**一种基于生成对抗网络的多模态配准图像生成的方法**

中山大学 瞿毅力

**技术领域**

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。

**背景技术**

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。

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和，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。

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。

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。

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。

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。

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。

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。

**发明内容**

为了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，本发明设计了一种基于SPP改进算子的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。本发明的具体技术方案如下：

1. 变步池化方法

SPP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SPP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。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。将变步池化分区后的最大池化操作改为卷积，则得到了变步卷积，变步卷积通过设置可以模拟普通卷积，同样实现了接收不定输入和生成固定输出。变步池化和变步卷积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。变步池化方法的具体处理过程如下：

1. 设定对输入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。
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。
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. 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。
5. 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。
6. 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。
7. 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。
8. 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的数量的特征图。

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，我们将上述池化操作过程称为变步池化。

坐标通道的添加根据使用场景是可选的。

1. 坐标预测方法

一个普通CNN中，不同形状的输入经过多层卷积和池化后得到不同形状的特征图，我们首先对得到的特征图添加坐标通道，即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。

在完成坐标通道的添加之后，我们得到不同形状的新特征图，在所有形状中，我们记录其中最大的横向尺寸为WMAX、最大的纵向尺寸为HMAX。然后我们将得到的不同形状的新特征图进行变步池化处理，得到相同尺寸的特征图。此时，可选择性地对该特征图进行卷积池化学习，然后将最后得到的特征图拉直成神经元向量，再将该神经元向量输入两个多层全连接网络，分别得到一个长度与WMAX相同的向量和一个长度与HMAX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。

1. 有效信息区域提取方法

我们接收到一对有效信息区域中心点坐标XC、YC和一个特征图。首先，我们设定变步池化的横纵方向的分区数量（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，最后将特征图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，具体情况处理过程如下：

1. 当W = H时，直接将特征图根据设置进行变步池化处理。
2. 当W > H时，我们以坐标XC对应的纵线为裁剪中线，裁取XC 对应的纵线两边各H/2宽的范围。就具体情况来说:
3. 当H/2 >= XC时，我们直接裁取 (0,0)、(0,H)、(H,H) 、(H,0)四个坐标点依次连线围成的区域；
4. 当H/2 >= W-XC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区域；
5. 当H/2 < XC < W- H/2时，我们裁取 (XC-H/2,0)、(XC-H/2,H)、(XC+H/2,H) 、(XC+H/2,0)四个坐标点依次连线围成的区域。

这样，我们得到一个横向纵向尺寸均为H的特征图块，然后将此特征图块根据设置进行变步池化处理。

1. 当W < H时，我们以坐标YC对应的横线为裁剪中线，裁取YC 对应的横线两边各W/2宽的范围。就具体情况来说:
2. 当W/2 >= YC时，我们直接裁取 (0,0)、(0,W)、(W,W) 、(W,0)四个坐标点依次连线围成的区域；
3. 当W/2 >= H-YC时，我们直接裁取 (H-W,0)、(H-W,W)、(H,W) 、(H,0)四个坐标点依次连线围成的区域；
4. 当W/2 < YC < H- W/2时，我们裁取 (YC-W/2,0)、(YC-W/2,W)、(YC+W/2,W) 、(YC+W/2,0)四个坐标点依次连线围成的区域。

这样，我们得到一个横向纵向尺寸均为W的特征图块，然后将此特征图块根据设置进行变步池化处理。

1. 根据预测有效信息区域进行概率分布向量预测的方法

在一个包含卷积池化部分和全连接层与损失函数部分的普通CNN中，我们首先分离其卷积池化部分和全连接层与损失函数部分，将不同形状的输入原网络的卷积池化部分经过多层卷积和池化得到不同形状的特征图，然后我们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。

当W 与H不相等时，我们对其进行如下处理：

1. 我们使用（2）中坐标预测方法得到X、Y坐标的概率分布向量，根据X、Y坐标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。然后针对如下两种情况：
2. 当W > 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；
3. 当W < H时， 将得到的概率最高的X坐标与概率前K的K个Y坐标组成K对坐标。
4. 然后，我们根据得到的这K对坐标将特征图按照（3）中的有效信息区域提取方法提取出K个特征图有效信息区域。
5. 我们将这K个特征图有效信息区域分别输入原网络的全连接层与损失函数部分逐层计算，直到得到原网络的损失函数层输出的K个损失值。我们比较这K个损失值，得到最小的损失值lossMIN以及其在前面K对坐标中对应的坐标(XMIN、YMIN)。预测时lossMIN 对应的概率分布向量即为预测结果。
6. 我们将得到的XMIN、YMIN作为标签，与①中坐标预测方法得到X、Y坐标的概率分布向量一起，分别计算交叉熵求和得到lossXY。计算公式如下：

lossXY = -(log(P(X=XMIN|X)) + log(P(Y=YMIN|Y)))

其中，P(X=XMIN|X)指X坐标的概率分布向量中，对应坐标为XMIN的概率值。

1. 最后我们将得到的lossMIN 与lossXY 求和得到的最终结果作为新的完整网络最后的损失loss，计算公式如下：

loss = lossMIN + lossXY 。

但当W = H时，我们可以简化上述过程为：

1. 直接将特征图根据（3）中的有效信息区域提取方法处理，得到特征图有效信息区域。
2. 我们将这个特征图有效信息区域输入原网络的全连接层与损失函数部分逐层计算，直到得到原网络的损失函数层输出的损失值loss，该值即作为新的完整网络最后的损失。预测时计算loss所用的概率分布向量即为预测结果。

