Neuron-level Selective Context Aggregation for Scene Segmentation

摘要

上下文信息为场景分割中区分视觉相似的像素提供了重要线索。在本文中，我们介绍了一个用于场景分割的神经元级的选择性上下文聚合（SCA）模块，由一个上下文依赖性预测器和一个上下文聚合运算器组成。依赖性预测器是隐性训练的，用于推断不同图像区域之间的上下文依赖关系。上下文聚合运算器用全局上下文来增强局部表征，根据其即时预测的依赖关系有选择地在每个神经元上聚合。所提出的机制使数据驱动的上下文依赖性推断成为可能，并促进了上下文感知的特征学习。所提出的方法在具有挑战性的场景分割数据集上改善了建立在VGG16基础上的强大基线，这证明了其在对上下文信息建模方面的有效性。

1. 介绍

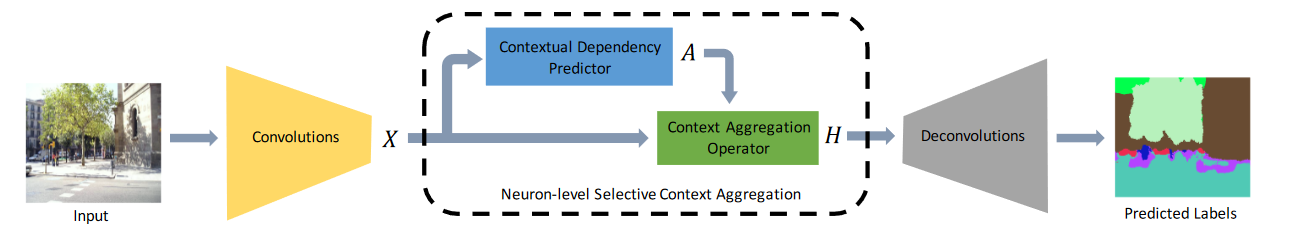
场景分割是计算机视觉中一个长期存在的基本问题，其目的是将语义对象的类别标签与场景图像中的每个像素联系起来。人们早已认识到，考虑到图像区域之间的语义背景可以显著提高场景分割算法的性能[8, 14, 21, 26]。正如计算机视觉的许多其他领域一样，场景分割的最先进技术已经通过使用深度神经网络得到了改变。Long等人[13]提出了一个完全卷积网络，它可以通过端到端的训练来预测一个密集的标签图，能够提取粗略的物体形状。Noh等人[15]对这种方法进行了扩展，引入了多个反卷积层，以逐步捕捉更精细的细节。在这些基于CNN的纯方法中，卷积特征图中的神经元表征是从有限的感受野中提取的，只捕捉到了局部信息。为了解决这个问题，Liu等人[12]提出使用一个层的平均特征作为全局背景来增强每个神经元的局部特征。Yu和Koltun[28]引入扩张卷积，在不损失分辨率的情况下系统地聚集多尺度的上下文信息。Zhao等人[30]通过金字塔集合，以不同的基于区域的上下文聚合来利用全局上下文信息。尽管这些方法提高了分割的准确性，但它们以预定义的方式聚集上下文，而这并不取决于输入图像。因此，它们缺乏处理不同图像区域间非均匀依赖关系的内部机制。在另一条研究路线中，Shuai等人[24]提出了有向无环图循环神经网络（DAGRNNs）来利用上下文信息，从而提高卷积特征的表示能力。他们的方法在三个具有挑战性的数据集上显示出比最先进的技术有了很大的进步。然而，由于RNNs的梯度消失问题[1, 16]，上下文信息只能传播到附近的神经元，因此不能很好地捕获长距离的依赖关系。事实上，Shuai等人[23]报告说，使用由8个连接的网格诱导的DAGRNNs，明显优于4个连接的情况。此外，由于RNN的使用，DAG-RNN的隐藏向量必须顺序更新，使得它的效率低于基于CNN的纯方法。在这项工作中，我们提出了一种新的方法，在每个神经元上选择性地聚合上下文信息，并通过引入一个新的上下文聚合模块，以输入依赖的方式进行。该模块将卷积层的特征图作为输入，并输出一个具有语境的特征图。具体来说，该模块由两个主要部分组成：一个上下文依赖预测器和一个上下文聚合运算器。依赖性预测器是一个辅助网络，它推断出代表每对神经元之间依赖程度的系数。请注意，这些系数是与输入有关的。上下文聚合器根据其预测的系数将每个输出神经元与所有输入神经元密集地连接起来，产生具有增强的上下文表示能力的特征图。然后，这些特征图被送入一个反卷积网络，以预测一个语义标签图（见图1）。我们把所提出的模块称为神经元水平的选择性聚合（SCA）模块，因为它的每个输出神经元都使用不同的权重集来聚合来自输入神经元的上下文特征。总之，我们的工作有三个主要贡献。  
- 我们提出了一个新的上下文聚合模块，该模块根据深度卷积特征图中神经元之间特定的图像依赖关系，有选择地将全局上下文信息注入到局部表征中。

