

文章编号: 2095-2163(2023)10-0127-06

中图分类号: TP183

文献标志码: A

# 基于轻量化融合损失网络的单目深度估计

周润民<sup>1</sup>, 王晓红<sup>1</sup>, 王辉<sup>1</sup>, 苏靖<sup>1</sup>, 刘璐<sup>2</sup>

(1 贵州大学 矿业学院, 贵阳 550025; 2 贵州大学 林学院, 贵阳 550025)

**摘要:** 针对现有监督学习的单目深度估计存在编码器模型参数量大、计算复杂度高问题, 本文构建了一种基于轻量化融合损失网络的单目深度估计模型, 该模型基于编码器-解码器设计。首先, 编码器采用网络搜索技术的 EfficientNet; 其次, 解码器设计一种结合双通道空间注意力机制的 SU-Net 结构模型; 最后, 在解码器的输出端基于该模型构建 ResNet50 损失网络, 对物体边缘、纹理和轮廓差异进行惩罚, 来弥补编码-解码阶段特征信息表达能力不足的问题。在 NYU-Depth V2 数据集上开展了试验, 试验结果表明: 该模型深度估计的均方根误差 (RMSE) 达到了 0.419, 优于原始基准模型。

**关键词:** 单目深度估计; 轻量化; 损失网络; 空间注意力机制

## Monocular depth estimation based on lightweight fusion loss network

ZHOU Runmin<sup>1</sup>, WANG Xiaohong<sup>1</sup>, WANG Hui<sup>1</sup>, SU Jing<sup>1</sup>, LIU Lu<sup>2</sup>

(1 College of Mining, Guizhou University, Guiyang 550025, China;

2 College of Forestry, Guizhou University, Guiyang 550025, China)

**[Abstract]** For supervised monocular depth estimation models, there are some problems such as large number of encoder model parameters and high computational complexity. This paper presents a monocular depth estimation model based on lightweight fusion loss network. The model is designed based on encoder - decoder architecture. The encoder adopts the EfficientNet and the decoder follows a SU-Net structural model combining the dual channel spatial attention mechanism. Finally, the ResNet50 loss network is constructed based on the model at the output end of the decoder, punishing the differences in the edges, textures and outline of objects to make up for the lack of expression ability of the feature information in the encoding-decoding stage. The experiments are conducted on NYU-Depth V2 dataset. The test results show that the root mean square error (RMSE) of depth estimation of this model is 0.419, which is better than the baseline model.

**[Key words]** monocular depth estimation; lightweight; loss network; spatial attention mechanism

## 0 引言

场景深度感知是重建周围环境的关键环节, 包含了目标的三维结构信息, 对于室内三维测图、无人机避障和自动驾驶等研究领域都具有重要的意义<sup>[1]</sup>。

目前深度信息的获取可以分为主动式和被动式。主动式获取方式有激光雷达、RGB-D 相机和 3D 结构光等, 均是通过反射物体表面来获取深度信息, 但通常设备价格昂贵、体积重及功耗高; 被动式获取是基于图像序列, 利用相机捕获同一场景下单摄或者多摄目标的影像, 通过三角测量方式获取深

度信息, 拍摄时要求时间同步, 通常拍摄设备体积较大, 不适用于微型机器人平台 (例如微型航空和地面自动驾驶), 因此研究成本低、体积小和能源效率高的单目相机进行场景深度感知具有重要的现实意义。当前采用的编码端模型臃肿、参数量大, 如何得到轻量化且精准的图像深度估计模型成为了关键性挑战。

