计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(11):365−372 [doi: 10.15888/j.cnki.csa.008790]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合残差上采样结构的 P²Net 无监督单目深度估计①

刘安旭, 黎向锋, 刘晋川, 赵　康, 李高扬, 左敦稳

(南京航空航天大学 机电学院, 南京 210016)
通信作者: 黎向锋, E-mail: fxli@nuaa.edu.cn

摘　要: 单目深度估计是计算机视觉领域中的一个基本问题, 面片匹配与平面正则化网络 (P²Net) 是现阶段最先进的无监督单目深度估计方法之一. 由于 P²Net 中深度预测网络所采用的上采样方法为计算过程较为简单的最近邻插值算法, 使得预测深度图的生成质量较差. 因此, 本文基于多种上采样算法构建出残差上采样结构来替换原网络中的上采样层, 以获取更多特征信息, 提高物体结构的完整性. 在 NYU-Depth V2 数据集上的实验结果表明, 基于反卷积算法、双线性插值算法和像素重组算法的改进 P²Net 网络相较原网络在均方根误差 *RMSE* 指标上分别降低了 2.25%、2.73% 和 3.05%. 本文的残差上采样结构提高了预测深度图的生成质量, 降低了预测误差.

关键词: 深度估计; 无监督; P²Net; 残差上采样结构; 深度学习

引用格式: 刘安旭,黎向锋,刘晋川,赵康,李高扬,左敦稳.融合残差上采样结构的 P²Net 无监督单目深度估计.计算机系统应用,2022,31(11):365–372. http://www.c-s-a.org.cn/1003-3254/8790.html

Unsupervised Monocular Depth Estimation with P²Net Incorporating Residual Upsampling Structure

LIU An-Xu, LI Xiang-Feng, LIU Jin-Chuan, ZHAO Kang, LI Gao-Yang, ZUO Dun-Wen

(College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Monocular depth estimation is a fundamental problem in computer vision, and the patch-match and plane-regularization network (P²Net) is one of the most advanced unsupervised monocular depth estimation methods. As the nearest neighbor interpolation algorithm, the upsampling method adopted by the depth prediction network of P²Net, has a relatively simple calculation process, the predicted depth maps have a poor generation quality. Therefore, the residual upsampling structure based on multiple upsampling algorithms is constructed in this study to replace the upsampling layer of the original network for more feature information and higher integrity of the object structure. The experimental results on the NYU-Depth V2 dataset reveal that compared with the original network, the improved P²Net based on the transposed convolution, bilinear interpolation, and PixelShuffle can reduce the root mean square error (RMSE) by 2.25%, 2.73%, and 3.05%, respectively. The residual upsampling structure in this study improves the generation quality of the predicted depth maps and reduces the prediction error.

Key words: depth estimation; unsupervised; patch-match and plane-regularization network (P²Net); residual upsampling structure; deep learning

深度估计旨在获取输入 RGB 图像对应的深度图像, 深度图中每个像素点的值可用来表示场景中某一点与相机成像平面的距离, 可用于机器人定位导航、自动驾驶和三维重建等领域[1]. 目前虽有一些硬件能够直接得到深度图, 但都有各自缺陷, 如激光雷达设备较为昂贵; 基于结构光的深度摄像头在室外无法使用, 且

① 基金项目: 国家自然科学基金联合基金 (U20A20293)
收稿时间: 2022-02-15; 修改时间: 2022-03-14; 采用时间: 2022-04-12; csa 在线出版时间: 2022-07-15

得到的深度图噪声较多, 需要进一步处理; 双目摄像头需要利用立体匹配算法, 计算量相对较大, 且对于低纹理场景的深度估计效果不好[2]. 与之相比, 单目摄像头成本最低, 设备也最为普及; 为此, 开展单目深度估计研究具有重要的研究意义.

传统单目图像深度估计是使用目标场景中存在的多种视觉线索来获取深度信息, 主要包括: 从明暗度变化的规律恢复形状 (shape from shading, SFS)[3]、从运动中恢复形状 (shape from motion, SFM)[4]、从对焦获取深度 (depth from focus, DFF) 及从离焦获取深度 (depth from defocus, DFD)[5] 等. 通过这些视觉线索, 可以初步推断出图像中的整体构成和视差等深度相关信息, 进而结合相机参数计算出对应的深度值. 但这些方法都需要额外的辅助设备或对应用场景有着特殊的要求, 且预测结果很容易受到环境因素的影响, 使用条件苛刻, 难以推广. 近年来, 得益于卷积神经网络强大的特征提取能力和硬件设备不断提升的计算性能, 使用深度学习的方法预测图像深度信息逐渐成为主流研究方向[6].

