等则跳转执行下一步;否则,跳转执行步骤10);

[0054] 4) 对特征图添加坐标通道后进行变步池化处理输出固定尺寸的新特征图;

[0055] 5) 根据新特征图,通过以特征图的最小的横向尺寸 W_{MAX} 、最小的纵向尺寸 H_{MAX} 设计的多个全连接层和一个softmax函数层构成的坐标预测网络确定包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量;

[0056] 6) 根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量获取K对坐标,根据K对坐标分别剪切出正方形特征图块并通过变步池化变成尺寸相同的K个特征图有效信息区域,将K个特征图有效信息区域分别通过全连接层、softmax函数层变换得到K个概率分布向量,并通过损失函数层输出K个损失值,找出最小的损失值作为最小预测损失值 $loss_{min}$ 及其对应的坐标(X_{MIN} 、 Y_{MIN});

[0057] 8) 将坐标(X_{MIN} 、 Y_{MIN})作为标签和包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 $loss_{XY}$;

[0058] 9) 获取交叉熵损失值 $loss_{XY}$ 、最小预测损失值 $loss_{min}$ 之和作为总损失值 $loss_1$,且以总损失值 $loss_1$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代,然后退出或者进入下一轮迭代过程;

[0059] 10) 特征图直接经过变步池化处理得到一个特征图有效信息区域,将该特征图有效信息区域通过全连接层、softmax函数层缩放为一个概率分布向量,并通过预设的损失函数计算输出一个损失值 $loss_2$;以该损失值 $loss_2$ 对总的损失函数求导,开启反向传播反向逐层计算出各层参数的梯度值,然后根据各层参数梯度更新这些参数,完成本轮迭代,然后退出或者进入下一轮迭代过程。

[0060] 如图2所示,其中虚线图框中的内容为本实施例可使具有全连接层的CNN接受不定形状输入的方法改造后的新增部分,其中现有部分如下:

[0061] 图片输入层,用于输入指定范围内的任意尺寸的图片。

[0062] 卷积池化部分,包括多个卷积层和池化层,用于将图片通过卷积池化处理得到特征图。

[0063] 卷积池化部分的原始后续层部分,包括三个全连接层、softmax函数层、损失函数层,用于将特征图有效信息区域分别通过全连接层、softmax函数层变换得到概率分布向量,并通过损失函数层输出损失值。

[0064] 新增部分如下:

[0065] 添加坐标层,用于对卷积池化处理得到特征图补充横纵坐标通道信息。

[0066] 变步池化层#1,用于对添加坐标通道后的特征图进行变步池化处理输出固定尺寸的新特征图。

[0067] 坐标预测网络,包括多个全连接层和一个softmax函数层,用于根据新特征图确定包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量,其中全连接层根据特征图的最小的横向尺寸 W_{MAX} 、最小的纵向尺寸 H_{MAX} 和变步池化层#1输出的新特征图拉直后的长度进行设计;

[0068] 提取特征图块,用于坐标预测网络预测中心点坐标的概率分布向量获取K对坐标,根据K对坐标分别剪切出正方形特征图块;

[0069] 变步池化层#2,用于将提取特征图块输出的K个特征图块通过变步池化变成尺寸

相同的K个特征图有效信息区域；

[0070] 坐标交叉熵损失函数层,用于将坐标(X_{\min} 、 Y_{\min})作为标签和坐标预测网络预测中心点坐标的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 loss_{XY} ；

[0071] 总损失函数层,用于获取交叉熵损失值 loss_{XY} 、最小预测损失值 loss_{\min} 之和作为总损失值 loss_1 。

[0072] 本实施例中,步骤3)中对特征图添加坐标通道的具体步骤包括:在其通道方向添加特征图元素在横纵方向上的X、Y两个坐标通道信息,X、Y坐标均以特征图左上角为原点,从0开始取整数值,特征图当前所有通道中相同位置的元素坐标值相同,X、Y两个坐标通道依次堆叠到现有通道的后面。坐标通道的添加根据使用场景是可选的。

[0073] 现有的SPP的池化策略通过将图片划分成固定数量的分区,对每个分区最大池化,再将池化结果拼接成向量的方式,实现了接收不定输入和产生固定输出,但是该分区方法破坏了分区与分区之间的关联性,为此需要进行多种不同尺度的分区再拼接这些输出。本实施例中的变步池化方法采用和SPP一样的分区方法,需要固定横纵方向上分区的次数外,同时还需要确定核的大小。在核大小大于分区子图大小时,其效果和普通池化一样,在核大小等于分区子图大小时,则与SPP一样。变步池化将对分区子图的处理结果按原位拼接,得到可继续采用卷积或池化学习的特征图。