- 我们引入了一个上下文依赖性预测器，它可以学习推断出神经元之间特定的图像依赖关系。由此产生的依赖系数允许每个神经元有选择地聚集来自整个图像的背景信息。

- 我们将所提出的网络应用于场景分割任务，并在两个大型数据集上证明其有效性和潜力。  
本文的其余部分组织如下。第2节讨论了与我们的方法有关的一些工作，第3节介绍了拟议的新模块的两个组成部分，第4节给出了实施细节，最后，第5节报告了我们的实验结果。

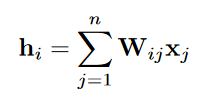
3.

上下文信息已被广泛用于计算机视觉应用中[8, 14, 21, 26]。一般来说，上下文可以指任何对表示能力有促进作用的全局信息。然而，请注意，一个局部表征可能会受到其背景中不同语义类别的像素的不同影响。例如，与 "山 "的像素相比，图像中 "海 "的像素的存在对于将附近的沙色像素分类为 "海滩 "是一个更重要的线索，因为 "海 "和 "海滩 "有更高的共同发生的相关性。然而，以前的方法[12, 28, 30]通常以固定和统一的方式将全局背景嵌入到局部表示中，因此缺乏内部机制来有效地考虑不同图像区域之间的非统一的依赖关系。为了提高处理非均匀上下文信息的能力，我们提出了一种与输入有关的方式，有选择地将上下文纳入本地表示。

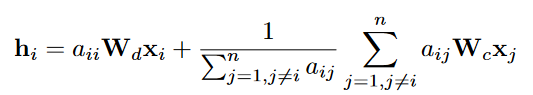
图1显示了我们提出的分割网络架构的概况。给定一个输入图像I，卷积层逐渐提取其抽象特征，产生一个高水平的特征图X。SCA模块由两部分组成。第一个部分是上下文依赖预测网络，它把X作为输入，通过几个隐藏层预测一个矩阵A，它由X中所有神经元对之间的依赖系数组成。依赖系数aij试图预测神经元xj在神经元xi的上下文中应该被考虑的程度。接下来，上下文聚合运算器使用预测的依赖矩阵A将输入特征图X转化为上下文感知的特征图H。最后，在H上应用几个去卷积层来预测最终的语义分割标签图。  


3.1.绿色部分

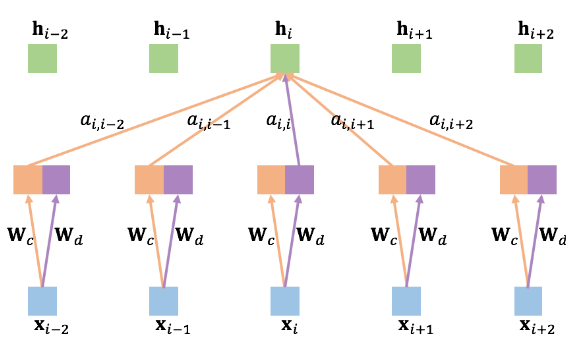
一般来说，情境聚合算子的目的是将全局情境信息编码到局部神经元表示中。它将一个特征图X作为输入，并产生一个具有相同空间维度（但可能具有不同通道数）的上下文感知特征图H。为了充分利用上下文，这个运算器的接受域应该覆盖整个图像。一个天真的方法是使用全连接层，其中每个输出神经元都与所有输入神经元相连。更具体地说，让xi∈RN和hi∈RM分别表示X和H中第i个神经元的特征向量。使用全连接层，H中的每个神经元由

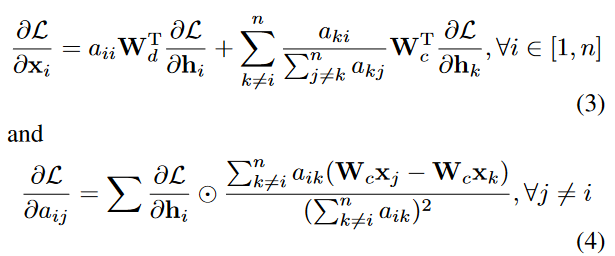


其中n是X中输入神经元的数量，Wij∈RM×N是权重矩阵。可以看出，Wij对第i个输入神经元和第j个输出神经元是特别的。因此，1.学到的权重是位置敏感的，这对密集标签预测来说是不可取的，因为同一类别的像素可以出现在图像的任何地方。此外，在训练后，Wij在测试时间内是固定的，因此2.（测试阶段）上下文聚合是与输入无关的。最后，全连接层消耗了大量的可学习参数，使得它3.很容易对训练数据进行过度拟合。为了解决所有这些问题，我们引入了一个新的聚合算子。我们的算子首先使用两个1×1的卷积来为每个神经元提取两个不同的特征，即身份特征和背景特征。我们的直觉是，用于描述神经元本身的特征和描述其对其他神经元背景的贡献的特征可能是不同的。然后，一个神经元的输出特征是由它自己的身份特征和所有其他神经元的背景特征的线性组合产生的根据一些输入依赖的系数。形式上，输出神经元hi的计算被定义为：