## 1 相关工作

传统单目深度估计依靠手工提取特征, 输出的深度图像比较模糊, 而且场景中物体细节部分信息丢失, 导致估计的图像深度信息不理想。随着深度

**基金项目:** 国家自然科学基金(41901225); 教育部规划基金项目(22YJZH083); 贵州省科技计划项目(黔科合支撑[2022]一般204)。

**作者简介:** 周润民(1995-), 男, 硕士研究生, 主要研究方向: 深度学习、三维测图与重建; 王晓红(1970-), 男, 博士, 副教授, 主要研究方向: 遥感数据处理与分析、3S 技术应用。

**通讯作者:** 王晓红 Email: xhwang@gzu.edu.cn

**收稿日期:** 2022-11-02

学习的发展,极大的促进了深度估计领域的发展。对于高精度的深度估计任务,文献[2]提出将卷积神经网络(Convolutional Neural Networks, CNN)用于预测单目图像深度估计,设计了一个由粗到细的两个堆栈组成的卷积神经网络,粗尺度网络预测图像的全局深度,细尺度网络提取局部深度并优化图像的局部细节;文献[3]将连续条件随机场(Conditional Random Field, CRF)和深度卷积神经网络相结合,构建深度卷积神经网络模型来提升深度估计效果;文献[4]同样采用双尺度网络模型进行深度估计,同时引入了感知损失使预测深度进一步提高;文献[5]构建了一种基于残差学习的全卷积网络架构用于预测图像的深度估计,该模型编码器采用 ResNet50 模型结构,该结构层次更深同时能够提取更加丰富的特征信息,但是更深的网络导致模型臃肿、参数量增加、计算复杂度高;文献[6]构建了一种轻量金字塔解码结构的单目深度估计模型,编码器使用 ResNet50 用于特征提取,解码器将轻量金字塔解码模块提取到的不同感受野特征图进行融合,并采用跳跃连接实现知识共享,以提升深度估计的性能。

近年来深度学习不断发展,网络模型结构的轻量化、高效越来越重要。文献[7]通过深度可分离卷积代替标准卷积,提出了一种轻量化的网络模型 MobileNet,网络的运行速度有很大的提升,但该网络模型结构简单,没有复用图像特征进行特征融合,

随后提出了更加高效的网络模型;文献[8]构建了一种以 MoblieNet 为基础,适合在嵌入式系统上使用的模型,该模型的推理速度得到了进一步提升,但是由于解码器采用的是传统 U-Net 结构上采样,导致预测深度图目标轮廓模糊、边缘细节不够锐利;文献[9]提出了一种可以进行量化的单目深度估计模型,并完成对该模型的量化压缩工作,该模型实现了网络模型的轻量化,但是以牺牲深度估计的准确性为代价的。

本文构建一种基于轻量化融合损失网络的模型结构,对预测结果在多个网络上进行特征约束。模型在编码阶段提取全局特征,解码阶段通过特征融合整合全局和局部特征信息,使得预测的深度图具有很好的深度连续性,最后预测的深度图通过损失网络更好地惩罚预测信息和目标信息,进一步提高深度估计精度。

## 2 整体模型框架

基于轻量化融合损失网络的模型结构如图 1 所示。在整体模型框架中,编码器阶段是图像特征的提取阶段,低维特征分辨率较高,包含丰富的特征信息,而高维特征分辨率较低,提取到的图像特征更加抽象;解码器阶段是提取特征的恢复阶段,将编码器提取的高维特征与跨层共享的低维特征进行特征融合,生成清晰准确的特征预测图;最后,预测深度传递到 ResNet50 损失网络进行融合感知。