这样我们实现了整个先根据预测坐标提取有效信息区域，再根据有效信息区域进行概率分布向量预测的过程。

**附图说明**

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。



图1 为本发明实施例采用的变步池化方法示意图，从左到右依次为：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



图2 为本发明实施例采用的坐标预测方法中特征图添加坐标通道再进行变步池化处理的示意图



图3 为本发明实施例采用的有效信息区域提取方法，包括W>H时根据XC裁取有效信息区域（上）和W<H时根据YC裁取有效信息区域（下）两种情况



图4 为本发明实施例采用的根据预测有效信息区域进行概率分布向量预测的方法流程图



图5 为本发明实施例采用方法的主流程



图6 AlexNet原网络结构（左）和应用本发明方法改进后的结构（右）

**具体实施方式**

下面结合附图及具体实施例对本发明的实施过程进行详细说明：

AlexNet是一个具有全连接层的经典模型，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。我们通过本发明中的方法对其进行改进，使其能接受不定形状的输入。AlexNet原网络结构和我们改进的结构对比图如图6所示。

1. 分离网络并计算卷积池化部分可能输出的最大尺寸值WMAX、HMAX

我们首先分离AlexNet的卷积池化部分和全连接层与损失函数部分。

每次迭代开始时，输入为一张任意形状的图片，经过AlexNet的卷积池化部分的计算，就得到一个特征图。我们记录特征图的横向尺寸W与纵向尺寸H，并根据数据集和应用范围计算或预估出，在可以得到的所有W和H中，最大的横向尺寸WMAX和最大的纵向尺寸HMAX。

1. 在卷积池化部分后添加变步池化层

该步骤以卷积池化部分网络输出特征图作为输入，先对特征图添加坐标通道，再对特征图进行变步池化处理，输出一个固定尺寸的新特征图。

1. 根据WMAX、HMAX和变步池化的输出尺寸设计坐标预测网络

坐标预测网络以变步池化层输出的新特征图为输入，然后输出两个坐标的概率分部向量。坐标预测网络连接在变步池化层之后，由一个近似AlexNet全连接网络的全连接网络组成。其中第一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AlexNet一致，以第一层的输出作为输入，输入输出长度均为4096；最后一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WMAX、HMAX，分别用来预测X坐标和Y坐标；最后一层全连接层之后是softmax函数，用来将最后一层全连接层的输出缩放为概率分布向量。

1. 根据坐标预测结果从卷积池化部分输出特征图提取有效信息区域

每一次的训练时，我们根据W、H的比较判断特征图的形状。

当W 与H不相等时，我们对得到的特征图进行如下处理：

1. 根据X、Y坐标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。然后针对如下两种情况：
2. 当W > 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；
3. 当W < H时， 将得到的概率最高的X坐标与概率前K的K个Y坐标组成K对坐标。
4. 然后，我们根据得到的这K对坐标将特征图提取出K个特征图有效信息区域：
5. 当W > H时，我们以坐标XC对应的纵线为裁剪中线，裁取XC 对应的纵线两边各H/2宽的范围。这样，我们得到一个横向纵向尺寸均为H的特征图块，然后将此特征图块根据设置进行变步池化处理。
6. 当W < H时，我们以坐标YC对应的横线为裁剪中线，裁取YC 对应的横线两边各W/2宽的范围。这样，我们得到一个横向纵向尺寸均为W的特征图块，然后将此特征图块根据设置进行变步池化处理。

当W = H时，我们可以简化上述过程为：

1. 直接将特征图进行变步池化处理，得到1个特征图有效信息区域，如图6中虚线所示。
2. 全连接层与损失函数部分对有效信息区域预测并计算损失

当W 与H不相等时，AlexNet的全连接层与损失函数部分有K个特征图有效信息区域输入，得到K个损失值：

1. 我们将这K个特征图有效信息区域分别输入AlexNet的全连接层与损失函数部分逐层计算，先得到AlexNet的全连接层输出并经过softmax函数缩放为K个概率分布向量，最后得到AlexNet的损失函数层输出的K个损失值。我们比较这K个损失值，得到最小的损失值lossMIN以及其在前面K对坐标中对应的坐标(XMIN、YMIN)。
2. 我们以得到的XMIN、YMIN作为标签，与坐标预测网络得到X、Y坐标的概率分布向量，一起作为坐标预测网络后的损失函数层的输入，分别计算两个坐标的交叉熵，求和得到lossXY。
3. 最后我们将得到的lossMIN 与lossXY 求和得到的最终结果作为新的完整网络最后的损失loss。
4. 训练时，在计算出总的loss值后，根据原网络设定的优化方法，依据loss的值对总的损失函数loss求导，根据求导链式法则开启反向传播，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。更新完成后开启下一次迭代，AlexNet的卷积池化部分的第一层接收下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，lossMIN 对应的AlexNet的全连接层输出概率分布向量即为预测结果。

当W = H时，AlexNet的全连接层与损失函数部分仅有1个特征图有效信息区域输入和1个损失值输出，我们可以简化上述过程为：

1. 我们将接收到的唯一一个特征图有效信息区域输入AlexNet的全连接层与损失函数部分逐层计算，得到AlexNet的全连接层输出并经过softmax函数缩放为1个概率分布向量，最后得到AlexNet的损失函数层输出的1个损失值loss，该值即作为新的完整网络最后的损失。
2. 训练时，在计算出总的loss值后，根据原网络设定的优化方法，依据loss的值对总的损失函数loss求导，根据求导链式法则开启反向传播，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。更新完成后开启下一次迭代，AlexNet的卷积池化部分的第一层接收下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，AlexNet的全连接层输出唯一一个概率分布向量即为预测结果。

AlexNet经过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。