[0074] 本实施例中,步骤3)中进行变步池化处理输出固定尺寸的新特征图的详细步骤包括:

[0075] 3.1) 设定对输入的特征图横向和纵向固定分区的数量,在通道方向不做处理,保留所有通道;

[0076] 3.2) 设定对每个分区进行池化的池化核的大小,核的尺寸可以大于、小于和等于分区的尺寸;

[0077] 3.3) 设定池化方式,可选的方式包括最大池化、平均池化和最小池化;

[0078] 3.4) 根据固定数量的分区,对每个分区按照设定池化方式进行池化;初始时池化核左上角与分区的左上角对齐,按先横向后纵向的顺序,逐行地对每个分区执行一次池化操作:

[0079] a) 当核的尺寸小于分区的尺寸时,分区内未被池化核覆盖的区域则不作处理而被丢弃;

[0080] b) 当核的尺寸等于分区时,特征图的每个元素刚好均被池化核覆盖一遍;当核的尺寸等于分区的尺寸且池化方式为最大池化时,算子处理效果与SPP和某些设置下普通池化算子的效果一致;

[0081] c) 当核的尺寸大于分区的尺寸时,核超出的部分将覆盖到其他分区,则池化范围扩大为池化核覆盖的范围;如果池化核覆盖区域超出下边界,则将池化核的下边界自动被调整为特征图的下边界;如果池化核覆盖区域超出右边界,则将池化核的右边界自动被调整为特征图的右边界;这样的池化效果与通常设置下普通池化算子的效果一致。

[0082] 图3(a)~(d)分别为本实施例中采用的变步池化过程中,池化核尺寸小于分区尺寸的情况、池化核尺寸等于分区尺寸的情况、池化核尺寸大于分区尺寸的情况、池化核尺寸大于分区尺寸时下边界和右边界的情况;

[0083] 3.5) 每个分区池化出的元素根据原来的分区位置进行拼接组装,得到一个尺寸为

设定的纵横方向上分区的数量的特征图。

[0084] 整个池化过程,可以看作是步数固定但步长随着特征图尺寸变化而变化的池化过程,因此上述池化操作过程称为变步池化。特征图添加坐标通道再进行变步池化处理的原理如图4所示,通过对特征图添加坐标通道后进行变步池化处理,最终输出固定尺寸的新特征图。

[0085] 本实施例中,步骤4)特征图的最长的横向尺寸 W_{MAX} 、最长的纵向尺寸 H_{MAX} 的获取步骤包括:针对图片数据集分别迭代输入任意形状的图片,将图片通过卷积池化处理得到特征图,并记录特征图的横向尺寸 W 与纵向尺寸 H ,最终计算或预估出特征图的所有横向尺寸 W 与纵向尺寸 H 中的最长的横向尺寸 W_{MAX} 和最长的纵向尺寸 H_{MAX} 。在完成坐标通道的添加之后,然后本实施例将得到的不同形状的特征图进行变步池化处理,得到相同尺寸的新特征图。此时,可选择性地对该新特征图进行卷积池化学习,然后将最后得到的新特征图拉直成神经元向量,再将该神经元向量输入包含两个输出层的多层全连接网络,分别得到一个长度与 W_{MAX} 相同的向量和一个长度与 H_{MAX} 相同的向量,再用softmax函数将两个向量变换为X坐标的概率分布向量和Y坐标的概率分布向量。

[0086] 参见图2,步骤4)的坐标预测网络的多个全连接层和一个softmax函数层中,第一层全连接层的输入长度均设定为变步池化层输出的新特征图拉直后的长度;第二层全连接层与基础模型一致且以第一层全连接层的输出作为输入(与AlexNet一致,输入输出长度均为4096);最后一层全连接层全连接层为两个并列的全连接层,均以第二层全连接层输出作为输入,两者输出长度分别设定为特征图的最长的横向尺寸 W_{MAX} 、最长的纵向尺寸 H_{MAX} ,分别用来预测X坐标和Y坐标;softmax函数层位于最后一层全连接层之后,用于将最后一层全连接的输出缩放为概率分布向量,从而输出包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量。

