文章编号:1671-4598(2025)04-0262-08 DOI:10.16526/j.cnki.11-4762/tp.2025.04.035 中图分类号:TP391.4 文献标识码:A

基于边缘设备的快速单目深度估计算法研究

王文帅[1]，韩 军[1]，邹小燕[2]，倪源松[1]，胡广怡[1]
(1. 上海大学 通信与信息工程学院，上海 200444；
2. 浙江华云信息科技有限公司，杭州 310051)

摘要：单目深度估算采用单一相机，安装方便，在机器人、无人机领域有广泛的应用；由于单目深度估计算法采用基于编码一解码的复杂的深度神经网络结构会导致边缘设备实时推理效率较低的问题，进而提出了一种可以在边缘设备上实时深度估计的网络架构；该架构采用倒置残差块设计的编码端，采用残差深度可分离卷积与最近邻插值重新设计的解码端，大大减少了模型的参数和计算量，并通过跨层连接将编码网络的特征与解码网络的特征相融合增强深度图中物体的边缘细节信息；实验结果表明，提出的网络架构参数量减少了82%，计算量减少了92%，在KITTI数据集上达到了先进的性能，并且在Jetson TX2上推理速度达到了50 FPS。

关键词：深度感知；单目相机；边缘设备；倒置残差；神经网络

Research on Fast Monocular Depth Estimation Algorithm Based on Edge Devices

WANG Wenshuai[1], HAN Jun[1], ZOU Xiaoyan[2], NI Yuansong[1], HU Guangyi[1]
(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China;
2. Zhejiang Huayun Information Technology Company, Hangzhou 310051, China)

Abstract: Monocular depth estimation, employing a single camera with easy installation, is widely applied in the fields of robotics and unmanned aerial vehicles. However, it adopts complex depth neural network structures based on encoder-decoder architectures in monocular depth estimation algorithms, which results in lower real-time inference efficiency on edge devices. Consequently, a network architecture is proposed to enable real-time depth estimation on edge devices. This architecture adopts an encoder designed with inverted residual blocks and a decoder redesigned with residual depth-wise separable convolution and nearest neighbor interpolation, significantly reducing the model's parameters and computational load. Moreover, through cross-layer connections, the features from the encoder and decoder networks are fused to enhance the representation of fine-grained edge details in the depth map. Experimental results show that the proposed network structure has an reduction of 82% in model parameters and a reduction of 92% in computational load, achieving state-of-the-art performance on the KITTI dataset. Notably, the proposed architecture achieves a real-time inference speed of 50 frames per second (FPS) on the Jetson TX2 platform.

Keywords: depth perception; monocular camera; edge device; inverted residual; neural network

0 引言

从单幅图像预测深度是当前研究的热点课题，因为它提供了一个多维度的信息，使机器能够更好地感知世界。现有的深度传感器，例如激光雷达、结构光传感器等，通常体积大、重量大且价格高昂，这些限制使它们

收稿日期:2024-01-17； **修回日期**:2024-02-23。
基金项目:国家自然科学基金项目(62371278,62371279)。
作者简介:王文帅(1996-),男,硕士研究生。
通讯作者:韩 军(1965-),男,博士,副教授,硕导。
引用格式:王文帅,韩 军,邹小燕,等.基于边缘设备的快速单目深度估计算法研究[J].计算机测量与控制,2025,33(4):262-269.

不适合在边缘设备中使用。单目深度估计（MDE，monocular depth estimation）是智能系统中获取场景深度的一种低成本方式。单目深度估计也是边缘计算、自动驾驶和机器人定位等需要实现的任务。随着边缘设备的日益普及，在边缘设备上进行快速深度估计任务变得很有必要。通常边缘设备体积较小，内存和计算能力也受到限制[1]。因此有必要研究一种轻量级、精确的单目深度估计方法。