根据训练数据集中是否带有深度标签, 即输入 RGB 图像对应的真实深度图, 可分为有监督学习的单目深度估计和无监督学习的单目深度估计. 其中, 有监督深度估计算法要求输入 RGB 图像都有对应的深度标签, 而深度标签的获取成本非常昂贵且得到的深度图存在一定的误差[7]. 无监督深度估计算法在训练时则不需要输入对应的真实深度图, 极大减轻了对大量深度标签的需求, 逐渐受到研究学者们的重视. Garg 等人[8] 于 2016 年最先提出了一个无监督单目深度估计网络框架, 通过类似于自动编码器的方式来训练网络, 输入为源图像和目标图像构成的图像对. 首先, 使用编码器预测目标图像的深度图, 并计算此图像对的相对位移. 然后解码器使用上一步预测的深度图和相对位移重构目标图像, 最后计算重构目标图像与原目标图像之间的误差来对网络进行监督. Wong 等人[9] 针对目标场景中物体边缘梯度不连续等问题, 提出了一种自适应正则化方案和左右视差间的双边一致性约束来降低物体边缘位置的深度误差, 并使得网络能够处理立体图像对中的共同可见区域和遮挡区域, 表现出了较好的泛化性能. Zhou 等人[10] 基于光场多向对极几何提出了一种无监督单目深度估计网络. 该网络从光场中心位置预测深度, 根据光场内部深度线索和几何约束, 提出了由光度损失、散焦损失和对称损失组成的新损失函数, 并验证了该算法对真实场景光场图像的有效性和通用性. 近期, Zhou 等人[11] 提出了一种基于光流的网络训练方法, 通过提供更清晰的训练目标和单独处理非纹理区域来降低无监督学习的难度, 并使用密集光流来监督深度和位姿的学习, 在室内场景中取得了较好的预测效果. 相比 Zhou 等人采用的方法, Yu 等人[12] 提出的面片匹配与平面正则化网络 (patch-match and plane regularization, P^{2}Net) 使用基于面片的光度损失来监督网络的训练, 并使用平面一致性损失来调整非纹理区域内的深度值, 进一步降低了预测误差. 但是其深度预测网络 DepthNet 所采用的上采样方法为计算过程较为简单的最近邻插值算法, 使得预测深度图的生成质量较差、整体效果不理想.

针对上述问题, 本文对 P^{2}Net 中主要组成模块之一的深度预测网络 DepthNet 进行改进, 基于多种上采样算法构建出残差上采样结构来替换原网络中的上采样层, 以进一步提高预测深度图中物体结构的完整性, 降低预测误差.

1 融合残差上采样结构的 P^{2}Net

1.1 P^{2}Net 整体网络结构介绍

P^{2}Net 整体网络框架如图 1 所示, 主要包含深度预测网络 DepthNet 和位姿预测网络 PoseNet 两个可学习的模块; 其中, DepthNet 用来预测目标图像 I_t 对应的深度图 D_t, PoseNet 用来预测目标图像 I_t 与源图像 I_s (即目标图像的相邻帧) 之间的相对位姿 $T_{t\to s}$. 之后, 使用相对位姿 $T_{t\to s}$、从目标图像 I_t 提取的关键点和深度图 D_t 来计算基于面片的光度损失; 并使用从目标图像 I_t 提取到的超像素来获得对应的分段平面, 再结合深度图 D_t 计算出平面一致性损失; 这两种损失即为 P^{2}Net 损失函数的主要组成部分, 用来对网络的训练过程进行监督.

P^{2}Net 网络提出基于面片的光度损失和平面一致性损失; 前者通过面片匹配法将一个关键点和以该点为中心的局部窗口相结合以提高其辨识能力, 并结合相对位姿 $T_{t\to s}$ 和深度图 D_t 计算该局部窗口的重构误差作为其光度损失, 其中图像关键点通过直接稀疏测距法 (direct sparse odometry, DSO)[13] 获得. 后者使用较大的超像素来代替具有相似属性的同质区域, 并计算超像素预测深度与拟合平面深度的绝对误差作为其平

面一致性损失, 其中超像素是由一些位置相邻、颜色和纹理相似的像素点组成的小区域, 通过 Felzenszwalb 法[14] 获得.