[0087] 有效信息区域提取时,接收到一对预测坐标 X_c 、 Y_c 和一个特征图。首先,我们设定变步池化的纵横方向的分区数量(也就是输出尺寸)和池化核尺寸,此处分区横向和纵向的尺寸要求相同、池化核横向和纵向的尺寸要求相同。然后获取特征图的横向尺寸 W 与纵向尺寸 H 以判断特征图的形状。最后根据特征图形状选用坐标,再根据坐标对特征图进行裁剪,得到包含有效像素区域的特征图块,最后将特征图块根据设置进行变步池化处理,此时变步池化处理前无需添加坐标通道。

[0088] 本实施例中,步骤6)中根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量获取K对坐标的详细步骤包括:根据包含有效像素区域的特征图块的中心点坐标(X,Y)的概率分布向量得到概率前K的K个X坐标和概率前K的K个Y坐标,根据获取步骤2)得到的特征图的横向尺寸 W 与纵向尺寸 H 以判断特征图的形状,当特征图的横向尺寸 W 大于纵向尺寸 H 时,将得到的概率前K的K个X坐标与概率最高的Y坐标组成K对坐标;当特征图的横向尺寸 W 小于纵向尺寸 H 时,得到的概率最高的X坐标与概率前K的K个Y坐标组成K对坐标。

[0089] 本实施例中,步骤6)中剪切出正方形特征图块的详细步骤包括:

[0090] 6.1) 针对每一对坐标,计算其对应特征图的横向尺寸 W 与纵向尺寸 H ;

[0091] 6.2) 将横向尺寸 W 和纵向尺寸 H 进行比较,如果横向尺寸 W 大于纵向尺寸 H ,则以坐标 X_c 对应的纵线为裁剪中线截取 X_c 对应的纵线两边各 $H/2$ 宽的范围作为正方形特征图块(得

到一个横向纵向尺寸均为H的特征图块),如图5所示;否则如果横向尺寸W小于纵向尺寸H,则以坐标 Y_c 对应的横线为裁剪中线裁取 Y_c 对应的横线两边各 $W/2$ 宽的范围作为正方形特征图块(得到一个横向纵向尺寸均为W的特征图块),如图6所示;其中, X_c 、 Y_c 为预测的特征图块中心点的坐标。

[0092] 本实施例中,步骤6.2)中以坐标 X_c 对应的纵线为裁剪中线裁取 X_c 对应的纵线两边各 $H/2$ 宽的范围具体是指:当 $H/2 \geq X_c$ 时,直接裁取 $(0,0)$ 、 $(0,H)$ 、 (H,H) 、 $(H,0)$ 四个坐标点依次连线围成的正方形特征图块;当 $H/2 \geq W - X_c$ 时,我们直接裁取 $(W-H,0)$ 、 $(W-H,H)$ 、 (W,H) 、 $(W,0)$ 四个坐标点依次连线围成的正方形特征图块;当 $H/2 < X_c < W - H/2$ 时,裁取 $(X_c - H/2, 0)$ 、 $(X_c - H/2, H)$ 、 $(X_c + H/2, H)$ 、 $(X_c + H/2, 0)$ 四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块中心点的坐标,W为特征图的横向尺寸,H为特征图的纵向尺寸。

[0093] 本实施例中,步骤6.2)中以坐标 Y_c 对应的横线为裁剪中线裁取 Y_c 对应的横线两边各 $W/2$ 宽的范围具体是指:当 $W/2 \geq Y_c$ 时,直接裁取 $(0,0)$ 、 $(0,W)$ 、 (W,W) 、 $(W,0)$ 四个坐标点依次连线围成的正方形特征图块;当 $W/2 \geq H - Y_c$ 时,直接裁取 $(H-W,0)$ 、 $(H-W,W)$ 、 (H,W) 、 $(H,0)$ 四个坐标点依次连线围成的正方形特征图块;当 $W/2 < Y_c < H - W/2$ 时,裁取 $(Y_c - W/2, 0)$ 、 $(Y_c - W/2, W)$ 、 $(Y_c + W/2, W)$ 、 $(Y_c + W/2, 0)$ 四个坐标点依次连线围成的正方形特征图块;其中, X_c 、 Y_c 为预测的特征图块中心点的坐标,W为特征图的横向尺寸,H为特征图的纵向尺寸。

