文章编号：2095-2163(2020)06-0042-06 中图分类号：TP391 文献标志码：A

基于稠密连接网络的单目深度估计

张顺然，吴克伟，洪 炎
（合肥工业大学 计算机与信息学院，合肥 230601）

摘 要：单目深度估计作为计算机视觉的基本问题，得到人们的广泛关注。目前的方法多集中在深度卷积神经网络的图像级信息上，训练时收敛速度较慢，精度下降，特别是在图像中拥有不同大小的多目标情况下。为此，本文基于一个编解码框架提出了一个新的卷积神经网络模型结构 DCDN（Deep Convolution DenseASPP Network），并将其应用到深度估计中。不同尺度的物体特征需要不同的卷积核去获取，对于多目标的图像，用不同的卷积核去获取他们的特性。本文采用稠密链接的空洞卷积组，利用不同扩张率的空洞卷积去强化多尺度目标的特性学习。实验结果表明，该方法在 NYU-Depth-v2 数据集上达到了 0.823 的准确率（阈值< 1.25），优于最先进的方法。

关键词：深度估计；卷积神经网络；空洞卷积；多尺度

Monocular depth estimation based on dense connected network

ZHANG Shunran, WU Kewei, HONG Yan
（School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China）

【Abstract】 As a basic problem of computer vision, monocular depth estimation has been widely concerned. At present, most methods focus on the image-level information of deep convolutional neural network, and the convergence speed is slow and the accuracy drops, especially in the case of multi-objects with different sizes in the image. For this reason, we propose a new DCDN (Deep Convolution DenseASPP Network) model structure based on a codec framework and apply it to depth estimation. We believe that different convolution kernels are needed to obtain the features of objects of different scales. For some multi-object images, different convolution kernels should be used to obtain their characteristics. In this paper, dense linked dilated convolution groups are used to enhance the characteristic learning of multi-scale targets by using the dilated convolution with different dilation rates. The experimental results show that our method achieves the accuracy of 0.823 (threshold < 1.25) on NYU-Depth-V2 data set, which is better than the most advanced method.

【Key words】 Depth Prediction; Convolutional Neural Network; Dilated Convolution; Multi-scale

0 引 言

单目深度估计作为计算机视觉领域中一个基本问题，可以从一个简单 RGB 图像中估计场景深度信息。获得的深度信息为推动其他任务的发展提供了重要的线索，如语义分割[1]，三维重建[2]，人体姿态估计[3]，目标检测[4]，和即时定位与地图构建[5]。随着深度传感技术的不断发展，构建出更具精确性的 RGBD 数据集，拓展了深度估计的研究领域。但是，由于图像场景的复杂性，场景物体的多样性和大量的视觉干扰等因素，使得对深度估计的研究仍具有挑战性。

随着深度学习的发展，深度神经网络通过端到端多层连接的方式，集成不同尺寸、不同级别的特征来学习新的深度估计模式。深度神经网络中不同层次的特征存在于不同的下采样层，通过卷积和池化操作来学习。然而，由于连续的卷积操作，使得特征的分辨率不断降低。在不断的池化过程中，这些需要被学习的特征在生成的过程中又不断的丢失所含的信息，这就导致了在接下来的训练中很难获得更好的结果。针对这个问题，深度估计网络引入了空洞卷积，来产生具有更大接受域的特征，同时不牺牲空间分辨率。而在众多网络结构中，具有多个空洞卷积[6]的 ASPP[7]（Atrous Spatial Pyramid Pooling），在某种程度上有效的缓解了这一问题，该网络结构将多个具有不同扩张率的空洞卷积的输出特征连接为最终的特征表示，从而汇集具有不同尺度的优势的特征。虽然，ASPP 能够生成多尺度特征，但是尺度轴上的分辨率并不足以满足更多复杂场景深度估计的需求。为此，在 ASPP 的基础上衍生了更为有效的 DenseASPP[8]（Dense Atrous Spatial Pyramid Pooling）网络结构。DenseASPP 同样使用不同扩张率的空洞卷积，但在特征汇集方式上不同于 ASPP，