其中aij是控制从xj到hi的信息量的系数，Wd∈RM×N和Wc∈RM×N分别是用于提取身份特征和背景特征的权重矩阵。在实践中，我们总是把aii设置为1，这样输出神经元就会提取其所有的身份特征。图2说明了在一维特征图上单个输出神经元的拟议算子的计算过程。（别人的贡献+自己的特征）

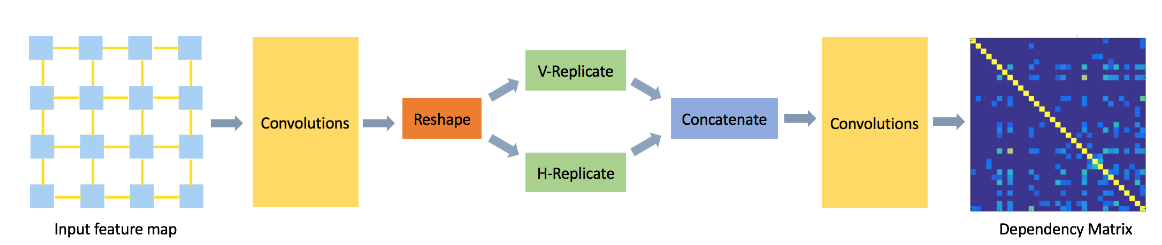
   
值得注意的是，这个新算子有几个吸引人的特性。首先，矩阵Wd、Wc是跨神经元共享的，将参数空间从n2N M（对于全连接层）减少到2N M。其次，该操作保留了空间信息，因此它适合应用于密集预测任务。第三，系数aij不是固定参数，能够代表不同神经元对之间的任意依赖关系。这些依赖关系可以根据一些先验知识手动定义，也可以针对特定的输入X自动推断。这里我们采用数据驱动的范式，并引入一个辅助的上下文依赖预测器（第3.2节），学习推断神经元之间的依赖关系，并为拟议的上下文聚合算子提供即时的输入依赖预测。为了允许通过拟议的运算器进行反向传播，我们必须定义损失L的梯度，即关于输入特征图的神经元xi和系数aij。这些梯度由以下公式给出



其中，L/∂hi表示从顶层收到的关于hi的损失梯度，表示元素相乘，（4）中最外层的和是在特征通道上。

3.2.蓝色部分

上下文依赖性预测器（CDP）的目的是估计输入特征的神经元之间的依赖关系。更确切地说，给定一对神经元（xi, xj），目标是预测从xj中提取的上下文特征在汇总神经元xi的上下文时应该被赋予的权重。由于每对神经元可能被赋予不同的权重，全局性的上下文信息被选择性地编码到每个神经元的局部表示中。神经元之间的依赖函数可以用神经网络来建模。通常情况下，孪生网络可以用来学习这样的函数，它在寻找两个输入之间的固有关系的任务中很受欢迎。训练孪生网络需要成对的输入和它们相应的基础事实。然而，我们的输入特征图并不是以神经元对的形式组织的，更重要的是，我们没有定义好的基础真理来说明这些神经元对之间的依赖关系。因此，我们的方法是通过训练CDP和上下文聚合算子（以及网络的其他部分）来隐含地学习依赖函数，以最小化最终的分割损失。CDP应该把有n个神经元的特征图X作为输入，并输出一个n×n矩阵A，其中每个元素aij代表神经元xi和xj之间的依赖系数，用于上下文聚合运算。为了实现这一目标，我们提出了一个新的网络结构，描述如下；也见图3。



首先，在X上应用几个1×1的卷积层，将每个神经元的表示映射到一个更高层次的特征空间。直观地说，这种映射的目的是防止网络简单地计算神经元之间的线性关联；相反，网络能够探索更高层次的语义关系，这些关系可能被证明对最终的语义分割有用。为了有效地预测所有神经元对的依赖关系，我们使用了一个简单的技巧来构建一个大的特征图，其中每个位置都存储了一个特定神经元对的映射特征的连接。具体来说，我们将二维特征图重塑为一维特征向量，分别使用水平和垂直复制将其扩展为两个二维方形特征图，然后通过深度通道将这两个复制的特征图连接起来。利用所得到的大特征图，我们只需再应用一个1×1卷积层就可以产生依赖矩阵。根据公式（4）CDP可以接收来自上层的梯度，这使得它可以与主网络联合训练而不需要使用额外的监督。它通过为不同的神经元对产生适当的依赖系数，学习如何有选择地纳入上下文信息，以帮助在训练中最小化整体分割成本。

下面开始技术总结：