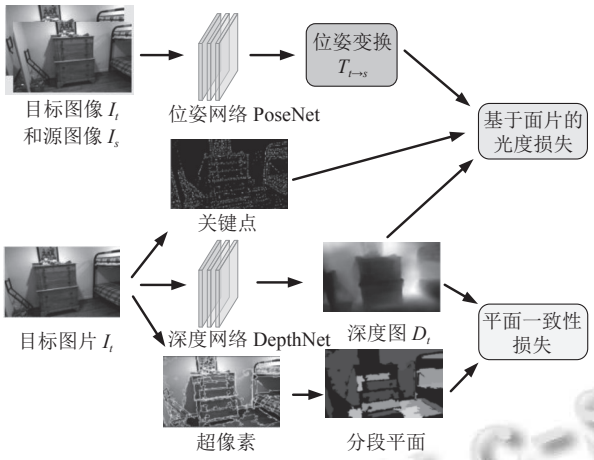


图 1 P²Net 网络框架

1.2 位姿预测网络 PoseNet 介绍

相机在现实三维场景中的运动为六自由度的刚性运动, 因此相机的位姿变化可以使用 6 个变量进行表示, 包括沿 x、y、z 轴的 3 个平移变量 t_x、t_y、t_z 和绕 x、y、z 轴旋转的 3 个欧拉角 r_x、r_y、r_z. PoseNet 的输出即为此 6 个变量, 之后利用这 6 个变量计算出对应的平移向量 T 和旋转矩阵 R, 最后构造出 4×4 的位姿变换矩阵 $T_{t\to s}$[15], 如式 (1) 所示. PoseNet 的详细结构如图 2 所示. 其中, 2×表示输入图像尺寸为当前图像尺寸的 2 倍, 图像块下方数字表示当前图像通道数.

$$T_{t\to s}=\begin{bmatrix}R & T\\ 0 & 1\end{bmatrix} \tag{1}$$

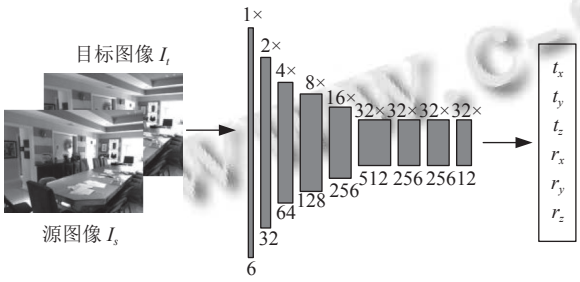


图 2 PoseNet 网络结构

1.3 深度预测网络 DepthNet 介绍

本文 DepthNet 为 U 形结构的网络, 由下采样部分 (Conv)、上采样部分 (Cat_UpProj 和 UpProj) 和连接这两部分的跳跃连接组成, 如图 3 所示. 随着下采样操作的进行, 输出图像尺寸不断缩小, 输出通道数不断增大, 网络模型可以提取到更高级别的语义信息. 上采样操作是对输入图像进行放大的过程, 在上采样的同时, 还融入了跳跃连接传递过来的信息, 如图 3 中虚线箭头所示. 跳跃连接有利于将下采样获得的各种尺度信息在上采样时进行整合, 更好地恢复图像的细节信息. 这样的 U 形结构有助于整个网络很好地“记住”图像的所有信息, 其中包含跳跃连接的 4 层上采样操作均使用 Cat_UpProj 实现, 最后一层不包含跳跃连接的上采样操作使用 UpProj 实现.

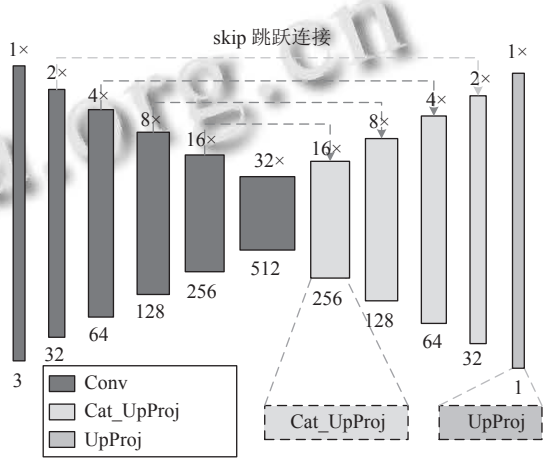


图 3 本文 DepthNet 网络结构

(1) DepthNet 的残差上采样结构

残差网络通过使用跨层的连接来避免误差梯度的消失, 使得前层的参数能够及时得到更新[16]. 如图 4 所示, 图中的 x 代表输入, $F(x)$ 代表经过激活函数后的输出. 普通网络中输出对输入的梯度为$\partial F(x)/\partial x$, 残差网络的梯度为$\partial(F(x)+x)/\partial x=\partial F(x)/\partial x+1$. 随着网络层数的不断增加, $\partial F(x)/\partial x$ 逐渐减小直至衰减为 0, 而残差网络的梯度能够一直维持在 1 附近, 从而避免了梯度消失现象的产生. 在误差梯度逐层反向传播时, 残差网络不仅可以使用普通网络的非线性层逐步传播, 还可以通过自身跨层的连接直接跨层传播, 显著提升了参数的更新效率.

