交通场景理解的空间卷积神经网络（SCNN）

1. 摘要

卷积神经网络通常通过一层一层的堆积卷积操作来生成。虽然卷积神经网络已经展示出了从原始像素中提取语义特征的强大能力，但是它通过一个图像的行和列来获取像素的空间相关性的能力没有完全展现出来。这些关系对于学习具有强烈形状先验但外观连贯性较弱的语义对象非常重要，例如交通线路，如图1（a）所示，它们经常被遮挡或甚至不涂在路面上。在本文中，我们提出了空间CNN（SCNN），它将传统的深层逐层卷积推广到特征映射中的切片卷积，从而实现层中行和列之间的像素之间的消息传递。这种SCNN特别适用于长期连续形状结构或大型物体，具有强烈的空间关系但较少的外观线索，如交通车道，杆和墙。

我们在新发布的极具挑战性的车道检测数据集和Cityscapse数据集上应用SCNN。结果显示SCNN能学习到物体结构空间相关性并且性能得到了很大的提高。实验表明我们的SCNN比基于RetNet的RNN和MRF+CNN（MRFNet）在道路监测中分别高出了8.7%和4.6%的准确率。而且我们的SCNN在TuSimple Benchmark Lane Detection比赛中以96.53%取得了第一的好成绩。

1. 课题介绍

在最近几年中，自动驾驶在学术界和工业界已经得到了很大的关注。而交通场景理解是自动驾驶最具挑战性的任务之一。它包括很多计算机视觉上的很多任务例如：车道检测和语义分割等。车道监测帮助指导车辆和能被用在驾驶辅助系统上，而语义分割提供周围物体例如车辆或者行人。更具体的位置。然而，在现实应用中考虑到苛刻的场景（例如：坏的天气状况，昏暗的或者耀眼的光线），这些任务是非常具有挑战性的。另一个交通场景理解的挑战是在很多应用中，特别是在车道检测中，我们需要先处理具有强大结构的物体，但是具有较少的外观线索，如车道标记和杆子等长的连续形状，可能会被遮挡。 例如，在图1（a）的第一个例子中，汽车在右侧完全封闭最右边的车道标记。

虽然。由于强大的表达学习能力，基于CNN的方法（Krizhevsky, Sutskever,and Hinton 2012; Long, Shelhamer, and Darrell 2015）已经把场景理解带到一个新的水平。但是它对于具有长结构区域并且可以被遮挡的物体，例如图1中红色边界框中所示的车道标记和杆。仍然表现的不是非常好。然而，人可以很容易的推断出它们的位置和能从周围环境中填补遮挡的部分。（既可见部分）

为了解决这个问题，我们提出了空间卷积神经网络（SCNN）既深度神经网络到丰富的空间层次的一个扩展。在一层叠一层的卷积神经网络中（CNN）一个卷积层的输入时前一层的使用卷积操作非线性激活函数的结果。该过程按顺序完成。相似地SCNN把特征图的行或者列当成层，然后顺序地在层上执行卷积操作非线性激活，然后求和操作，这样就形成了深度卷积神经网络。用这种方法在同一层的神经元能够互相传播信息。它对有结构物体例如：车道，标杆，卡车，特别有用，因为可以通过层内传播来增强空间信息。如图1所示，在CNN不连续或杂乱的情况下，SCNN可以很好地保持车道标记和杆的平滑性和连续性。在我们的实验中，SCNN比其它基于RNN或者MRF/CRF的方法表现的更出色。也比深度ResNet-101获得了更好的成绩.

相关工作，对于车道检测，大多数存在的算法使基于手工的低阶特征（Aly 2008; Son et al.2015;jung, youn,and sull2016）,这就限制了他们处理极端状况的能力。只有Huval et al(2015)在车道检测中采用深度学习的方式进行了首次的尝试。但是当时没有大的通用数据集。然而

对于语义分割，基于CNN的方法已经成为了主流，也取得了巨大的成功（Long, Shelhamer, and Darrell 2015; Chen et al.2017）

当然也有其他的尝试去利用在神经网络中的空间信息。Visin et al.2015和Bell et al.2016Visin等人使用递归神经网络沿每行或每列传递信息，因此在一个RNN层中，每个像素位置只能从同一行或列接收信息。Liang et al.2016a,b提出了LSTM的变体来利用语义对象解析中的上下文信息，但是这样的模型在计算上是昂贵的。学者还有尝试把CNN与MRF或CRF等图形模型结合起来，在这种方法中信息通过大卷积核的卷积操作来实现传递的（Liu et al.2015; Tompson et al.2014; Chu et al.2016）。与上述方法相比，CNN有三个优点:（1）顺序消息传递方案比传统的密集MRF / CRF计算效率更高，（2）消息作为残差传播，使SCNN易于训练，（3）SCNN灵活，可应用于任何级别 深度神经网络。