作者简介：张顺然（1993-），男，硕士研究生，主要研究方向：计算机视觉；吴克伟（1984-），男，博士，副教授，主要研究方向：计算机视觉；洪炎（1996-），女，硕士研究生，主要研究方向：知识图谱。

收稿日期：2020-03-12

采用了类似 Densenet[9] 的密集连接的方式，使得不同尺度的特征重用性更强，从而提高不同尺度物体的深度估计精度。

目前常见的深度学习模型主要是采用 ASPP 结构，促使发掘出一种采用 DenseASPP 模块的网络模型来进行深度估计任务。在这种模型训练中，不仅可以获得多尺度的信息，还可以覆盖更大尺度范围，没有显著增加模型的大小，而且可以获得更好的估计结果。其中本文的主要贡献如下：

（1）提出了一种结合 DesenASPP 网络结构的新型网络模型，解决了目前大多神经网络模型对多尺度场景进行深度预测过程中的不同尺度分辨率不足和对不同尺寸物体估计精度不足的问题。

（2）将 DenseASPP 结构融入并构造出不同的网络模型，以发现 DenseASPP 结构和不同网络结构的相适性，探寻出最佳的网络模型结构。

（3）提出的网络模型经过实验证明，可以有效的提高含有多尺度物体图像的深度估计精度，并在 NYU-V2 数据集上进行测试，效果比当前最佳方法高 1%。

1　场景深度估计现状

在过去十年中，单目深度估计问题引起了广泛的关注。尽管早期的方法主要基于手工制作的特征[10]，但是受到深度学习在各种视觉任务中成功的启发，深度卷积神经网络已经广泛应用于各主流的深度估计方法中。对于单幅图像的单目场景深度估计，其神经网络模型对特征的处理方法起着至关重要的作用，而不同的网络具有各自的特点。

最早期的出现的深度估计网络模型为全卷积无多尺度的深度模型。Shir Gur 等人在一项开创性的工作中，通过应用新的点扩散函数卷基层，根据每个图片产生特定的内核去学习深度，并取得很好的效果。在卷积神经网络对场景估计深度描绘的基础上，提高预测深度图质量的另一个方向是将卷积神经网络与条件随机场相结合。Liu 等人使用 CNN 和 CRF 简单融合的方式获得深度估计网络模型；Lee 等人提出了一种新的损失函数，并对单图像深度估计问题进行傅里叶分析的研究；Yan 等人利用 CRF 模型加入物体表面法向量的约束条件，在超像素级和像素级上估计多层次场景的深度；Liu 等人在超像素基础上，使用卷积神经网络提取场景深度特征，构建像素池化的 CRF 模型，来进行场景深度估计；Wang 等人在全局布局指导下，将图像分解为局部区域，以卷积神经网络为基础，构建层次 CRF 模型，进行场景深度和语义预测；Li 等人进行深度估计的同时，添加了曲面法线的约束条件，对深度图进行细化；Laina 等人提出并采用一种上采样策略，通过新的上采样方式来增加还原分辨率时的精度；Cao 等人在使用全卷积残差网络的基础上，把深度估计视为一个像素级的分类任务。

但在应用全卷积网络进行深度估计时，会出现物体边缘混淆、层次不明显、目标深度与背景融合等问题。在处理图像时，可以利用图像分割获得多个场景块，利用分割信息优化估计结果。在处理视频的深度估计中，引入了光流线索，并结合多个线索来估计图像的深度。Yang Wang 等人使用 UnOS（Unsupervised optical - flow and Stereo - Depth Estimation）来估计图像深度，这是一种针对光流线索的无监督三维深度估计模型。Eddy Ilg 等人提出光流线索结合遮挡、视差，光流多线索联合估计场景深度。