由于原始 DepthNet 使用的上采样结构为类似图 4 (a) 所示的普通结构, 且使用的上采样方法为计算过程较为简单的最近邻插值算法, 使得预测深度图的生成质量较差、整体效果不理想. 而残差结构在加深网络层数提高特征信息获取能力的同时, 能够避免网络退化现象 (即网络模型的精度随着层数的增加逐渐饱和甚至降低) 的产生, 因此本文基于图 4 (b) 的基本残差单元构建出图 5 所示的残差上采样结构, 即 Cat_

UpProj 模块. 该模块中 X 为上采样层的输入特征图, *skip* 为跳跃连接传递的特征图. UpSample 层为上采样方法, 分别使用反卷积算法、双线性插值算法和像素重组算法实现. Conv、BN 和 ReLU 分别为卷积操作、批标准化和激活函数, 这 3 部分组成一个小的特征提取单元. 批标准化 BN 层对数据的分布进行修正, 能够在一定程度上缓解过拟合问题. 激活函数 ReLU 层能够增加网络的非线性, 降低计算量. 图 5 中右侧的 UpProj 模块用于图 3 中最后一层不包含跳跃连接的上采样.

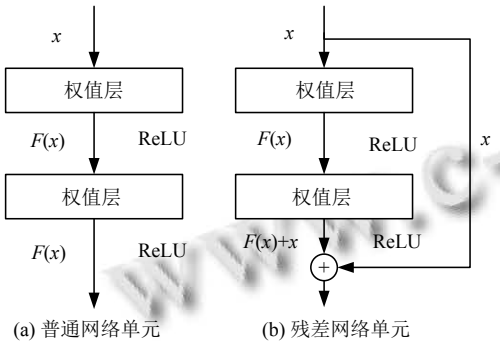


图 4 普通网络单元与残差网络单元

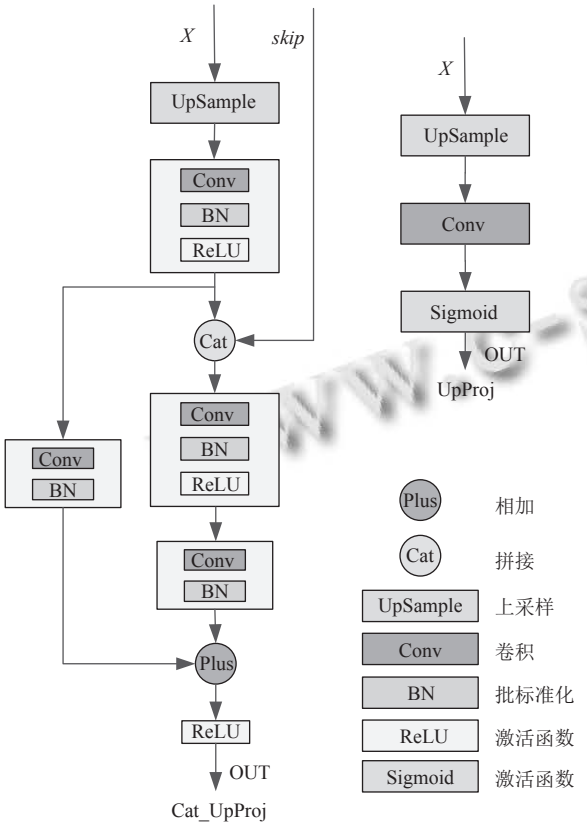


图 5 本文 DepthNet 中的上采样模块

(2) 多种上采样算法

图 5 中的上采样层 UpSample 分别采用反卷积算法、双线性插值算法和像素重组算法实现, 并进行实验对比分析.

1) 反卷积算法

输入图像通过标准卷积计算后, 得到的输出图像尺寸会变小, 而有时我们需要将输出图像尺寸增大以顺利进行下一步的操作. 增大图像尺寸, 将小分辨率图像扩展到大分辨率图像的操作, 即为上采样. 反卷积 (transpose)[17] 为上采样方法中的一种, 也称为转置卷积. 先将卷积核反转, 然后按照设定值对输入图像进行补 0 以增大图像尺寸, 最后使用反转后的卷积核对补 0 后的图像进行标准卷积.

① 卷积核反转. 此处的反转并不是线性代数中的转置操作, 而是对 4 方向的参数进行逆序操作. 反转前后对比如图 6 所示.

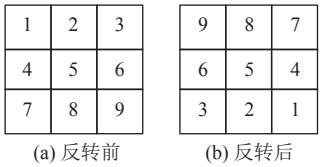


图 6 卷积核反转

② 对输入进行补 0 操作, 补 0 个数为步长减 1, 如果步长为 1, 则不补 0. 如图 7 所示, 输入尺寸为 3×3, 步长为 2, 输出为 6×6, 补 0 后的矩阵为 5×5.