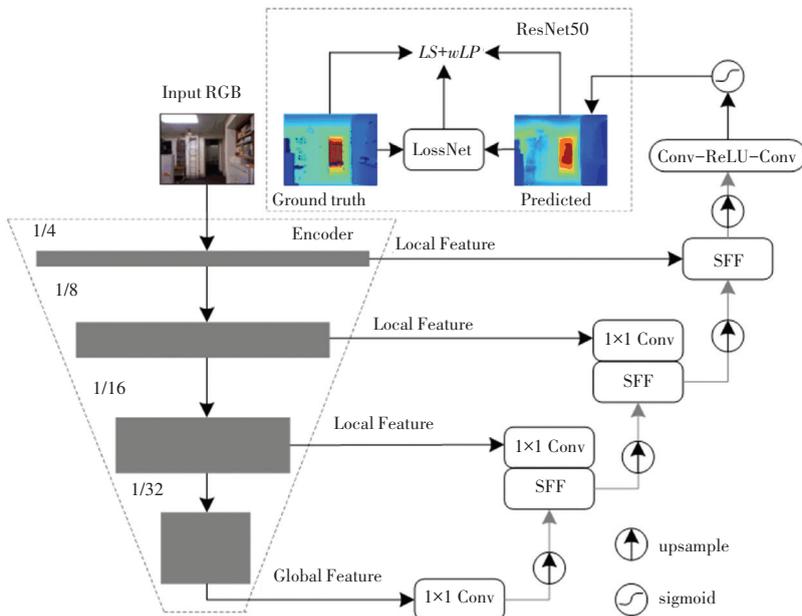


图 1 基于轻量化融合损失网络的模型结构

Fig. 1 Model structure based on lightweight fusion loss network

## 2.1 编码器结构

在样本数据较少的情况下,为获得更高的准确率,通常会扩展深度学习模型的深度、宽度以及分辨率。常规卷积神经网络尽管可以任意调整其中的几个维度,但是都需要手动调节参数,且很可能伴随着模型性能和效率的下降。EfficientNet 是一种结合神经网络搜索技术的多维度混合放缩模型,在 EfficientNet-B0 的基线上调整深度、宽度以及分辨率缩放获得<sup>[10]</sup>。式(1):

$$\begin{cases} d = \alpha^\theta, \alpha \geq 1 \\ w = \beta^\theta, \beta \geq 1 \\ r = \gamma^\theta, \gamma \geq 1 \end{cases} \quad (1)$$

其中,  $\alpha, \beta, \gamma$  为常数,是模型深度、宽度和分辨率的网络搜索技术参数,缩放系数  $\theta$  是模型的扩张量。

目前深度估计任务编码器大部分采用 ResNet 作为特征提取器,通过不断加深网络的深度以获取更高的精度。但随着网络的加深,模型变得十分庞大,计算效率低。因此,从实际应用的角度,综合考虑参数尺寸及模型复杂度,本文选择 EfficientNet-B2 模型并运用迁移学习加载预训练模型,提高深度预测性能的同时加速网络收敛,防止梯度弥散和梯度爆炸等问题。

## 2.2 解码器结构

解码器通过编码器输出的特征图进行上采样,并在空间维度上恢复特征映射,生成与输入尺寸相同的深度图。传统的解码器采用卷积与插值的 U-Net 结构恢复深度图,本文使用一种全新的解码器结构 SU-Net (Selective Feature Fusion-based U-Net),主要包含选择特征融合模块 (Selective Feature Fusion, SFF),相邻 SFF 模块之间通过跳跃连接和双线性插值上采样恢复到对应尺度,其中  $1 \times 1$  卷积层使输入值的通道维度减少,与编码器对应的跳跃连接输出值在通道维度上相等,然后经过两个卷积层和一个 Sigmoid 函数得到预测深度图<sup>[11]</sup>。

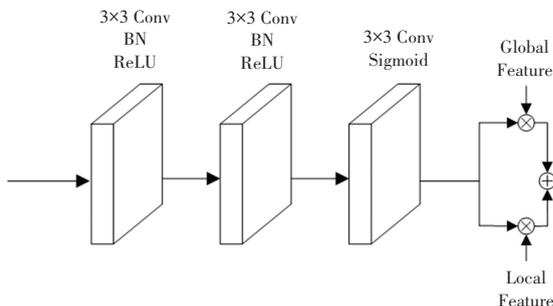


图 2 选择特征融合模块

Fig. 2 Feature selection and fusion module

SFF 模块类似于空间注意力模块如图 2 所示,

将浅层局部特征、深层全局特征分别表示为  $f_1 \in R^{W \times H \times C_1}$ ,  $f_2 \in R^{W \times H \times C_2}$ , 空间位置集合  $R = \{(x, y) \mid x = 1, 2, \dots, W; y = 1, 2, \dots, H\}$ ,  $(x, y)$  为空间特征的坐标信息。为了获取空间特征的注意力图,将局部特征和全局特征在通道维度上进行特征拼接,将拼接特征通过三层  $3 \times 3$  卷积操作得到双通道空间注意力图,通过 Sigmoid 函数运算将其映射至  $[0, 1]$ ,然后将局部特征和全局特征分别与空间注意力通道矩阵相乘,得到专注于空间关系的有效特征,最后使两种特征矩阵相加。SFF 模块生成的双通道特征注意力图能够自适应地选择和融合特征值,使网络从局部和全局信息出发,选择对网络性能好的空间特征权重,抑制无效权重。双通道空间注意力特征和选择特征融合模块计算如式(2)和式(3)所示:

$$M_s(F) = \sigma(\text{Conv}[f_1, f_2]) \quad (2)$$

$$F_{SFF}(f_1, f_2) = M_{S_1}(F) \cdot f_1 + M_{S_2}(F) \cdot f_2 \quad (3)$$

其中,  $\sigma$  表示 Sigmoid 激活函数;  $\text{Conv}$  为卷积层运算;  $M_s(F)$  表示双通道注意力特征。

## 3 损失函数

损失函数是神经网络优化中至关重要的一部分。选择合适的损失网络不仅可以加快网络收敛,同时能够保证模型寻找到全局最优解,衡量模型预测能力, L1、L2 函数是目前应用比较广泛的损失函数。本文根据单目深度估计以往研究方法,采用尺度不变对数误差损失函数和基于本文构建的残差感知损失函数联合相加得到联合损失函数。联合损失函数的表达式(4):

$$L_{total}(Y, \hat{Y}) = L(Y, \hat{Y}) + wL_p(Y, \hat{Y}) \quad (4)$$

其中,  $L_{total}$  表示联合损失函数;  $L$  为尺度不变对数误差损失函数;  $L_p$  为残差感知损失函数;  $w \in [0, 1]$  为残差感知损失函数的平衡因子,试验中设置为 0.5。

### 3.1 尺度不变对数误差损失函数

在模型训练期间,目标深度图中的物体边缘、弱纹理区域和镜面反射等周围会出现丢值现象。本文利用简单的掩膜方法来处理这些缺失值,即通过目标深度图获取一个掩膜矩阵,重新计算预测与真值深度图的像素值,使模型在有效点上计算损失,排除无效点的影响。损失函数表达式(5):

$$L(Y, \hat{Y}) = \frac{1}{n} \sum_i d_i^2 - \frac{\lambda}{n^2} \left( \sum_i d_i \right)^2 \quad (5)$$

式中:  $\lambda \in [0, 1]$  为尺度因子,  $d_i = \log_{10} Y - \log_{10} \hat{Y}$ ,

$Y, \hat{Y}$  分别为模型预测深度图与真实深度图的集合,  $Y = [y_1, y_2, \dots, y_i]$ ,  $\hat{Y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_i]$ , 试验设置  $\lambda = 0.5$ 。

### 3.2 残差感知损失函数

目前有监督的深度估计任务核心是对预测深度图与真实深度图进行差异评估, 传统的解决方法是建立逐像素损失函数, 这种方式简单高效。最近也有部分研究将感知损失用于遥感影像去噪、医学影像去噪和单目深度估计, 以上均是采用预训练的 VGG (Visual Geometry Group) 网络, 对输入图像提取高维度特征, 随后使用优化的感知损失函数评估特征之间的差异<sup>[12-13]</sup>。采用 VGG 网络模型预测图像, 预测结果图质量好, 物体边缘锐利, 但是训练优化模型时开销时间较大。本文综合以上的优点, 构建一种具有残差的感知损失函数, ResNet50 相对于其他网络具有更深的网络层数和更抽象的高层特征。当预测深度  $\varphi(Y)$  的特征偏离目标深度  $\varphi(\hat{Y})$  时, 感知损失会惩罚预测深度 ( $Y$ ) 的特征, 惩罚边缘、纹理和轮廓等差异。残差感知损失网络结构见表 1, 感知损失表达式 (6):

$$L_p(Y, \hat{Y}) = \frac{1}{n} \sum_i \|\varphi(Y) - \varphi(\hat{Y})\|_2^2 \quad (6)$$

其中,  $\varphi$  表示损失网络,  $\varphi(Y)$ 、 $\varphi(\hat{Y})$  分别表示真实深度图和预测深度图经过第  $i$  层后提取到的特征集合。

表 1 残差感知损失网络结构参数

Tab. 1 Residual perceived loss network structural parameters

卷积层	参数设置
Convolution	7×7 Conv, Stride 2
Pooling	3×3 max pool, Stride 2
Resblock(1)	$\left. \begin{array}{l} 1 \times 1 \text{ Conv, } 64 \\ 3 \times 3 \text{ Conv, } 64 \\ 1 \times 1 \text{ Conv, } 256 \end{array} \right\} \times 3$
Resblock(2)	$\left. \begin{array}{l} 1 \times 1 \text{ Conv, } 128 \\ 3 \times 3 \text{ Conv, } 128 \\ 1 \times 1 \text{ Conv, } 512 \end{array} \right\} \times 4$