随着对多尺度信息关注的增多和 ASPP 模块的提出，由于 ASPP 采用不同扩张率的空洞卷积来解决信息提取中不同尺度特征信息问题极具优势，使得带有多尺度模块的全卷积深度模型成为深度估计领域的主流。其中，Xu 等人构建深度估计模型，采用多尺度策略，并对每层多尺度进行 CRF 平滑操作；Eigen 等人提出了一种多尺度的神经网络结构来获取多尺度信息，从而精确多尺度物体的深度估计结果；Lee 等人结合多尺度相对图与分解-融合策略进行深度估计任务；Chen，L.C 等人构建模型，并使用 ASPP 采取多尺度特征集合；Fu 等人使用 ASPP 和跨通道的方法学习多尺度信息，使用逐像素有序回归损失函数完成深度估计任务。

然而，在全卷积的多尺度深度模型中，ASPP 的局限性阻碍了深度估计精度的提升。当增加网络深度时，还存在严重的梯度退化问题。分析了一些解决办法，深度卷积模型的密集卷积模型 DenseNet 解决了消失梯度问题，在此基础上还加强了信息传递，鼓励特征重用，大大减少了训练参数的规模；Gao Huang 等人引入了稠密卷积网络（DenseNet），该网络以前馈方式将每一层与每一层连接起来；Fu 等利用多尺度级联 CNN 将深度估计的 RCCN 模型训练为离散深度分类任务；Simon Jégou 等人将 DenseNet 的密集级联网络向下采样、向上采样应用于分割任务；Maoke Yang 等人将 ASPP 与稠密级联方法结合，构建出 DenseASPP 模块，利用该模块进行多尺度采样，可以获取更多的特征信息。

通过上述试验分析,带多尺度的全卷积网络可以获得比不带多尺度的全卷积网络有更好的效果,ASPP 和 DenseNet 模块能有效的获取不同优势的特征信息,DenseNet 可解决梯度消失问题。现有方法虽然已经研究了很多方法结合的网络结构,但是并没有探究 ASPP 和 DenseNet 结合网络 DenseASPP 来完成深度估计任务,受此问题启发,本文提出一种采用 DenseASPP 结构的网络模型 DCDN(Deep Convolution DenseASPP Network)。

2 模型框架

从单目 RGB 图像中预测深度的 DCDN 模型结构,如图 1 所示。模型框架主要有两个部分组成(编码器和解码器),其中编码器部分(如图 1 中(a)部分)主要功能为特征提取;而解码器部分(如图 1 中(b)部分)主要功能为逐步回复空间信息,得到与 RGB 图像等大的深度估计结果图。在基础模型中扩展了空洞卷积的使用,通过应用不同扩张率的空洞卷积,对图像级特征进行多尺度的卷积特征探测,引入 DenseASPP 处理特征,得到更为精确的多尺度信息。

2.1 编解码器

采用含有空洞卷积的主干网络 ResNet-101 来提取图片的基础特征作为基础特征,将基础特征作为解码器的一类输入。另外,在编码器部分增加 DenseASPP 模块,利用图像级特征,通过不同速率的卷积对卷积特征进行多尺度的探测,在 DenseASPP 模块中的 Bottleneck 为多个卷积、池化组合。使用 DensASPP 的最后一个特征图作为编码器-解码器结构中的编码器另一类输出,编码器输出特征图包含 256 个通道和丰富的信息。