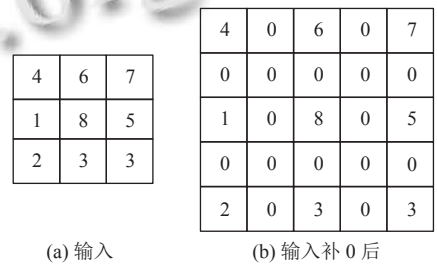


图 7 根据步长补 0

③ 对补 0 后的输入矩阵再做整体补 0. 以补 0 后的 5×5 输入矩阵作为输入, 反转以后的卷积核作为权重, 此时的步长参数变为 1. 按照普通卷积的填充方式计算补 0 参数, 并将计算出的补 0 参数赋给各自的反方向, 填充后的矩阵如图 8 所示.

④ 使用反转后的卷积核对整体补 0 后的矩阵进行步长为 1 的标准卷积运算, 即可完成反卷积的计算.

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	4	0	6	0	7	0
0	0	0	0	0	0	0	0
0	0	1	0	8	0	5	0
0	0	0	0	0	0	0	0
0	0	2	0	3	0	3	0
0	0	0	0	0	0	0	0

图 8 整体补 0

2) 双线性插值算法

双线性插值算法 (bilinear)[18] 是通过对待求点的邻近 4 点求两次线性插值, 将所得结果作为待求点像素值的方法, 图 9 给出了双线性插值的计算示意图.

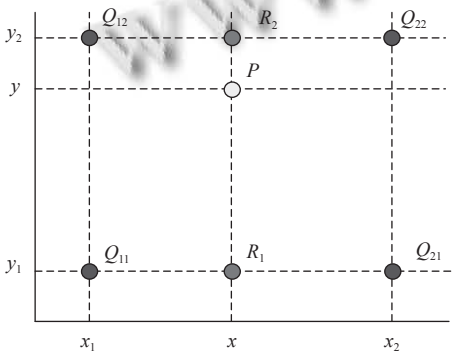


图 9 双线性插值示意图

图 9 中, Q_{11}、Q_{12}、Q_{22}、Q_{21} 为待求点 $P=(x, y)$ 的邻近 4 点, 首先对 Q_{11}、Q_{21} 和 Q_{12}、Q_{22} 分别在 x 方向上进行插值操作, 得到点 $R_1=(x, y_1)$ 和 $R_2=(x, y_2)$ 的像素值. 计算公式如下:

$$f(R_1)\approx\frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{11}) \tag{2}$$

$$f(R_2)\approx\frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22}) \tag{3}$$

然后再对 R_1 和 R_2 在 y 方向上进行插值, 得到 P 点的像素值, 计算公式为:

$$f(P)\approx\frac{y_2-y}{y_2-y_1}f(R_1)+\frac{y-y_1}{y_2-y_1}f(R_2) \tag{4}$$

3) 像素重组算法

像素重组算法 (PixelShuffle) 主要是使用尺寸不变卷积和多通道间像素重组来将低分辨率输入图像采样为高分辨率输出图像, 其中多通道间像素重组操作也称为亚像素卷积[19]. 该算法的主要作用是将一个尺寸为 $H\times W$ 的输入图像变为 $rH\times rW$ 的输出图像, r 为放大倍数. 首先通过中间卷积层进行多次尺寸不变卷积后得到 r^2 个通道的特征图, 然后通过轮询的方式依次选取每个通道对应位置的像素进行重组以得到高分辨率的输出图像, 如图 10 所示.

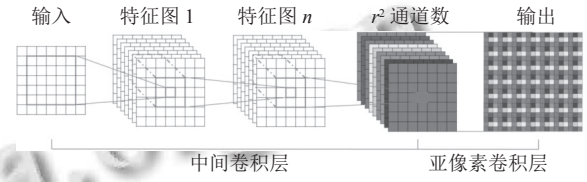


图 10 PixelShuffle 处理过程

1.4 损失函数

P^{2}Net 在每个关键点 p_t 的局部窗口上定义了一个支持域Ω_{p_t}, 然后在每个支持域上计算其光度损失. 将基于面片的光度损失 L_{ph} 定义为关键点支持域上的 $L1$ 范数和结构相似性损失 $SSIM$ 的组合, 如式 (5) 所示. 其中, $I_t[\Omega_{p_t}]$为原目标图像中的支持域, $\hat{\Omega}_{p_t}$为结合相对位姿 $T_{t\to s}$ 计算出的转换后支持域, $I_s[\hat{\Omega}_{p_t}]$为根据 $\hat{\Omega}_{p_t}$得到的重构目标图像中相应的区域[12]. α为加权因子, 设置为 0.85.