当前 MDE 的研究方法分为两类：1）有监督的 MDE；2）无监督的 MDE。早期的研究基于有监督学习，通过应用真实深度值来计算损失，进而训练模型。第一个有监督学习方法是文献［2］提出的，它包括一个用于执行全局预测的粗尺度网络和一个用于完善局部预测的细尺度网络。文献［3］提出了一个更深的残差网络编码器和包含 Up-projection（反卷积）的解码器，并采用反向 Huber 损失进行单目深度估计。文献［4］设计了一种新颖的局部平面引导层的网络架构，通过使用局部平面引导层引导每个特征，有效地将输入特征引导到所需深度，提高了网络预测的性能。文献［5］提出较先进的基于有监督学习的深度估计方法，引入了自适应单元宽度估计器模块，将深度划分为单元，其中心值根据图像自适应估计。接着使用单元中心值的线性组合来计算深度。尽管有监督的深度估计能够实现较高的深度预测精度，但需要大规模的地面真实深度标签，并且建立这些标签值成本较高。为了摆脱这个问题，研究人员开始探索 MDE 的无监督学习方法。

无监督学习模型在训练时的输入可以是单目视频图像序列或立体图像对，立体图像对最初用于训练无监督学习模型[6-9]，这些模型使用同一场景的左右两张图像，通过计算这两个图像对应像素之间的位移来计算视差图。近几年的研究[10-14]提出了使用单目视频图像序列的无监督学习模型。经典的无监督深度估计算法 Monodepth2[12] 的深度图预测是基于姿态估计网络（PoseNet）与深度估计网络（DepthNet），由于 Monodepth2 方法不仅是近几年无监督 MDE 方法的 SOTA 模型，而且该方法也是 MDE 无监督学习方法中广泛使用的基线模型。PoseNet 用于目标图像的重建，DepthNet 根据 PoseNet 的输出预测深度图。允许模型完全用单目视频图像序列进行训练，虽然在训练过程中同时使用深度网络和姿态网络，但在测试过程中它们可以分开运行。Monodepth2[12] 提出了最小光度误差损失函数，提高了物体遮挡边界的清晰度，从而显著提高了模型的精度。但是 Monodepth2 网络使用参数量和计算量均较大的 DNN 网络架构。文献［15］提出了一种特征金字塔网络架构 PyD-Net，用于 CPU 上快速推断深度图，该方法参数量减少为 1.9 M，但该方法模型精度较低，深度图较模糊。文献［16］提出了一种基于 Jetson TX2 板的实时轻量级架构，修剪后网络使用的参数总数为 1.34 M，但该算法基于室内场景，没有考虑室外场景，并且该算法输入图片的分辨率较低，模型推理精度较低。文献［17］通过使用网络修剪方法去除训练模型中不重要特征，进而开发轻量级的单目深度模型，该模型参数总量为 5.9 M，但模型推理精度较差（精度仅有 0.827），深度图物体边缘也较模糊。文献［18－20］中提出的深度估计方法也存在网络精度较低和深度图较模糊等不足。

为了提高 Monodepth2[12] 基线模型的深度估算效率，采用倒置残差块重新设计 Monodepth2 网络中深度估计网络（DepthNet）编码端，优化网络的编码效率；使用残差深度可分离卷积与最近邻插值重新设计解码端，最后采用 UNet 网络的跳跃连接，使解码端融合了编码端每个阶段的特征，使得模型最终生成的深度图中物体边缘细节更加清晰。在嵌入式 Jetson TX2 边缘设备上来验证优化的算法，最终确保单目深度估计的精度并实现低延迟深度估算。

1　优化的单目深度估计方法

优化的无监督单目图像视频帧快速深度估计方法是由一个深度网络（DepthNet）和一个姿态网络（PoseNet）组成。DepthNet 用于估计输入单目图像的深度图，PoseNet 用于估计两对输入相邻单目图像视频帧之间相机的相对姿态信息，如图 1 所示。

图 1 中采用相对较深的 ResNet-18 网络作为 PoseNet 编码端部分，准确估计的相机相对姿态信息对于 DepthNet 进行准确的深度预测很重要，PoseNet 解码端采用具有 ReLU 非线性激活函数的常规二维卷积层。在训练阶段，将 3 个连续的单目图像视频帧，送入网络，其中为目标图像，和为源前后图像帧，步骤如下：