1. 空间卷积神经网络
2. 数据集介绍：

**车道检测数据集**

在本文中，我们提出了一个具有挑战性的大型数据集用于行车道检测。尽管交通车道检测重要而且困难，但是存在的数据集不是太小就是太简单，我们需要一个大的公开的标记好的基准数据集来比较不同的方法（Bar Hillel et al. 2014）。KITTI（Fritsch, Kuhnl, and Geiger 2013）和CamVid（Brostow et al. 2008）包括对车道标记的像素级别的标注，但是仅仅只有几百张照片，对于深度学习来说太小了。Caltech Lanes Dataset(Aly 2008)和最近才开源出来的TuSimple Benchmark(TuSimple 2017)分别包含1224张和6408张带车道标注的图片，而交通处于受限制的情况下，交通量很小，车道标记清晰。此外，这些数据集中没有一个注释由于划分而被遮挡或看不见的车道标记，而这种车道标记可以由人推断并且在实际应用中具有高价值。

为了去收集数据，我们在不同的天，我们让不同的司机开着安装了照相机的六辆车在北京城里跑来录像。收集了超过55小时的视频，提取了133,235帧，这是TuSimple数据集的20多倍。我们把收集到的数据集分成88880张的训练集和9675张的验证集和34680张的测试集。这些图像使用（Scaramuzza，Martinelli和Siegwart 2006）中的工具不失真，分辨率为1640×590。 图2（a）示出了一些示例，其包括城市，乡村和高速公路场景。 作为一个最大的最拥挤的城市之一，北京提供了对道路检测的很多具有挑战性的交通场景。我们将测试集分为正常类和8个具有挑战性的类别，这些类别对应于图2（a）中的9个示例。 图2（b）显示了每种情况的比例。可以看出，8个具有挑战性的场景占据了数据集的大部分（72.3％）。

对于每一帧图像，我们使用三次样条（cubic splines）手动注释交通车道。如前所述，在许多情况下，车道标记被车辆遮挡或看不见。在实际应用中，重要的是车道检测算法即使在频繁发生的这些挑战性场景中也可以从上下文估计车道位置。因此，对于这些情况，我们仍然根据上下文注释通道，如图2（a）（2）（4）所示。我们也希望我们的算法可以区分道路障碍，如图2（a）（1）所示。因此，屏障另一侧的车道没有注释。

在本文中，我们将注意力集中在四个车道标记的检测上，这些标记在实际应用中最受关注。其他车道标记未注释。

1. 空间卷积神经网络

传统方法建立空间关系模型的方法是基于马尔可夫随机场（MRF）或者条件随机场（CRF）（Krahenbuhl and Koltum 2011）.最近（Zheng et al.2015; liu et al.2015 chen et al.2017）将它们与CNN结合起来的工作都遵循图3的流程。其中均值场算法可以用神经网络实现。特别地，这个处理过程是：（1）正则化：CNN的输出被视为unary potentials，并通过Softmax操作进行归一化，（2）消息传递，可以通过使用大卷积核的通道方式卷积来实现（对于密集的CRF，卷积核大小将覆盖整个图像，卷积核权重取决于输入图像）（3）兼容性转换，可以用1×1卷积层实现（4）增加unary potentials。 该过程迭代N次以给出最终输出。