[0094] 本实施例中,步骤8)将坐标 (X_{MIN}, Y_{MIN}) 作为标签,根据包含有效像素区域特征图块的中心点坐标 (X, Y) 的概率分布向量计算两个坐标的交叉熵并求和得到交叉熵损失值 $loss_{XY}$ 时,交叉熵损失值 $loss_{XY}$ 的函数表达式如式(1)所示;

$$[0095] \quad loss_{XY} = -\left(\sum_{i=1}^{W_{MAX}} q_X(X_i) \log(p_X(X_i)) + \sum_{j=1}^{H_{MAX}} q_Y(Y_j) \log(p_Y(Y_j))\right) \quad (1)$$

[0096] 式(1)中, W_{MAX} 为最大的横向尺寸, H_{MAX} 为最大的纵向尺寸, $p_X(X_i)$ 是指预测得到X坐标的概率分布向量中对应坐标为 X_i 的概率值; $q_X(X_i)$ 是指将X坐标的真实标签通过独热编码(one-hot编码)转换得到概率分布向量中对应坐标为 X_i 的概率值; $p_Y(Y_j)$ 是指预测得到Y坐标的概率分布向量中对应坐标为 Y_j 的概率值, $q_Y(Y_j)$ 是指将Y坐标的真实标签通过独热编码转换得到概率分布向量中对应坐标为 Y_j 的概率值。

[0097] 在针对一个数据集进行训练时,卷积池化部分的第一层可以接收一张不定形状的图片。如此往复,直到整个数据集训练完成。预测时,如果输入为正方形,那么AlexNet的全连接层输出唯一一个概率分布向量即为预测结果;否则,AlexNet的全连接层输出的概率分布向量中,与坐标预测网络中概率最大的横纵坐标对应的向量即为预测结果。AlexNet经过本实施例前述可使具有全连接层的CNN接受不定形状输入的方法上述的改进之后即可实现对不定形状输入的训练和预测,整个改进在模型设计阶段完成,模型训练和预测时,除去不同迭代批次可以接受的输入形状不同外,其他要求不变,包括一个迭代批次内的输入尺寸一致的限制。此外,本实施例还提供一种可使具有全连接层的CNN接受不定形状输入的系统,包括计算机设备,该计算机程序被编程以执行本实施例前述可使具有全连接层的CNN接受不定形状输入的方法的步骤。

[0098] 以上所述仅是本发明的优选实施方式,本发明的保护范围并不局限于上述实施例,凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理前提下的若干改进和润饰,这些改进和润饰也