1）将 I_t 目标图像输入到图 1 所示的 DepthNet 中，预测输出 I_t 目标图像的深度图 D_t。将 I_t 与 I_{t-1} 和 I_t 与 I_{t+1} 输入到 PoseNet 中，预测输出 I_t 与 I_{t-1} 和 I_t 与 I_{t+1} 之间具有 6-DoF 的相机相对姿态信息 $T_{t\to t-1}$ 和 $T_{t\to t+1}$。

2）联合生成目标图像 I_t 的深度图 D_t 和每个源图像 I_s 的相对相机位姿 $T_{t\to s}$，通过公式（1）计算变换，将源图像 I_s 中的像素 p 投影到目标图像 I_t 对应的像素 p' 上。

$$p' = KT_{t\to s}D_t(p)\,K^{-1}\,p \tag{1}$$

其中：$s\in\{t-1,\ t+1\}$，K 表示相机的内参。

3）通过双线性插值算法，从源图像 I_s 重建目标图像 I_t，计算过程如公式（2）所示：

$$I_s^w = I_s[p'] \tag{2}$$

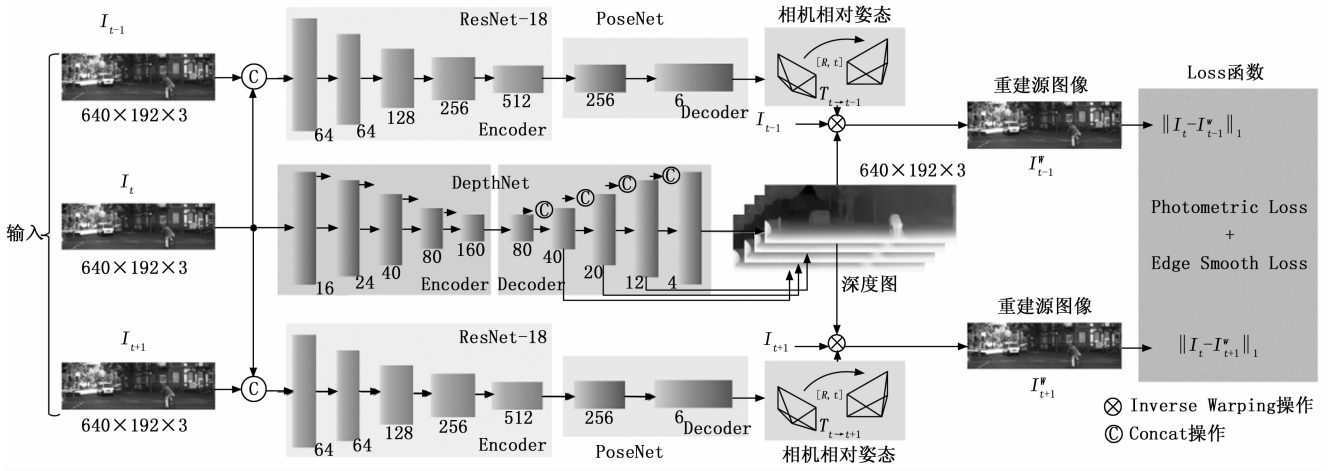


图 1 单目视频帧深度估计原理总框图

其中：[·] 为双线性插值操作，I_s^w 为源图像 I_s 重建的目标图像。

4) 源图像重建的目标图像与目标图像进行最小光度误差损失计算。DepthNet 生成的深度图进行边缘感知平滑损失计算，采用最小光度误差损失与边缘感知平滑损失以一定的权重比例得到的联合损失函数进行训练模型，联合损失代替了监督学习中 Groundtruth 的作用。