## 4 试验结果与分析

### 4.1 试验设置与数据集

本文采用纽约大学公开的室内环境数据集 NYU-Depth V2, 原始 RGB-D 数据由 Kinect 相机拍摄采集的 464 个室内场景, 包括 12 万对彩色图像与深度图像, 场景的深度范围为 0~10 m, 在 249 个场景中选取约 2 万张子集用于训练, 其余 215 个场景

用官方划分的 654 张图像子集作为测试。本文网络结构的原始图像输入大小为 480×640, 在训练过程中对输入的彩色图片按随机概率 50% 进行水平翻转、随机裁剪、颜色变换等, 并且引入 CutDepth 策略进行数据增强。

本文试验操作系统为 Ubuntu, 试验基于 Pytorch 1.9 深度学习框架, 试验硬件配置为 Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz 处理器, 内存为 16 GB, 模型在 GPU 显存为 12 G NVIDIA TITAN XP 的显卡上训练, 具体试验参数见表 2。

表 2 试验参数

Tab. 2 Test parameters

参数项	参数设置
Batch size	8
Optimizer	Adam
Optimizer momentum	$(\beta_1 = 0.9, \beta_2 = 0.999)$
epoch	30
learning rate strategy	OneCycleLR

### 4.2 定量评价指标

本文按照 6 项指标来评价本文构建模型的试验结果, 将模型预测与真实标签估计效果进行比较。

(1) 绝对值相对误差 (Abs Rel), 式 (7):

$$AbsRel = \frac{1}{|N|} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{\hat{y}_i} \quad (7)$$

(2) 均方根误差 (RMSE), 式 (8):

$$RMSE = \sqrt{\frac{1}{|N|} \sum_{i=1}^N \|y_i - \hat{y}_i\|^2} \quad (8)$$

(3) 对数平均误差 (log), 式 (9):

$$\log = \frac{1}{|N|} \sum_{i=1}^N |\lg d_i - \lg \hat{d}_i| \quad (9)$$

(4) 阈值相似度  $\delta$ , 式 (10):

$$\delta = \max \left( \frac{y_i}{\hat{y}_i}, \frac{\hat{y}_i}{y_i} \right) < thr \quad (10)$$

其中, 阈值  $thr = \{1.25, 1.25^2, 1.25^3\}$ ;  $i$  是像素索引;  $d_i, \hat{d}_i$  分别为像素  $i$  处的预测深度值和真实深度值;  $N$  是测试集深度图中所有像素数目的总和。

### 4.3 消融试验

为了验证数据增强、选择特征融合模块和损失网络 3 个消融因素对本文模型的有效性, 本文采用以下的方式进行消融试验分析: 编码器均统一采用 EfficientNet-B2, 解码器为 U-Net (模型 1)、解码器为 U-Net 采用 CutDepth (模型 2)、解码器为 SU-Net (模型 3)、解码器为 SU-Net 采用 CutDepth (模型 4)

以及构建的损失网络(模型 5)。3 种因素对网络消融试验的结果见表 3, 结果表明各项评估指标在解码器 SU-Net 比 U-Net 有一定的精度提升, 这归因于选择特征融合模块中双通道空间注意力图, 通过对全局与局部路径的空间特征权重的重标定, 对特征图进行自适应性的恢复校准, 从而提高模型识别恢复路径的准确性。通常情况下, 数据增强策略使样本多样性增多, 模型精度越高。模型 2 基于 U-Net 与模型 4 基于 SU-Net, 在数据增强策略下, 模型的误差小、准确度高, 与实际情况相符。

表 3 不同模块误差对比

Tab. 3 Error comparison of different modules

模型	误差			准确率		
	Rel	RMSE	log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
模型 1	0.128	0.436	0.054	0.845	0.975	0.995
模型 2	0.125	0.420	0.052	0.853	0.977	0.995
模型 3	0.125	0.430	0.053	0.851	0.975	0.995
模型 4	0.124	0.424	0.053	0.853	0.976	0.995
模型 5	0.120	0.419	0.051	0.860	0.978	0.996