应视为本发明的保护范围。

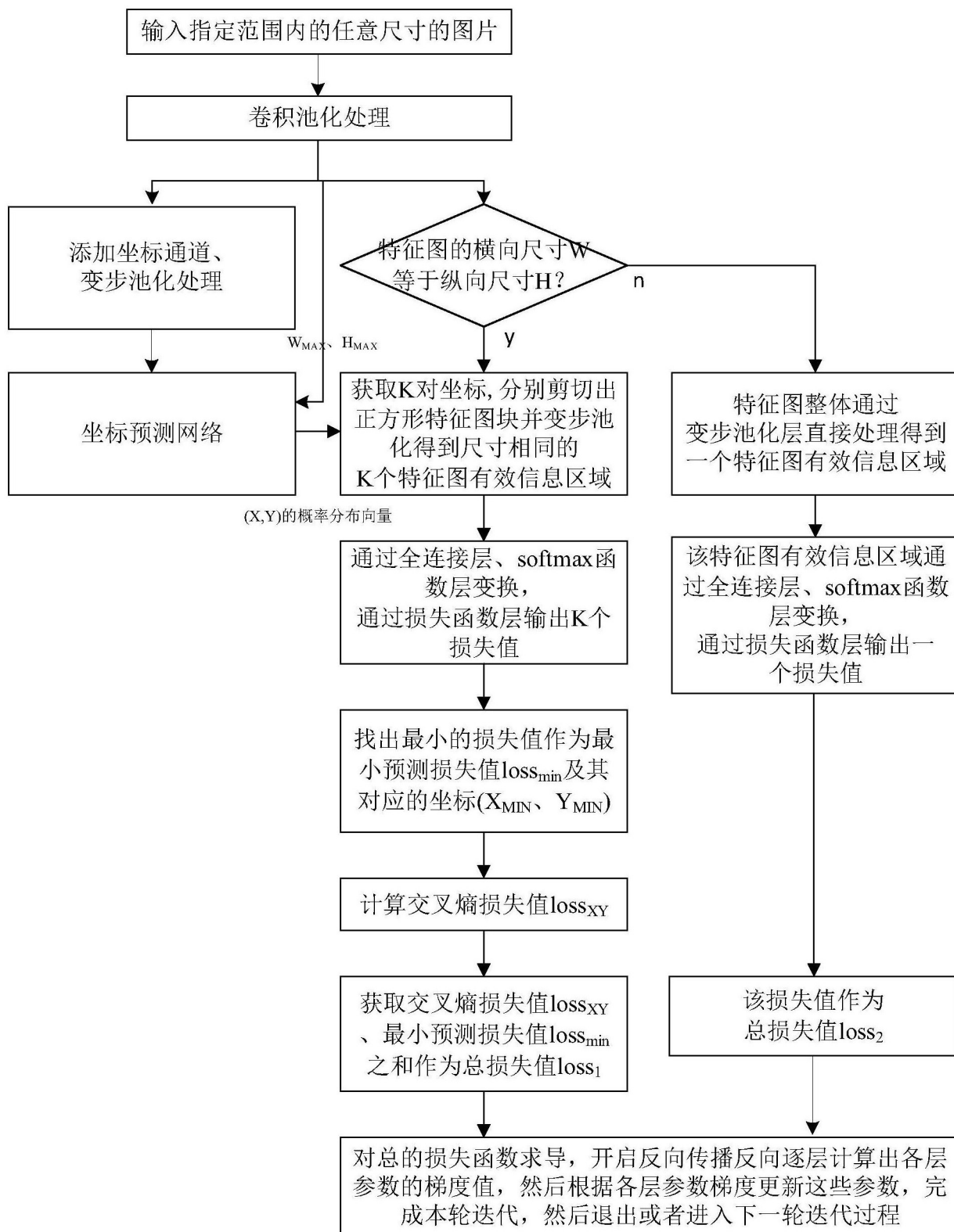


图1