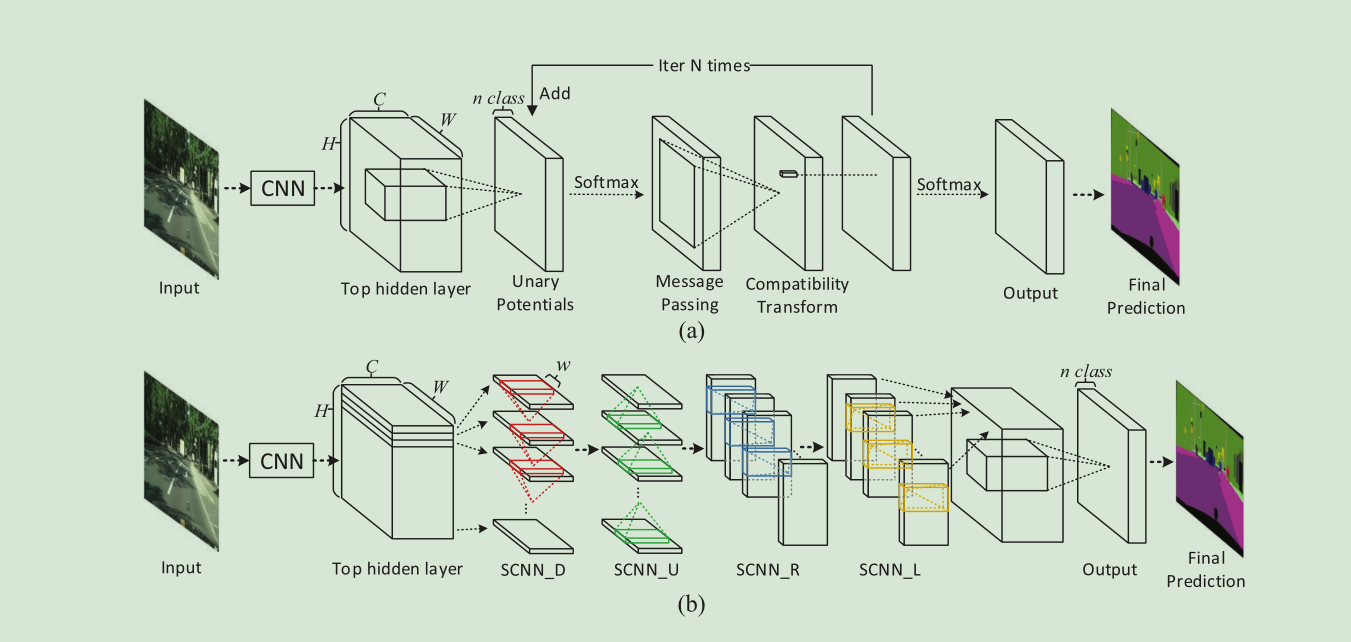


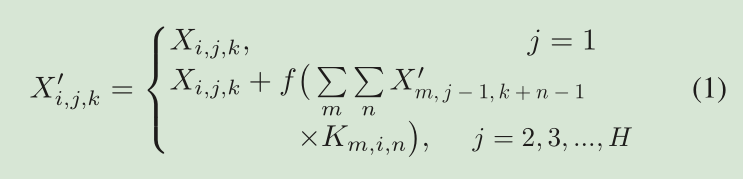
Figure3.（a）基于MRF/CRF的方法（b）SCNN的实现。

可以看出，在传统方法的消息传递过程中，每个像素接收来自所有其他像素的信息，这非常耗费计算并且难以在自动驾驶中用于实时任务。对于MRF，大卷积内核很难学习，并且通常需要仔细初始化(Tompson et al. 2014; Liu et al. 2015). 此外，这些方法应用于CNN的输出，而包含更丰富信息的顶层隐藏层可能是建模空间关系的更好地方。

为了解决这些问题，并且为了更有效地学习车道标记的空间关系和平滑连续的先验，或者在驾驶场景中的其他结构化对象，我们提出了空间CNN。请注意，此处的“空间”与“空间卷积”中的“空间”不同，但表示通过特殊设计的CNN结构传播空间信息。

如图3（b）的'SCNN D'模块所示，考虑应用于大小为C×H×W的3-D张量的SCNN，其中C，H和W分别表示通道数量，行的数量， 和列的数量。这个张量被分成H个小切片，然后第一个小切片被送到一个具有C个大小为C\*w核卷积层, w是这个卷积核的宽度。在传统的卷积神经网络，一个卷积层的输出是被送入到下一层的，然而这里我们把这个输出添加到下一个切片去生成一个新的切片。这个新的切片然后被送到下一个卷积层，这个过程直到最后一个切片被更新为止。

特别地，假设我们有一个3维的卷积核K，其中的元素Ki,j,k表示最后一个切片第i个通道元素与当前的切片的第j个通道元素之间的权重加上两个元素之间的k个偏移量。我们也设输入3维的张量X，其中元素为Xi,j,k，其中i，j，k分别表示通道，行和列的索引。那么SCNN前向传播计算是：



其中f为一个非线性激活函数ReLU。这个加上求导符号的X表示这个元素已经被更新了。说明一下：卷积核权重在所有切片之间共享，因此SCNN是一种递归神经网络。也需要说明一下SCNN是有方向的。在图3（b）中，带有后缀“D”，“U”，“R”，“L”的四个“SCNN”模块分别表示向下，向上，向右和向左的SCNN。