$$\begin{aligned}L_{ph}=&\alpha SSIM(I_t[\Omega_{p_t}],I_s[\hat{\Omega}_{p_t}])\\&+(1-\alpha)\left\|I_t[\Omega_{p_t}]-I_s[\hat{\Omega}_{p_t}]\right\|_1\end{aligned} \tag{5}$$

此外, P^{2}Net 假设大多数同质颜色区域都是平面区域, 并采用 Felzenszwalb 法进行超像素分割. 对于从输入图像提取的某一超像素 SPP_m 和对应的预测深度$D(p_n)$, 可以通过三维平面拟合方法计算出超像素 SPP_m 的拟合平面深度$D'(p_n)$. 之后, 将超像素 SPP_m 的预测深度 $D(p_n)$与拟合平面深度$D'(p_n)$的绝对误差作为平面一致性损失, 其中 M 表示超像素数量, N 表示每个超像素中的像素数量.

$$L_{spp}=\sum_{m=1}^{M}\sum_{n=1}^{N}\left|D(p_n)-D'(p_n)\right| \tag{6}$$

除上述基于面片的光度损失 L_{ph} 和平面一致性损失 L_{spp} 之外, 最终的损失函数还添加了边缘感知的梯度平滑损失 L_{sm}.

$$L_{sm}(d_t, I_t) = \sum_{p_t} |\partial_x d_t| \mathrm{e}^{-|\partial_x I_t|} + |\partial_y d_t| \mathrm{e}^{-|\partial_y I_t|} \quad (7)$$

其中, $d_t = D_t / \bar{D}_t$, D_t为 DepthNet 输出的深度图, $\bar{D}_t$为深度图中像素的平均值. ∂_x和∂_y为当前像素水平和垂直方向上的梯度. 最终的损失函数 L 如式 (8) 所示, 其中 λ_1和λ_2分别设置为 0.001 和 0.05.

$$L = L_{ph} + \lambda_1 L_{sm} + \lambda_2 L_{spp} \quad (8)$$

2 实验结果与分析

2.1 实验数据集与评价指标

本文采用纽约大学公开的室内图像数据集 NYU-Depth V2[20] 对网络模型进行训练与测试. 该数据集由微软的 Kinect 摄像机拍摄采集, 包含 464 个场景, 场景的深度范围为 0–10 m. 本文使用 P^{2}Net 提供的方法对官方数据集进行处理, 得到约 238 个场景, 共 15 543 个连续帧图像对, 每个连续帧图像对包含 9 张连续图片. 采用 Adam 优化器, 总训练批次 epochs 为 41, 批量大小 batch_size 为 12. 前 25 个 epoch 的初始学习率设置为 10^{-4}, 然后在接下来的每 10 个 epoch, 乘于一次 0.1. 在训练过程中采用随机翻转等预处理操作, 所有输入图像的大小调整为 288×384. 在测试过程中, 对预测深度图进行上采样, 使其恢复到原始尺寸 640×480. 在网络训练之前, 需要先使用 Felzenszwalb 超像素分割算法对目标图片提取超像素. 之后, 在网络训练过程中会基于提取的超像素重建分段平面, 同时也会使用 DSO 算法对目标图像提取关键点.

单目深度估计的评价指标主要分为两种[21], 一种为误差指标, 本文主要采用均方根误差 (root mean squared error, *RMSE*) 和平均相对误差 (absolute relative error, *REL*); 另一种为准确率, 本文采用阈值内准确度 δ, 其中阈值 $thr = 1.25, 1.25^2, 1.25^3$, 阈值越小, 说明评价指标越严格. 这 3 个客观指标表达式如下:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - g_i)^2} \quad (9)$$

$$REL = \frac{1}{n}\sum_{i=1}^{n}\frac{|d_i - g_i|}{g_i} \quad (10)$$

$$\delta = \max\left(\frac{d_i}{g_i}, \frac{g_i}{d_i}\right) < thr \quad (11)$$

其中, d_i 表示预测值, g_i 表示真实值, n 表示每张图片中像素的数量.

2.2 实验结果分析

表 1 为原 P^{2}Net 网络和基于本文残差上采样结构得到的 3 种改进 P^{2}Net 网络的实验结果, 其中 Res-T、Res-B 和 Res-P 分别代表基于反卷积算法、双线性插值算法和像素重组算法构建出的残差上采样结构, Nearest 为原网络中最近邻插值上采样结构. 误差指标均方根误差 *RMSE* 和平均相对误差 *REL* 的值越小表示效果越好, 准确度指标 $\delta_{1.25}$、$\delta_{1.25}^2$ 和 $\delta_{1.25}^3$ 的值越大效果越好, 表 1 中加粗字体为每列最优值. 图 11 为实验结果可视化图片示例, 图像颜色越深表示距离越近, 反之越远.