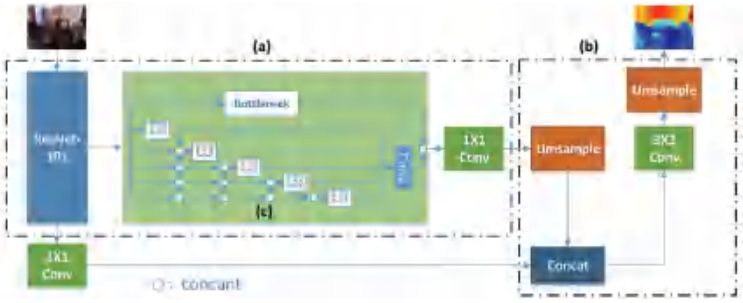

图 1 模型的整体框架

Fig. 1 The overall framework of the model

编码器特性通常用输出步长为 16 来计算。Upsample 将特征向上采样 16 倍,可以认为是一个朴素的解码器模块。而这个简单的解码器模块可能无法成功地恢复深度估计细节。因此,采用一个简单而有效的解码器模块(如图 1 中(b)部分)。编码器的特征首先被上采样 4 倍,然后与来自具有相同空间分辨率的网络主干的相应一类输入特征连接起来。应用另一个 1×1 卷积来减少通道数,因为相应的低级特征通常包含大量的通道(例如,256 或 512)。将上述的两类输入进行连接操作,应用几个 3×3 的卷积来细化连接得到的特征,再进行一个简单的双线性上采样,采样倍数为 4。编码器模块使用输出步长为 16 在速度和精度之间取得了最好的平衡。当编码器模块使用输出步长为 8 时,性能略有提高,但代价是增加了额外的计算复杂度。

本文设计了 3 种不同编码器构成的网络模型结构,(1)下采样采用 Resnet 提取特征的 RD 网络模型,如图 2(a)所示。(2)下采样采用 Densenet 的 DD 网络模型,如图 2(b)所示。

(a) RD 网络结构图

(b) DD 网络结构图

图 2 不同下采样结构网络示意图

Fig. 2 The network structure of the different down-sampling

2.2 空洞卷积

深度卷积神经网络(Deep Convolutional Neural Networks, DCNNs)以全卷积方式部署,在深度估计和分割任务中表现出良好的效果。然而,在这些网络的连续层上重复组合最大池化和步长会显著降低最终得到的特征图的空间分辨率,在最近的 DCNNs 中,通常每个方向的分辨率会降低 32 倍。空洞卷积是一个强大的工具,它允许显式地控制深度卷积神经网络计算的特征的分辨率,在不改变特征图分辨率的情况下,利用空洞卷积来增加接受域。通过控制深度卷积神经网络计算特征的分辨率,调整接受域来获取多尺度信息。在一维情况下,考虑二维信号,对于输出特征图上的每个位置 i,用 $y[i]$ 表示输出信号,用 $x[i]$ 表示输入信号,空洞卷积操作可以表示为公式(1):

$$y[i] = \sum_{k=1}^{K} x[i + d \cdot k] \cdot w[k]. \tag{1}$$

其中,d 为扩张率,决定采样输入的步长,$w[k]$ 为滤波器的第 k 个参数,K 为滤波器的大小。当 $d = 1$ 时,方程简化为标准卷积。空洞卷积相当于将输入 x 与上采样滤波器进行卷积,上采样滤波器是在两个连续的滤波器值之间插入 $d - 1$ 个 0 产生的。因此,更大的扩张率意味着更大的接受域。而对于一个空洞卷积层,它的接受域大小 R 可以通过公式(2)计算:

$$R = (d - 1) \times (K - 1) + K. \tag{2}$$

在不同的场景中,对象通常有很多不同的大小。为了处理这种情况,所得的特征图必须能够覆盖不同规模的接受域。针对这问题,有两种策略可以应对,即使用不同扩张率空洞卷积的并行或级联的多层空洞卷积层。在级联模式下,由于上一层接受下一层的输出,可以有效的产生大量的接受域。在并行模式下,由于多个空洞卷积层接受相同的输入,并且它们的输出连接在一起,因此得到的输出实际上是输入的一个采样,具有不同的接受域尺度。这种并行模式的正式名称为 ASPP,(Atrous Spatial Pyramid Pooling)。