1. **分析**

与传统方法相比SCNN有三大主要优势,如下所示：

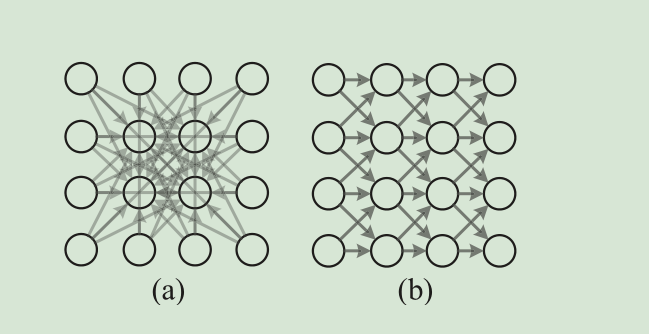


图4.消息传递方向a是密集MRF/CRF b是SCNN。对于a仅显示传递到内部4个像素的消息以进行清除。

1. 高效的计算性能。如图4所示，在密集的MRF/CRF中每一个像素接受其他所有像素的信息。这会有很多冗余信息。CNN消息传递以顺序传播方案实现。特别地，假设一个有H行W列的张量，在密集MRF/CRF中，每两个WH像素之间都有消息传递。经过n次迭代，这个信息传递的数量为n\*W2N2。在SCNN中，每一个像素仅仅收到来自w个像素的信息。因此这个信息传递的数量为ndir\*WHw，ndir为在SCNN中传播方向的数量，w为卷积核的宽度。ndir能在10到100取值，然而在这个论文中ndir是4，也就是4个方向。而w通常不比10大（在图4b中w=3）。从中可以看出，对于有成百个行和列的通道，SCNN能节省很多计算量，而且每一个像素仍然能够收到通过四个方向的其他所有像素的消息。
2. 消息作为残差。

在MRF / CRF中，通过所有像素的加权和来实现消息传递，根据前一段，其是计算上昂贵的。考虑到如此多的行或列，基于递归神经网络的方法可能会受到梯度下降（Pascanu，Mikolov和Bengio 2013）的影响。然而，深度残差网络却能轻松的在非常深的神经网络上训练。相似的，在我们的SCNN中消息作为残差传播。既是ReLU的输出在图1中。这个残差也可以被视为对原始神经元的一种修改。正如我们的实验展示的那样，如此信息传递的方法取得了比基于LSTM的方法更好的效果。

1. 灵活性。

由于SCNN的计算效率，它可以很容易地合并到CNN的任何部分，而不仅仅是输出。通常，顶层隐藏层包含丰富且高语义的信息，因此是应用SCNN的理想场所。通常，图3显示了我们在LargeFOV（Chen et al.2017）模型上实现SCNN。在顶部隐藏层（'fc7'层）之后依次添加四个空间方向上的SCNN以引入空间消息传播。

1. 实验

我们在车道检测数据集和Cityscapes上评估SCNN（Cordts等人2016），在两个任务中，我们使用标准的SGD训练模型，批量大小为12，基本学习率为0.01，动量为0.9，权重衰减为0.0001。学习率策略是“poly”，功率和迭代次数分别设置为0.9和60K。我们的模型基于LargeFOV模型进行了修改（Chen et al.2017）。前13个卷积层的初始权重是从ImageNet训练的VGG16（Simonyan and Zisserman 2015）中复制的（Deng et al.2009）。所有实验都在Torch7（Collobert，Kavukcuoglu和Farabet 2011）框架上实施。

**车道监测**

车道监测模型 与仅需要边界框的普通物体检测任务不同，车道检测需要精确的曲线预测。一个自然的想法是模型应该输出这些曲线的概率图（probmaps），因此我们生成像素级目标来训练网络，就像在语义分割任务中一样。一个自然的想法是模型应该输出这些曲线的概率图（probmaps），因此我们生成像素级目标来训练网络，就像在语义分割任务中一样。我们希望神经网络能够区分不同的车道标记，而不是将不同的车道标记视为一个类并在之后进行聚类，这样可以更加稳健。因此，这四个车道被视为不同的类别。 此外，然后将问题发送到小型网络以预测车道标记的存在。

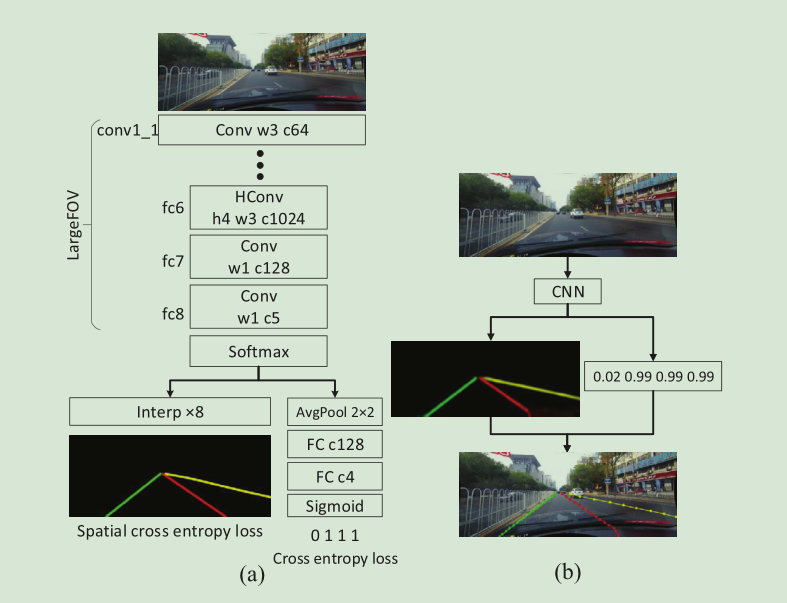
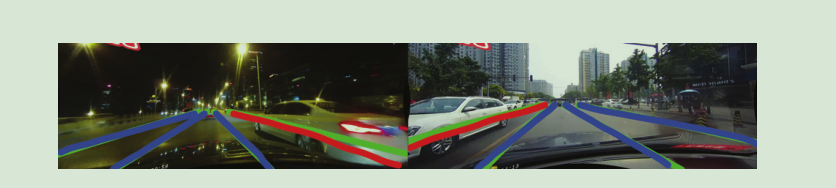