5) 训练过程中，每一组 batchsize 的数据将损失反向传播，调整网络参数来优化生成的深度图，通过不断往复直至联合损失降到最小停止训练。

在测试阶段，DepthNet 根据输入的单目视频帧图像生成对应的深度图，因此，DepthNet 的复杂性决定了单目深度估计算法推理时间的长短。目前的无监督方法通过设计更加复杂的 DepthNet 网络提高了算法的精确度，但耗时的问题也更加凸显。Monodepth2[12] 算法的 DepthNet 网络编码端采用 Resnet-18 网络，参数量为 11.2 M，计算量为，解码端采用 5 层 3×3 常规卷积块，参数量为 3.2 M。对于当前有 GPU 服务器设备，其网络结构并不复杂，有足够的计算和存储资源来处理这样规模的网络，能够在这些设备上进行实时推理。然而在边缘设备上通常受限于有限的计算资源和内存容量，这个网络结构较复杂。因此需要对网络进行进一步的优化，以确保在边缘设备上能够实现实时深度估算。

1.1 改进的 DepthNet 网络

针对当前无监督单目深度估计 DepthNet 网络存在的复杂性，提出了一种快速单目深度估计 DepthNet 网络模型，如图 2 所示。模型采用倒置残差块来设计编码端，采用残差深度可分离卷积与最近邻插值来设计解码端。采用倒置残差块进行图像特征提取，避免了采用过多下采样导致图像细节信息丢失的情况发生，解决了编码端网络参数量和计算量较大的问题。采用残差深度可分离卷积设计的解码端进一步降低了 DepthNet 网络架构的参数量，与使用反卷积进行上采样解码操作不同，该架构采用最近邻插值算法进行上采样，不仅不增加额外的参数量而且解码速度较快，同时该架构也加入了从编码端到解码端的跳跃连接来融合编码端低层多尺度特征以提升生成深度图的质量。

1.1.1 编码端网络的优化

如图 2 所示，优化的 DepthNet 编码端由一个标准

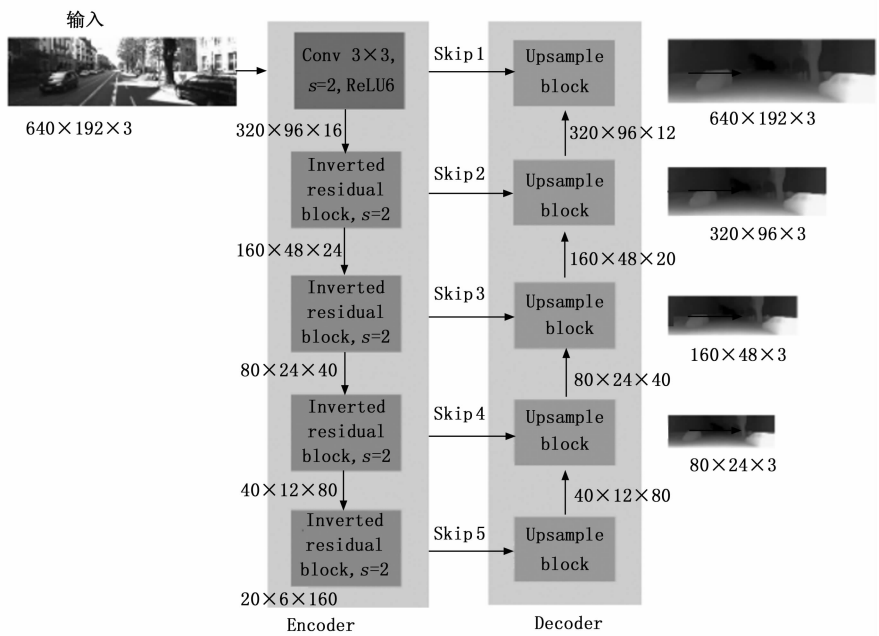


图 2 DepthNet 网络模型

卷积层和 4 个倒置残差块组成。编码端的第一层为标准的 33 卷积，步长为 2，采用 ReLU6 整数运算激活函数，该激活函数计算代价较低，更适用于嵌入式或移动设备。编码端的其余层为倒置残差块，该模块的基础组成如图 3 所示。