2.3 DenseASPP

DenseASPP 的结构如图 1 中的(c)部分所示。与 ASPP 相比,DenseASPP 形成了一个更为密集的特征金字塔。这个结构不仅可以得到更好的尺度多样性,其中卷积操作也会涉及更多的像素。这主要是由于 DenseASPP 中不同扩张率和不同层位空洞卷积的密集卷积结果。DenseASPP 将所有的空洞卷积层堆叠在一起,并将它们紧密地连接在一起,以获得更大的接受域和更多尺度的特征图。在这个过程中,卷基层的连接会促使接受域增大。例如:有两个卷积核,它们拥有不同的大小(K_1 和 K_2),则叠加之后的新的接受域为公式(3):

$$K = K_1 + K_1 - 1. \tag{3}$$

对于 DenseASPP 的结构,它是以带有不同扩张率的空洞卷积层以级联方式组织,每一层的扩张率逐层增加(3,6,12,18,24)。输入来自 ResNet 的 2048 维特征图,每一层的输出是一个 64 维和等大小的特性。将前层的输入和输出与后层的输入联合起来为下一层的输入。而 DenseASPP 最后结果是由具有多尺度和不同扩张率的卷积核进行卷积操作得到的新特征。为了可以直观的表达 DenseASPP 中各层的功能,给出表达式(4)如下:

$$y_l = H_{K,d_l}([y_{l-1}, y_{l-2}, \cdots, y_0]). \tag{4}$$

式中,d_l 为第 l 层的扩张率,$[\cdots]$ 为级联操作。$[y_{l-1}, y_{l-2}, \cdots, y_0]$ 表示将所有前一层的输出连接起来形成的特征图。比较原始的 ASPP,DenseASPP 堆叠所有的空洞卷积,以一个密集连接的方式将它们联合起来。这种变化带来两个好处:增大接受域和增强多尺度信息的汇集。

2.4 损失函数

对于回归类型的任务,一个常见且默认的损失函数选择是均方误差(L_2)。L_2 对训练数据中的异常值很敏感,因为对较大的错误惩罚更重。其极大降低了估计值 $\tilde{y}$ 和真实值 y 之间的欧几里德范数的平方,公式(5):

$$L_2(\tilde{y} - y) = \| \tilde{y} - y \|_2^2. \tag{5}$$

在实验中,使用 L_2 为训练的默认损失函数。

3 实验

为了证明深度估计模型的有效性,提供了具有挑战性的户外数据集 NYU Depth v2 的实验结果,以与最先进的作品进行比较。

3.1 NYU Depth V2 数据集

在室内场景数据集之一的 NYU Depth v2 上进行评估,原始数据集由 464 个场景构成,由 Microsoft Kinect 捕捉,其中数据集中 249 个被用于训练,215 个被用于测试。在实验中,考虑了 1449 对 RGB-D 的子集,其中 795 对用于训练,其余用于测试。特别地,实验还进行了数据扩充(缩放,旋转,颜色变换,翻转)来扩大训练集,共得到了 87 000 对训练数据。利用下采样操作将原始图像的像素由 640×480 变换为原来一半,经过中心裁剪为 304×228 像素,图

像输入到网络中进行训练。

3.2 实现细节

实验使用开放的深度学习框架 Pytorch 框架，实验工作站配置为 CoreX i7-6800k 6 核，3.4GHz CPU，两块 NVIDIA 1080 8GB 显卡和 16GB 内存。作为特征提取的编码器，使用了 ResNet-101，并使用 ILSVRC 数据集对经过预处理的权值进行分类[11]。随机梯度下降(SGD)算法开始时学习率为 0.001，每隔 6 个 epoch 下降 10 倍。进行 24 个 epoch 的训练。动量衰减和重量衰减分别设置为 0.9 和 0.000 5。批量大小的数量设置为 8。在 NYU Depth v2 数据集上训练网络(DCDN)花费了大约 14 个小时。为了避免过拟合，在输入到网络之前，使用随机水平翻转、随机对比度、亮度和颜色调整[0.6，1.4]的范围内增加图像，其中每个操作有 50%的几率实施。还使用了在弧度[-5，5]的范围中，对输入进行随机旋转的方法。在规模为 304×228 的纽约大学深度 V2 数据集的随机作物上训练网络。

3.3 评价标准

对于定量评估，使用以下错误度量得到的错误，这些错误已经被广泛使用。g 为地面真值深度，$\tilde{g}$ 为估计深度，T 为图像中所有点的集合。本文使用评价标准[12]体包括包括：

平均相对误差(rel)：$\frac{1}{|T|}\sum_{\tilde{g}\in|T|}\frac{|\tilde{g}-g|}{d}$.