图5.a训练模型b车道预测过程 'Conv'，'HConv'和'FC'分别表示卷积层，萎缩卷积层（Chen等2017）和完全连接层。’c’, ’w’, 和 ’h’分别表示输出的通道数目、卷积核的宽、萎缩卷积层的速率

在测试过程中，我们仍然需要从问题转到曲线。如图5（b）所示，对于存在值大于0.5的每个车道标记，我们每隔20行搜索相应的probmap，以获得响应最高的位置。然后通过三次样条连接这些位置，这是最终的预测。

如图5a所示，我们的基线模型和LargeFOV之间的详细差异是：（1）fc7层输出的通道数为128.（2）'fc6'的atrous卷积层的'rate'设置为4。（3）在每个ReLU层之前添加批量标准化（Ioffe and Szegedy 2015）（4）添加一个小网络来预测车道标记的存在。

在训练期间，目标的线宽设置为16像素，输入和目标图像重新缩放为800×288，考虑到背景和车道标记之间的不平衡标签，背景损失乘以0.4。



图片6 基于IoU的评估。 绿线表示基本事实，而蓝线和红线表示TP和FP。

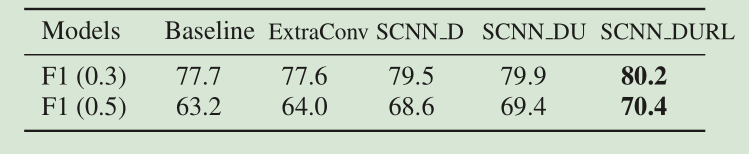


表1 CNN上具有不同方向设置的实验结果F1表示F1测量值，括号中的值表示IoU阈值。 后缀“D”，“U”，“R”，“L”分别表示向下，向上，向右和向左。

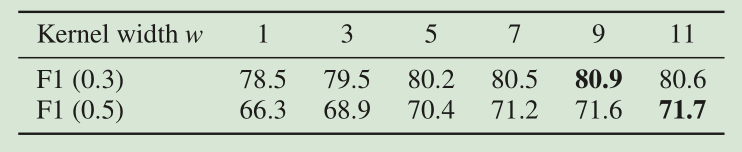


表2 SCNN在不同卷积核宽度上的实验结果。

**评价**

为了判断是否成功检测到车道标记，我们将车道标记视为宽度等于30像素的线，并计算地面实况和预测之间的交叉联合（IoU）。正如图6所示IoUs大于特定阈值的预测被视为真实的正面（TP）在这里，我们考虑0.3和0.5阈值对应松散和严格的评估。然后我们使用作为最后的评价索引。其中β=1对应于调和平均值（F1测量）。

**消融研究**

在2.2节中，我们提出了空间CNN来实现空间消息传播，为了验证我们的方法，我们将在本小节中进行详细的消融研究，我们的SCNN实现如图3所示。

1. 多向SCNN的有效性。首先，我们研究了SCNN中方向的影响。我们尝试具有不同方向实现的SCNN，结果如表1所示。这里SCNN的内核宽度w设置为5。可以看出，随着更多方向的增加，性能也会提高。为了证明改进不是来自更多参数，而是来自SCNN带来的消息传递方案，我们在顶部隐藏层之后添加了一个额外的5×5内核宽度的卷积层。基线模型并与我们的方法进行比较。从结果我们可以看出，额外的卷积层只能带来很小的改进，这证明了SCNN的有效性。
2. 卷积核宽度w的影响。

我们进一步尝试基于“SCNN DURL”模型的不同内核宽度的SCNN，如表2所示。这里内核宽度表示像素可以从中接收消息的像素数，并且w = 1的情况与方法类似（Visin etal. 2015; Bell et al. 2016）。 结果表明，较大的w是有益的，并且w = 9给出了令人满意的结果，其超过基线的显着边际8.4％和3.2％对应于不同的IoU阈值。

1. 空间CNN在不同的位置。如前所述，SCNN可以添加到神经网络的任何位置。这里我们考虑应用于（1）输出的SCNN DURL模型和（2）顶部隐藏层，其对应于图3。结果见表3指示包含比输出更丰富的信息的顶部隐藏层被证明是应用SCNN的更好位置。