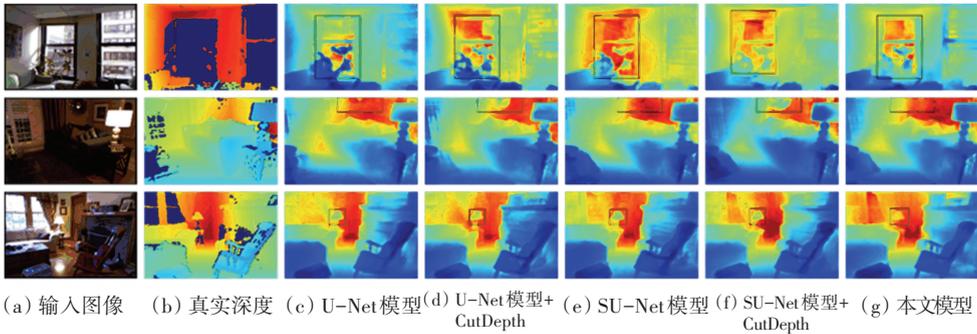


图 3 构建模型深度效果

Fig. 3 Depth results of proposed model

#### 4.4 对比试验

为了验证本文模型的深度估计精度, 将其与近年来其他应用深度学习的单目深度估计在 NYU-Depth V2 数据集上的对比试验结果见表 4。试验结果表明, 本文模型的绝对值相对误差 (Abs Rel)、均方根误差 (RMSE)、对数平均误差 (log) 以及不同阈值下的准确率 ( $\delta$ ) 在对比试验中优于其他方法。文献[14]采用 ResNet101 模型作为深度估计的基础网络并结合 DenseASPP 部分结构, 提出了一个 DCDN (Deep Convolution DenseASPP Network) 新的卷积神经网络模型; 而本文构建模型将 EfficientNet-B2 结构作为编码器, 其模型参数量更加的轻量, 在绝对值相对误差、均方根误差以及对数平均误差分别减少了 15.8%、13.1%、11.8%, 在阈值为  $\delta < 1.25$ 、 $\delta < 1.25^2$ 、 $\delta < 1.25^3$  的准确度分别提升 4.3%、1.4%、

本文构建的模型与其余 4 种模型相比, 误差更小, 准确率更高。这得益于模型不仅汲取了数据增强与 SU-Net 的两个优点, 还得到了损失网络对预测输出的进一步感知。损失网络结合迁移学习中的高级特征信息, 通过对预测深度与真实深度的融合感知, 提高了网络对标签信息的重构使用, 提取更为详细的细节信息, 弥补编码-解码阶段特征信息表达能力不足的问题。

5 种消融试验预测深度如图 3 所示, 可见本文最终构建模型相比前 4 种消融试验在深度预测更加接近真实深度, 存在镜面反射时预测深度有明显的提升, 对于目标的轮廓清晰可见, 边缘更加锐利, 这得益于损失网络强大的损失感知, 使预测深度图感知到与真实深度图之间存在的差异, 通过每一次感知损失函数的修正进一步提升预测精度。

0.2%。

表 4 NYU Depth v2 数据集上的试验结果对比

Tab. 4 Comparison of test results on NYU Depth v2 dataset

方法	误差			准确率		
	Rel	RMSE	log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
文献[2]	0.213	0.759	0.087	0.650	0.906	0.976
文献[3]	0.127	0.573	0.055	0.811	0.953	0.988
文献[5]	0.141	0.528	0.060	0.815	0.951	0.985
文献[6]	0.160	0.604	-	0.771	0.946	0.986
文献.[9]	0.158	0.641	-	0.769	0.950	0.988
文献.[14]	0.115	0.509	0.051	0.828	0.965	0.992
本文模型	0.139	0.474	0.057	0.823	0.964	0.994