根均方误差(rmse)：$\sqrt{\frac{1}{|T|}\sum_{\tilde{g}\in|T|}\|\tilde{g}-g\|^2}$.

对数误差(log10)：$\frac{1}{|T|}\sum_{\tilde{g}\in|T|}|\log_{10}(g)-\log_{10}(\tilde{g})|$.

特别地，

阈值精度：

$$\delta_i=\frac{card\left(\left\{\tilde{g}_i:\max\left\{\frac{\tilde{g}_i}{g_i},\frac{g_i}{\tilde{g}_i}\right\}<1.25^i\right\}\right)}{card(\{g_i\})}. \quad (6)$$

其中，$card$ 是集合的基数，而 δ_i 的值越高，则表示结果越好，本文中阈值参数设置为 $\delta_1<1.25$，$\delta_2<1.25^2$，$\delta_3<1.25^3$。

3.4 实验结果及分析

3.4.1 对比方法

本文所述实验的训练是基于整幅图像对网络进行有监督的指导，对图像的每个像素点进行回归处理。将对 NYU Depth v2 数据集评估本文的方法。为验证所提框架的有效性，将其与现有的方法进行比较，并对所提出的方法展开分析，对其精度和计算效率进行评价。

表 1 显示了本文模型方法与当前最佳方法就 NYU Depth V2 数据集比较的结果。从表中可以看出，与最先进的方法相比，DCDN 网络在阈值精度和 rmse 方面获得了更好的深度估计性能。Xu 等人在 log10 和 AbsRel 拥有更好的结果，这主要是因为其使用了连续条件随机场优化，对整体的场景分布会有更好的效果，而不是反应局部区域的特性，对小目标的估计参考价值有限。本文模型在 NYUv2 数据集上的准确率达到了 82.3%(阈值小于 1.25)，优于最先进的方法。采用带空洞卷积的残差网络做为主干网络，在训练过程中为浅层网络提供更多的信息，学习多种模式，并与密集连接方式的 DenseASPP 相结合，使深度网络具有更多的特性来弥补过程的损耗。同时，该方法也有效地缓解了梯度消失问题，获得了更好的效果。

表 1 不同方法识别性能对比

Tab. 1 Comparison of recognition performance of different methods

对比方法	精度			误差		
	δ_1	δ_2	δ_3	AbsRel	RMSE	log10
Saxena et al.	0.447	0.745	0.897	0.349	1.214	-
Wang et al.	0.605	0.890	0.970	0.220	0.824	-
Liu et al.	0.650	0.906	0.976	0.213	0.759	0.087
Eigen et al.	0.769	0.950	0.988	0.158	0.641	-
Li et al.	0.789	0.955	0.988	0.152	0.611	0.064
Laina et al.	0.811	0.953	0.988	0.127	0.573	0.055
Xu et al.	0.811	0.954	0.987	**0.121**	0.586	**0.052**
Lee at al.	0.815	0.963	0.991	0.139	0.572	-
Cao et al.	0.819	**0.965**	0.992	0.141	0.573	0.055
Ours	**0.823**	0.964	**0.994**	0.139	**0.474**	0.057