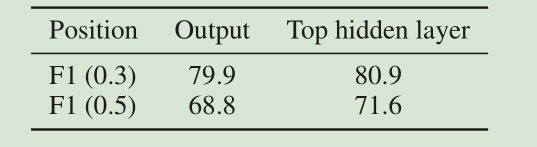


表3空间CNN在不同位置的实验结果，w = 9。

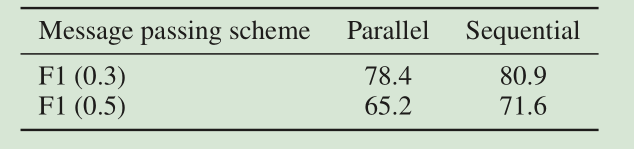


表4. 顺序和并行消息传递方案的比较，对于w = 9的SCNN DULR。

1. 顺序传播的有效性

在我们的SCNN中，信息以顺序方式传播，即，切片不将信息传递到下一个切片，直到它从前切片接收到信息。为了验证该方案的有效性，我们将其与并行传播进行比较，即，每个片在更新之前同时将信息传递给下一个片。对于并行的情况下这个等式1右边的‘被去掉。如表。 如图4所示，顺序消息传递方案明显优于并行方案。该结果表明，在SCNN中，像素不仅仅受到附近像素的影响，而且确实从其他位置接收信息。