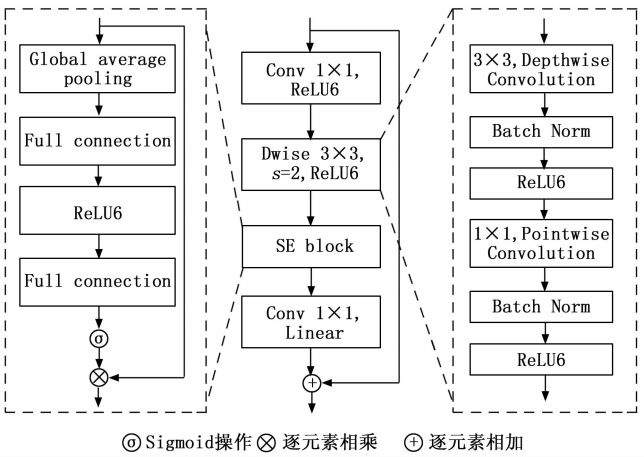


图 3　倒置残差块基础组成

采用了线性瓶颈结构减少网络参数量和计算量。首尾均为 1×1 卷积层，首部卷积层用于特征的升维，尾部卷积层用于特征的降维，能更好地整合编码端提取的特征信息。中间部分为 3×3 深度可分离卷积，包括 3×3 深度卷积和 11 逐点卷积，采用 ReLU6 激活函数。

常规卷积和深度可分离卷积参数量和计算量计算，如公式（3）：

$$\begin{cases} Parameter_{\text{conv}} = M\times N\times D_k\times D_k \\ Parameter_{\text{Dwiseconv}} = M\times D_k\times D_k + M\times N \\ \dfrac{Parameter_{\text{conv}}}{Parameter_{\text{Dwiseconv}}} = \dfrac{M\times N\times D_k\times D_k}{M\times D_k\times D_k + M\times N} = \dfrac{1}{\dfrac{1}{N}+\dfrac{1}{D_k^2}} \approx D_k^2 \\ FLOPs_{\text{conv}} = D_k\times D_k\times M\times N\times D_F\times D_F \\ FLOPs_{\text{Dwiseconv}} = D_F\times D_F\times M(D_k\times D_k + N) \\ \dfrac{FLOPs_{\text{conv}}}{FLOPs_{\text{Dwiseconv}}} = \dfrac{D_k\times D_k\times M\times N\times D_F\times D_F}{D_F\times D_F\times M(D_k\times D_k+N)} = \dfrac{1}{\dfrac{1}{N}+\dfrac{1}{D_k^2}} \approx D_k^2 \end{cases} \tag{3}$$

其中：M 为输入通道数，N 为输出通道数，D_k 为卷积核大小，D_F 为输入特征图尺寸。$Parameter_{\text{conv}}$ 为常规卷积参数量，$Parameter_{\text{Dwiseconv}}$ 为深度可分离卷积参数量，$\dfrac{Parameter_{\text{conv}}}{Parameter_{\text{Dwiseconv}}}$ 为两者参数量的比值，$FLOPs_{\text{conv}}$ 为常规卷积计算量，$FLOPs_{\text{Dwiseconv}}$ 为深度可分离卷积计算量，$\dfrac{FLOPs_{\text{conv}}}{FLOPs_{\text{Dwiseconv}}}$ 为两者计算量的比值。在相同卷积核大小和输入特征尺寸条件下，常规卷积的参数量和计算量均为深度可分离卷积的 D_k^2 倍，因此采用倒置残差块的编码端大大减少了参数量和计算量。原 DepthNet 网络编码端与改进后编码端网络对比，如表 1 所示。

表 1　原编码端与改进后编码端网络对比

Layer name	原编码端	改进后编码端
conv1	7×7,64,s=2	3×3,16,s=2
conv2	3×3maxpool,s=2	
(3×3,64)×4	(倒置残差块)×2	
conv3	(3×3,128)×4	(倒置残差块)×2
conv4	(3×3,256)×4	(倒置残差块)×3
conv5	(3×3,512)×4	(倒置残差块)×8
FLOPs	1.8×10^9	1.4×10^8
Parameters/M	11.2	2.8