表 1 实验结果对比

网络	采样结构	*RMSE*	*REL*	$\delta_{1.25}$	$\delta_{1.25}^2$	$\delta_{1.25}^3$
P^{2}Net	Nearest	0.621	0.169	75.2	93.7	98.2
P^{2}Net-1	Res-T	0.607	0.166	75.7	94.0	98.3
P^{2}Net-2	Res-B	0.604	**0.164**	**76.2**	94.1	**98.5**
P^{2}Net-3	Res-P	**0.602**	0.166	76.0	**94.3**	**98.5**

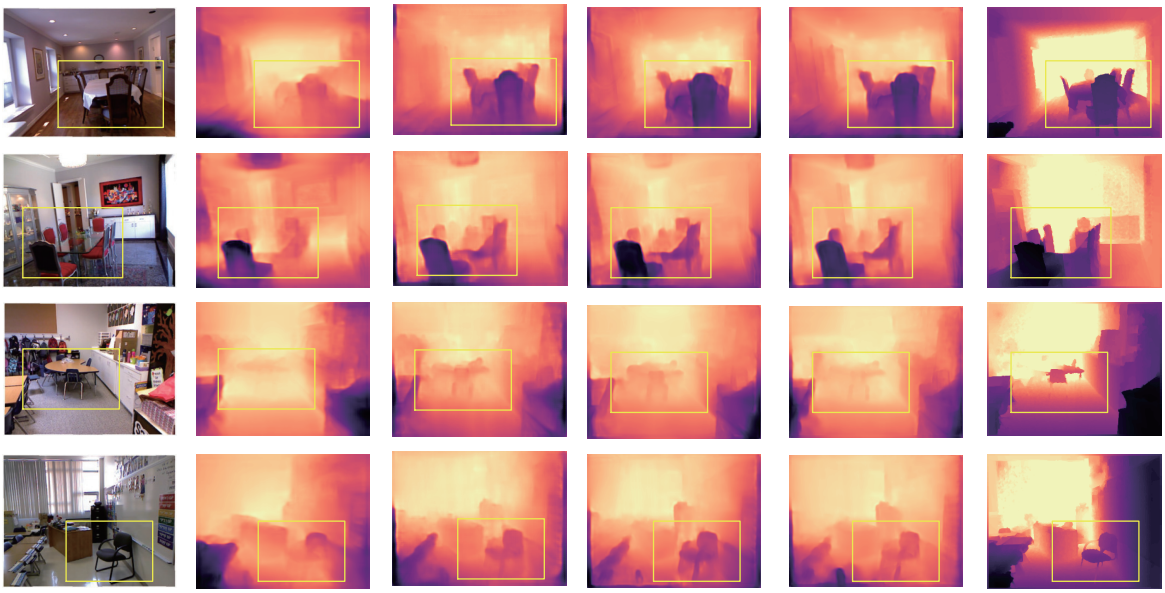
由表 1 可知, 3 种改进 P^{2}Net 网络在所有评价指标上均优于原 P^{2}Net 网络, 表明本文新的残差上采样结构相比原上采样层有着更强的特征提取能力和更好的预测效果. 与原 P^{2}Net 网络相比, 改进网络 P^{2}Net-1、P^{2}Net-2 和 P^{2}Net-3 在均方根误差 *RMSE* 指标上分别降低了 2.25%、2.73% 和 3.05%; 在平均相对误差 *REL* 指标上分别降低了 1.77%、2.96% 和 1.78%. 3 种改进网络中, 基于 Res-B 和 Res-P 得到的 P^{2}Net-2 和 P^{2}Net-3 网络均有 3 项评价指标取得了最优结果, 且整体表现都要优于 P^{2}Net-1 网络.

由图 11 中可以看出, 相比于原 P^{2}Net 网络, 本文 3 种改进网络的预测深度图整体数值分布更接近于真实深度图. 同时, 从图中虚线框可以看出, 3 种改进网络的预测深度图中物体结构更加完整, 细节信息更加丰富, 说明本文的改进网络能够获取更多特征信息, 提高图像的生成质量. 而 3 种改进网络中, P^{2}Net-2 预测的深度图中物体结构更为清晰, 整体预测效果要略优于 P^{2}Net-1 和 P^{2}Net-3 网络.