将本文构建模型与文献[5]的模型进行定性评价, 得到的预测深度图如图 4 所示。从图 4 可以看出, 采用本文模型, 小型物体轮廓明显边缘更加锐

利,远距离的深度预测效果有较大改善,说明本文模型对于提取目标轮廓边缘有显著提升,解决了小型

物体轮廓模糊的问题,同时提高了深度估计的性能,具有更好的预测深度能力。

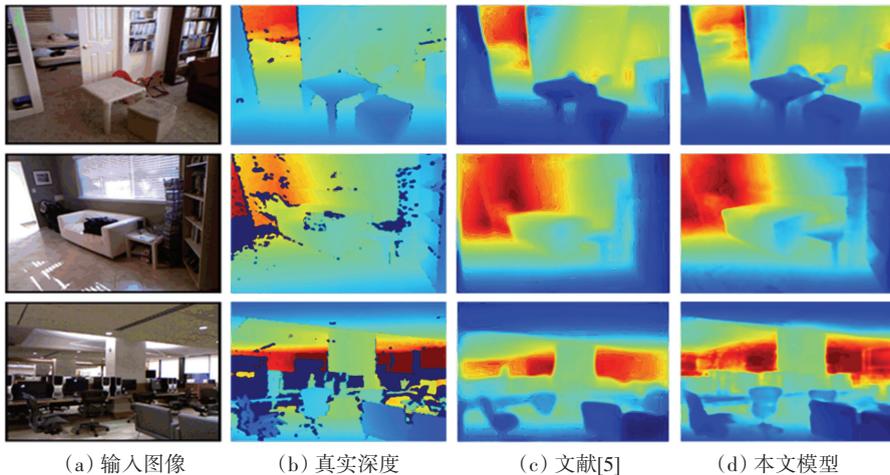


图4 深度预测图对比

Fig. 4 Comparison of different depth maps

## 5 结束语

针对现有的深度估计任务中存在的编码器模型参数量大和计算复杂度高问题,本文构建了一种基于轻量化融合损失网络的单目深度估计模型,并利用 NYU-Depth V2 公开数据集开展了试验研究。试验结果表明,该模型中损失网络汲取到了图像的高级特征信息,能对预测深度与真实深度进行融合感知,提高模型对信息的重构使用,弥补编码-解码阶段特征信息表达能力不足的问题;该模型不仅误差更小,准确率更高,而且优于现有的部分深度估计模型。

## 参考文献

- [1] 王科俊, 赵彦东, 邢向磊. 基于深度学习的无人机单目视觉避障算法[J]. 智能系统学报, 2018, 13(1): 55-69.
- [2] EIGEN D, PUHRSCHE C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[C]//Proceedings of the Proceedings of the 27<sup>th</sup> International Conference on Neural Information Processing Systems - Volume 2. Cambridge, MA, USA: MIT Press, 2014: 2366 - 2374.
- [3] LIU F, SHEN C, LIN G, et al. Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields [J]. IEEE Trans Pattern Anal Mach Intel, 2016, 38(10): 2024-2039.
- [4] 张蕾, 王园宇, 张文涛. 融合感知损失的单幅雾霾图像深度估计[J]. 计算机工程与科学, 2022, 44(3): 486-493.

- [5] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks [C]// Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), USA: IEEE Press, 2016: 239-248.
- [6] 贾瑞明, 李彤, 李阳, 等. 轻量金字塔解码结构的单目深度估计网络[J]. 计算机应用研究, 2021, 38(1): 293-297.
- [7] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision [EB/OL]. (2017-04-17) [2022-05-17]. <https://arxiv.org/abs/1704.04861>.
- [8] WOFK D, MA F, YANG T J, et al. Fastdepth: Fast monocular depth estimation on embedded systems [C]//Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), USA: IEEE Press, 2019: 6101-6108.
- [9] 胡坤, 陈迟晓, 李伟, 等. 深度估计网络的可学习步长轻量化研究 [J]. 小型微型计算机系统, 2022, 43(1): 50-54.
- [10] TAN M, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks [EB/OL]. (2019-05-28) [2022-05-17]. <https://arxiv.org/abs/1905.11946>.
- [11] KIM D, GA W, AHN P, et al. Global-local path networks for monocular depth estimation with verticalcutdepth [EB/OL]. (2022-01-19) [2022-05-17]. <https://arxiv.org/abs/2201.07436v1>.
- [12] 张意, 阚子文, 邵志敏, 等. 基于注意力机制和感知损失的遥感图像去噪[J]. 四川大学学报(自然科学版), 2021, 58(4): 45-55.
- [13] 郭志涛, 苏逸, 袁金丽, 等. 基于双注意力机制和复合损失的 LDCT 去噪方法[J]. 激光与光电子学进展, 2022, 59(2): 154-165.
- [14] 张顺然, 吴克伟, 洪炎. 基于稠密连接网络的单目深度估计[J]. 智能计算机与应用, 2020, 10(6): 42-47, 50.