3.4.2　消融实验

分析消融实验的结果见表 2。分析表 2 发现：(1)本文提出的关于深度估计的网络模块在多个基础网络上进行实验，试验效果较好的 RD(下采样模块采用 Resnet)和 DD(下采样模块 Densenet)，对其进行相同的上采样操作，在相同的位置添加 DenseASPP 模块，其实验结果对比 DCDN 具有更为优异的效果。(2)对添加 DenseASPP 是否有效果进行了实验。对比的消融实验结果如表 2 中 DeeplabV3 和 DCDN 所示。DeeplabV3 是空白对照组，对比具有 ASPP 的 DeeplabV3，从表 2 中数据可以得知，添加 DenseASPP 的结构比 ASPP 略高一个百分比，效果对比如图 3 所示。说明本文网络模型的具有优异性。

表 2　消融实验结果

Tab. 2　The results of Ablation Study

方法	精度			误差		
	$\delta_1 < 1.25$	$\delta_2 < 1.25^2$	$\delta_3 < 1.25^3$	AbsRel	RMSE	log10
(RD)	0.776	0.952	0.988	0.158	0.567	0.067
(DD)	0.798	0.952	0.989	0.149	0.562	0.064
Deeplabv3	0.813	0.962	0.993	0.149	0.477	0.060
DCDN	0.823	0.964	0.994	0.139	0.474	0.057

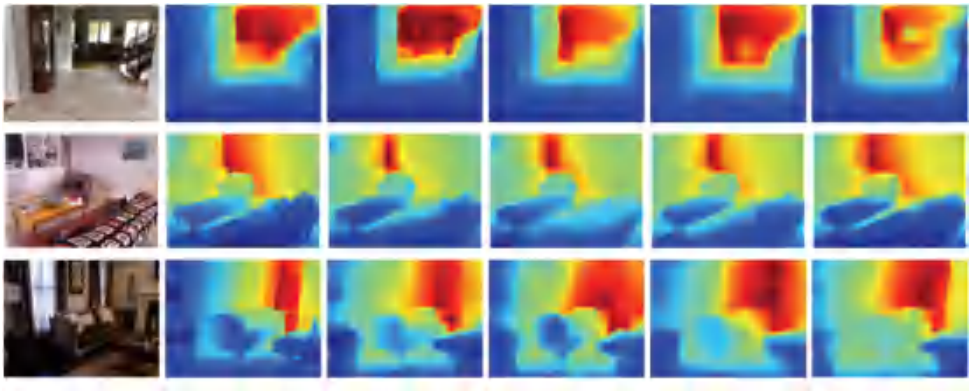

图 3　消融实验效果图

Fig. 3　The figures of ablation studies

4　结束语

针对目前大多神经网络模型对多尺度场景进行深度预测过程中的不同尺度分辨率不足和对不同尺寸物体估计精度不足的问题，本文提出一种结合 DenseASPP 的新型网络模型。通过一系列的消融实验，展示出了该模型的优异性和合理性。对比基础网络，模型不仅仅覆盖了更多的尺度范围，获得更多的多尺度信息，而且通过连接不同尺度下的诸多特征，使得不同尺度特征有更高的重用性。通过利用 NYU-Depth V2 公认数据集对该模型进行验证，实验结果表明其高于现有的大部分深度估计方法。

参考文献

[1] LIU B, GOULD S, KOLLER D. Single image depth estimation from predicted semantic labels[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 1253-1260.

[2] KELLER M, LEFLOCH D, LAMBERS M, et al. Real-time 3d reconstruction in dynamic scenes using point-based fusion[C]// 2013 International Conference on 3D Vision-3DV 2013. IEEE, 2013: 1-8.

[3] BORGHI G, VENTURELLI M, VEZZANI R, et al. Poseidon: Face-from-depth for driver pose estimation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4661-4670.

[4] CHENG Y, ZHAO X, HUANG K, et al. Semi-supervised learning and feature evaluation for RGB-D object recognition[J]. Computer Vision and Image Understanding, 2015, 139: 149-160.

[5] TATENO K, TOMBARI F, LAINA I, et al. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. arXiv 2017[J]. arXiv preprint arXiv:1704.03489.

[6] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.

[7] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

(下转第 50 页)