3 结论与展望

由于 P^{2}Net 中深度预测网络 DepthNet 所采用的上采样方法为计算过程较为简单、图像生成质量较差的最近邻插值算法, 使得预测深度图中物体完整性较

差、整体效果不理想. 因此, 本文基于多种上采样算法构建出残差上采样结构来替换原网络中的上采样层, 进一步提高了预测深度图中物体结构的完整性, 降低了预测误差. 本文训练与测试的输入数据均为连续帧图像对, 并未对视频数据进行训练与分析, 下一步将会由图像深度估计扩展到视频深度估计进行研究.



(a) RGB 图像 (b) P²Net (c) P²Net-1 (d) P²Net-2 (e) P²Net-3 (f) 真实深度图

图 11 实验结果可视化

参考文献

1 张敏. 基于多线索信息的深度估计算法研究与实现 [硕士学位论文]. 大连: 大连理工大学, 2021. [doi: 10.26991/d.cnki.gdllu.2021.003186]

2 公冶佳楠, 李轲. 基于光场图像序列的自适应权值块匹配深度估计算法. 计算机系统应用, 2020, 29(4): 195–201. [doi: 10.15888/j.cnki.csa.007387]

3 Tankus A, Sochen N, Yeshurun Y. Reconstruction of medical images by perspective shape-from-shading. Proceedings of the 17th International Conference on Pattern Recognition. Cambridge: IEEE, 2004. 778–781. [doi: 10.1109/ICPR.2004.1334644]

4 Skarbek W. Shape from motion revisited. Proceedings of the 10th International Conference on Active Media Technology. Warsaw: Springer, 2014. 383–394. [doi: 10.1007/978-3-319-09912-5_32]

5 Zhang XD, Liu ZQ, Jiang MS, *et al*. Fast and accurate auto-focusing algorithm based on the combination of depth from focus and improved depth from defocus. Optics Express, 2014, 22(25): 31237–31247. [doi: 10.1364/OE.22.031237]

6 张喆韬, 万旺根. 基于 LRSDR-Net 的实时单目深度估计. 电子测量技术, 2019, 42(19): 158–163. [doi: 10.19651/j.cnki.emt.1902912]

7 詹雁, 张娟, 金昌基. 联合语义感知与域适应方法的单目深度估计. 传感器与微系统, 2021, 40(5): 60–63. [doi: 10.13873/J.1000-9787(2021)05-0060-04]

8 Garg R, Kumar BGV, Carneiro G, *et al*. Unsupervised CNN for single view depth estimation: Geometry to the rescue. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 740–756. [doi: 10.1007/978-3-319-46484-8_45]

9 Wong A, Soatto S. Bilateral cyclic constraint and adaptive regularization for unsupervised monocular depth prediction. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5637–5646. [doi: 10.1109/CVPR.2019.00579]

10 Zhou WH, Zhou EC, Liu GM, *et al*. Unsupervised monocular depth estimation from light field image. IEEE Transactions on Image Processing, 2020, 29: 1606–1617. [doi: 10.1109/TIP.2019.2944343]

11 Zhou JS, Wang YW, Qin KH, *et al*. Moving indoor: Unsupervised video depth learning in challenging environments. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 8617–8626. [doi: 10.1109/ICCV.2019.00871]

12 Yu ZH, Jin L, Gao SH. P²Net: Patch-match and plane-

regularization for unsupervised indoor depth estimation. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 206–222. [doi: 10.1007/978-3-030-58586-0_13]

13 Engel J, Koltun V, Cremers D. Direct sparse odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611–625. [doi: 10.1109/TPAMI.2017.2658577]

14 Felzenszwalb PF, Huttenlocher DP. Efficient graph-based image segmentation. International Journal of Computer Vision, 2004, 59(2): 167–181. [doi: 10.1023/B:VISI.0000022288.19776.77]

15 高昊昇. 基于深度学习的无监督单目图像序列深度估计 [硕士学位论文]. 南京: 东南大学, 2020. [doi: 10.27014/d.cnki.gdnau.2020.000546]

16 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]

17 Zeiler MD, Krishnan D, Taylor GW, *et al*. Deconvolutional networks. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010. 2528–2535. [doi: 10.1109/CVPR.2010.5539957]

18 Jaderberg M, Simonyan K, Zisserman A, *et al*. Spatial transformer networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 2017–2025.

19 Shi WZ, Caballero J, Huszár F, *et al*. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1874–1883. [doi: 10.1109/CVPR.2016.207]

20 Silberman N, Hoiem D, Kohli P, *et al*. Indoor segmentation and support inference from RGBD images. Proceedings of the 12th European Conference on Computer Vision. Florence: Springer, 2012. 746–760. [doi: 10.1007/978-3-642-33715-4_54]

21 Cheng XJ, Wang P, Yang RG. Depth estimation via affinity learned with convolutional spatial propagation network. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 108–125. [doi: 10.1007/978-3-030-01270-0_7]

(校对责编: 牛欣悦)