1. 与最先进的方法进行比较。

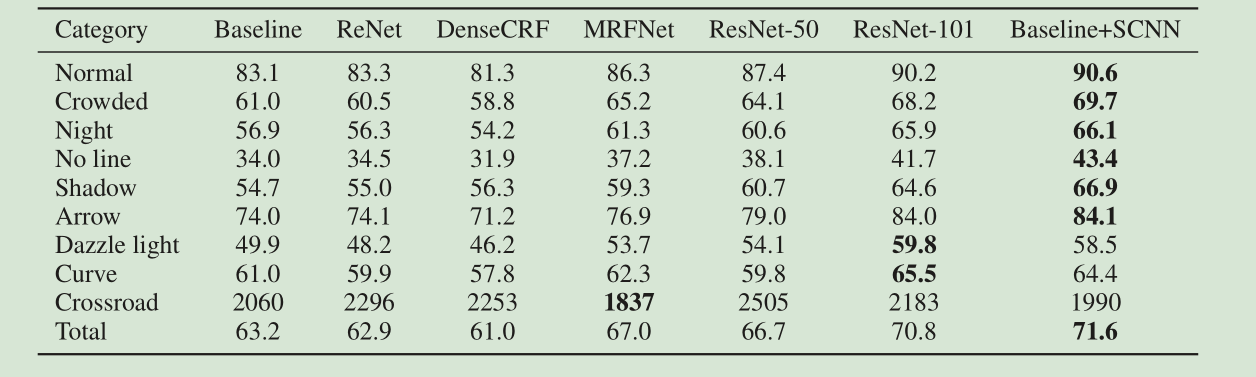


表5. 与其他方法比较，IoU阈值= 0.5。 对于十字路口，仅显示FP

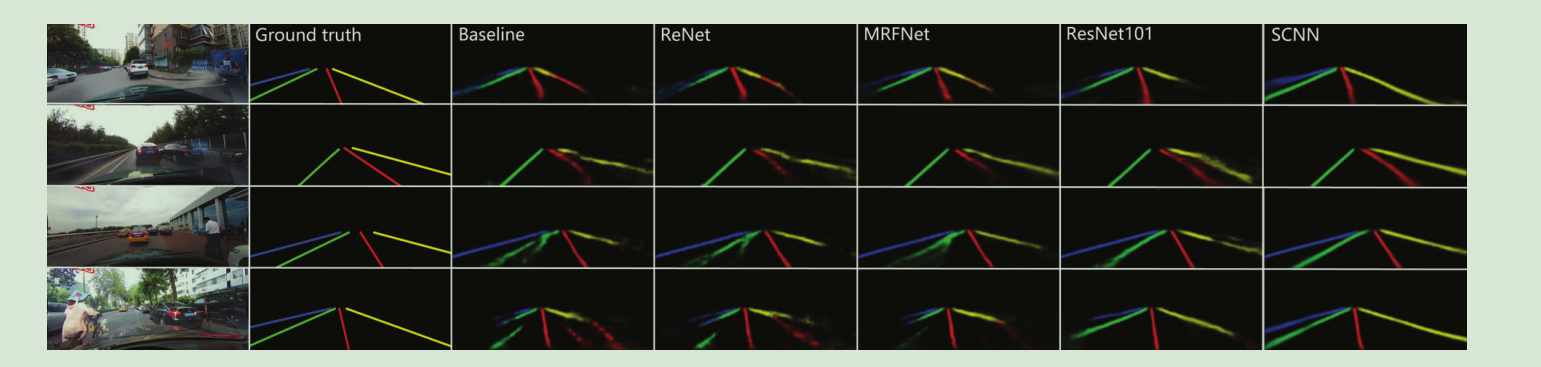


图7. 基线，ReNet，MRFNet，ResNet-101和SCNN的问题清单之间的比较。

为了进一步的验证SCNN在监测车道的有效性。我们与几个最先进的方法进行了比较：基于rnn的ReNet（Visin等人2015），基于MRF的MRFNet，DenseCRF（Krähenbühl和Koltun 2011），以及非常深的残留网络（He et al.2016）对于基于LSTM的ReNet，我们用两个ReNet层替换图3中的“SCNN”层：一层传递水平信息，另一层传递垂直信息。对于DenseCRF，我们使用密集CRF作为后处理，并采用10个均值场迭代，如（Chen et al.2017）。对于MRFNet，我们使用图3（a）中的实现，迭代时间和消息传递内核大小分别设置为10和20 MRF与CRF的主要区别在于，在训练期间学习消息传递内核的权重而不是依赖于图像。对于ResNet，我们的实现与（Chen et al.2017）相同，只是我们不使用ASPP模块。对于SCNN，我们将SCNN DULR模块添加到基线，并且内核宽度w为9.不同场景的测试结果如表5所示，可视化在图7中给出。

从结果中我们可以看出，ReNet的性能甚至与SCNN DULR的w = 1无法比较，表明我们的剩余消息传递方案的有效性。有趣的是，DenseCRF在这里导致更糟糕的结果，因为车道标记通常具有较少的外观线索，因此密集的CRF无法区分车道标记和背景。相比之下，从数据中学习核心权重，MRFNet可以在一定程度上平滑结果并改善性能，如图7所示，但仍然不是很令人满意。此外，我们的方法甚至优于更深层次的ResNet-50和ResNet-101。尽管ResNet-101有超过一百层和非常大的感受野，但在具有挑战性的情况下它仍会产生混乱或不连续的输出，而我们的方法只有16个卷积层加上4个SCNN层，可以保持通道的平滑性和连续性。 线条更好。这表明SCNN在传统CNN上捕获对象之前的结构的能力强得多。

1. 计算效率高于其他方法。

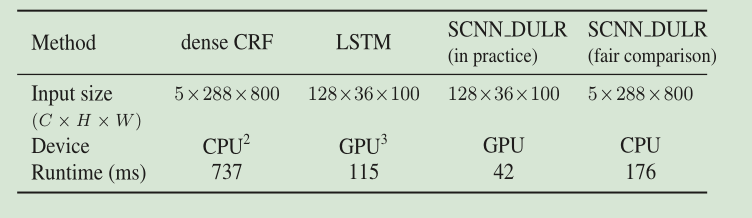


表6

在分析部分，我们对SCNN在密集CRF上的计算效率进行了理论分析。 为验证这一点，我们通过实验比较它们的运行时间。结果如表6所示，其中还给出了ReNet中LSTM的运行时间。这里运行时不包括骨干网络的运行时。对于SCNN，我们使用与密集CRF相同的设置测试实际案例和案例。在实际情况中，SCNN应用于顶层隐藏层，因此输入具有更多通道但较少高度和宽度。在公平比较的情况下，输入大小被修改为与密集CRF中的大小相同，并且两种方法都在CPU上进行测试。结果表明，即使在案例的公平比较中，SCNN也比密集CRF快4倍，尽管在（Krähenbühl和Koltun 2011）中有效实施了密集的CRF。这是因为SCNN显着减少了消息传递中的冗余，如图4所示。此外，SCNN比LSTM更有效，LSTM的门机制需要更多的计算。