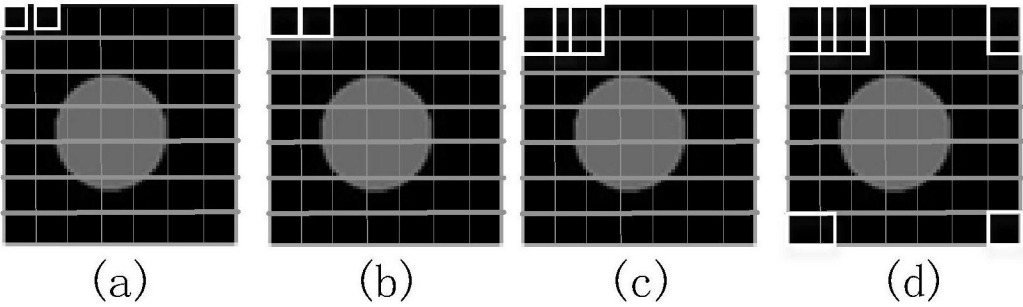
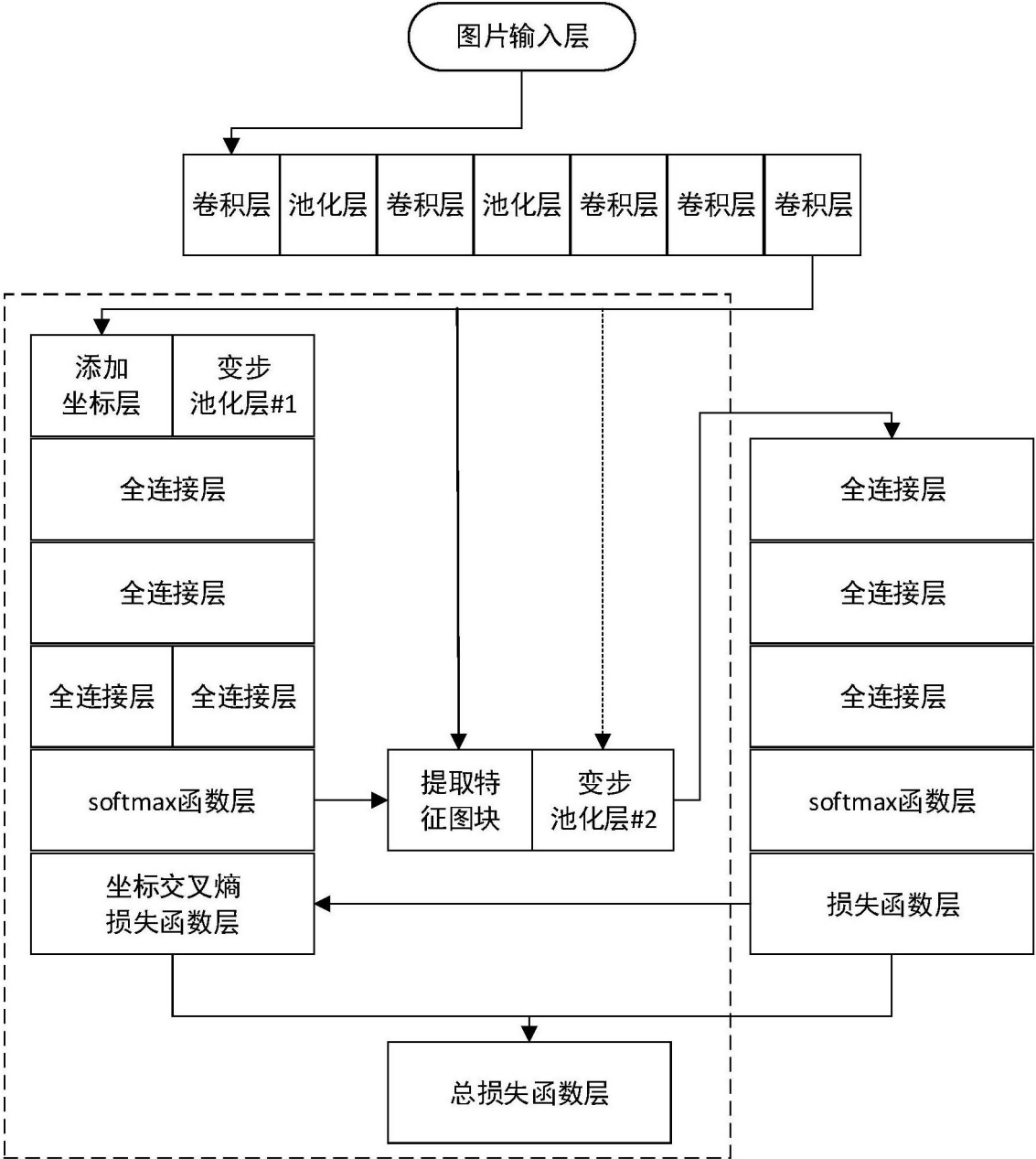