1. 城市景观的语义分割

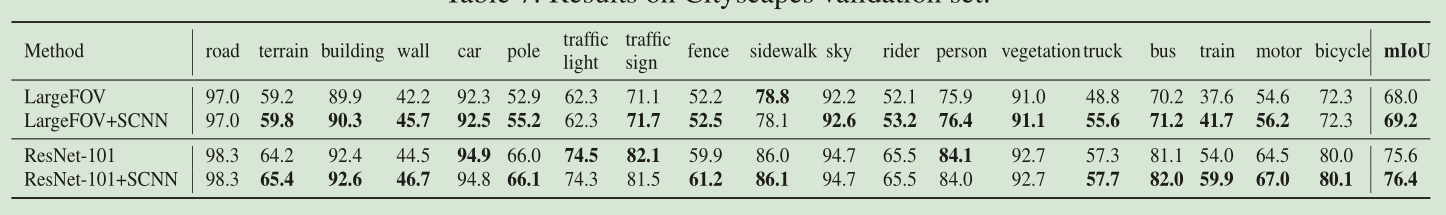


表7. Cityscapes验证集的结果。

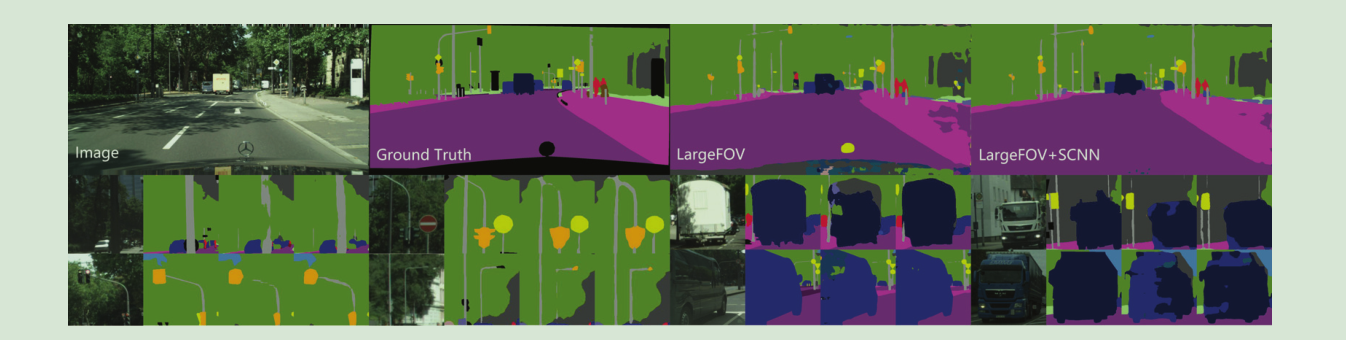


图8 Cityscapes验证集的视觉改进，对于每个示例，从左到右分别是：输入图像，基础事实，LargeFOV的结果，LargeFOV + SCNN的结果。

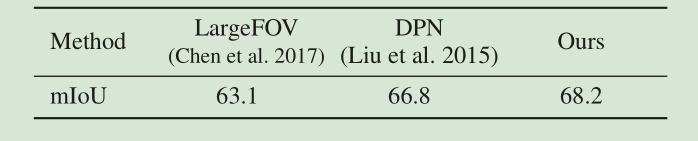


表8我们的SCNN和其他基于MRF / CRF的Cityscapes测试集方法之间的比较。

为了证明我们方法的一般性，我们还评估了Cityscapes上的空间CNN（Cordts等人2016）Cityscapes是城市交通场景语义分割的标准基准数据集。它包含5000个精细注释图像，包括用于训练的2975，用于验证的500和用于测试的1525。定义了19个类别，包括东西和物体。我们使用两个经典模型，DeepLab中的LargeFOV和ResNet-101（Chen等人，2017）作为基线。批量标准化层（Ioffe和Szegedy 2015）被添加到LargeFOV以实现更快的收敛。对于这两种型号，顶部隐藏层的通道编号被修改为128以使它们更紧凑。

我们以与车道检测相同的方式将SCNN添加到基线模型中。表7中显示了基线与结合内核宽度w = 9的SCNN DURL模型的基线之间的比较。可以看出，SCNN还可以改善语义分割结果。随着SCNN的增加，所有级别的IoU至少与基线相当，而“墙”，“极”，“卡车”，“总线”，“火车”和“电机”类别实现了显着的改进。这是因为对于像火车和杆这样的长形物体，SCNN可以捕获其连续结构并连接断开的部分，如图8所示。对于可能占据大图像区域的墙壁，卡车和公共汽车，SCNN的扩散效应可以根据上下文纠正错误分类的部分。这表明SCNN不仅适用于长薄结构，而且适用于需要正确分类全局信息的大型对象。另一个有趣的现象是，在训练期间忽略标签的图像底部的车辆头部在LargeFOV中处于混乱状态，而在添加SCNN时，它被归类为道路。这也是由于SCNN的扩散效应，其将道路信息传递到车辆头部区域。为了将我们的方法与其他基于MRF / CRF的方法进行比较，我们在Cityscapes测试集上评估LargeFOV + SCNN，并与也使用VGG16（Simonyan和Zisserman 2015）作为骨干网络的方法进行比较。结果如表8所示。这里LargeFOV，DPN和我们的方法分别使用密集CRF，密集MRF和SCNN，并且共享几乎相同的基本CNN部分。结果表明，我们的方法实现了显着更好的性能。

1. 结论

在本文中，我们提出空间CNN，一种类似CNN的方案，以实现空间级别的有效信息传播。SCNN可以很容易地融入深度神经网络并进行端到端训练。它在交通场景理解中的两个任务进行评估：车道检测和语义分割。结果表明，SCNN可以有效地保持长薄结构的连续性，而在语义分割中，其扩散效应也被证明有利于大型物体。具体来说，通过将SCNN引入LargeFOV模型，我们的20层网络优于ReNet，MRF和非常深的ResNet-101进行车道检测。最后但同样重要的是，我们认为我们提出的大型挑战车道检测数据集将推动自动驾驶的研究。

1. 致谢

这项工作得到了SenseTime Group Limited的支持。我们要感谢Xiahang Zhan，Jun Li和Xudong Cao在构建车道检测数据集方面所做的有益工作。