图3

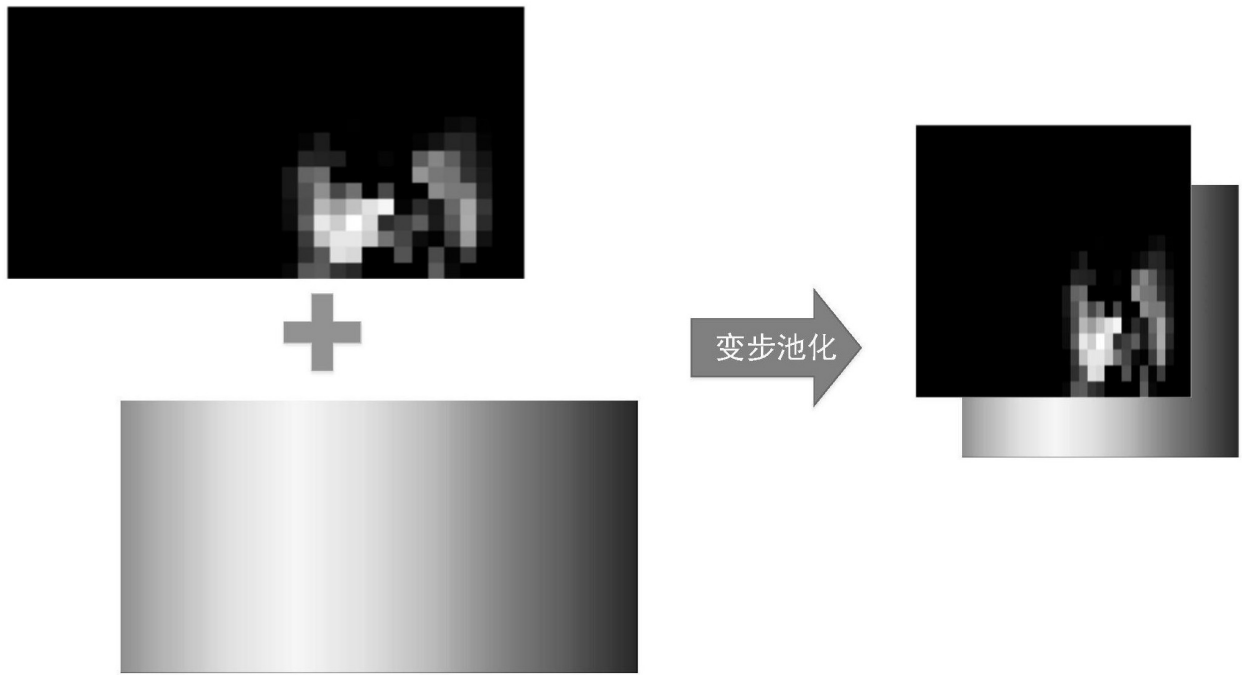


图4

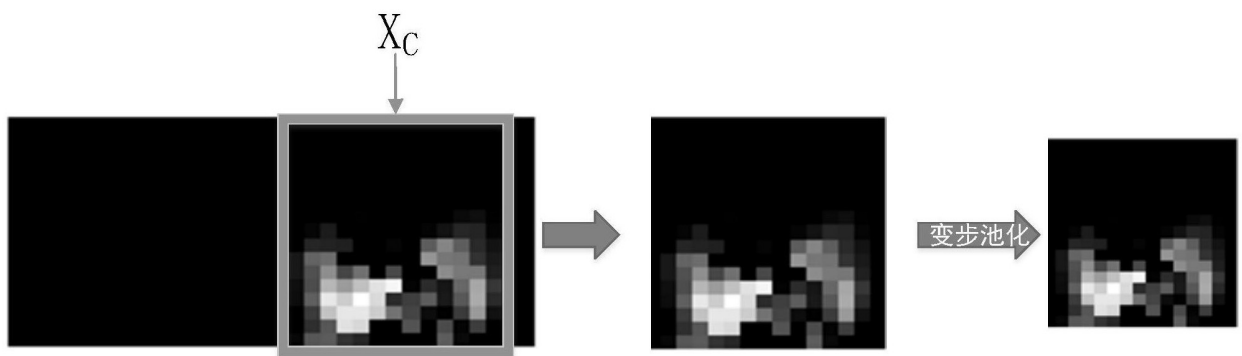


图5

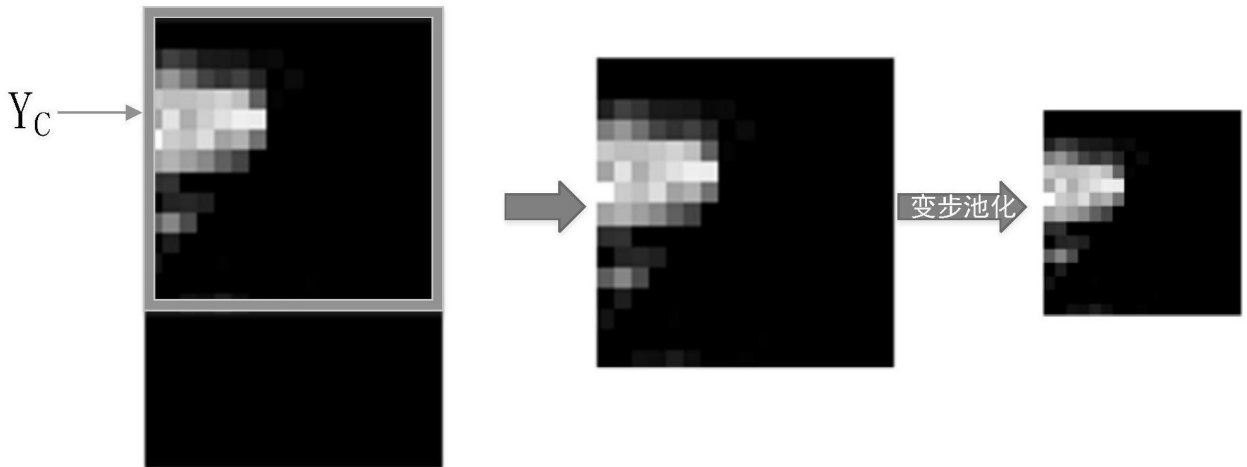


图6