深度可分离卷积之后为 SE（Squeeze and Excitation Network）网络模块，该模块通过引入一个全局平均池化层，将每个通道的特征图降为一个单一的数值，进一步降低计算复杂度。接下来，将得到的每个通道的标量输入到一对全连接层中，这对全连接层用于学习通道之间的权重关系，即对每个通道的重要性进行建模，从而增强了网络对重要特征的感知能力。由于深度估计网络依赖于像素信息，采用 SE 通道注意力网络通过关注重要特征信息和空间信息，保证了 DepthNet 网络预测的精确度。

1.1.2　解码端网络的优化

在解码阶段利用编码端的语义特征来推断高质量的深度图。为了满足高精度和实时性的需求，设计了一种如图 4 所示的新颖高效上采样模块。提出的轻量级上采样块由 3 个残差深度可分离卷积块、最近邻上采样组成。这些残差深度可分离卷积块是由深度卷积（Depthwise Convolution）和逐点卷积（Pointwise Convolution）组成，参数计算与深度可分离卷积相同。采用最近邻插

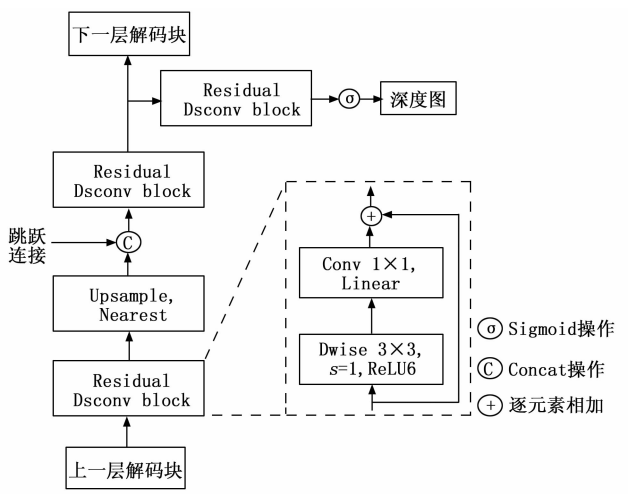


图 4　解码端上采样块

值算法进行上采样，不使用传统的反卷积进行上采样，避免了增加额外的参数数量，并且解码效率也比传统的反卷积高。改进后解码端网络由 5 个上采样块组成，对编码端的输出特征图进行上采样和聚合跳跃连接的特征信息。原 DepthNet 网络解码端与改进后解码端网络对比，如表 2 所示，大大减少了解码端的参数数量和运算成本。

表 2 原解码端与改进后解码端网络对比

Layer name	原解码端	改进后解码端
conv1	(3×3,16)×3	(解码端上采样块)×1
conv2	(3×3,32)×3	(解码端上采样块)×1
conv3	(3×3,64)×3	(解码端上采样块)×1
conv4	(3×3,128)×3	(解码端上采样块)×1
conv5	(3×3,256)×3	(解码端上采样块)×1
Parameters/M	3.2	0.21

1.1.3 跳跃连接的应用

编码端网络通常包含许多层以逐渐降低空间分辨率并从输入中提取更高级别的特征。而传统的编码端到解码端的输出变成了一组低分辨率特征，其中可能会丢失许多图像细节，从而使解码端更难恢复像素级（密集）数据。如图 2 所示，在编码端－解码端网络架构中使用跳跃连接（Skip1-Skip5），能够使得网络在每一级的上采样过程中，将编码端对应位置的特征图在通道上进行融合。通过底层特征与高层特征的融合，解码端能够保留更多高层特征图蕴含的高分辨率细节信息，从而提高了单目深度估计深度图的质量。

1.2 损失函数

与文献［12］类似，采用最小光度误差损失和边缘感知平滑损失联合监督 DepthNet 和 PoseNet 网络的训练过程。光度误差由公式（4）计算，其由结构相似性指数 $SSIM$ 和 L_1 范数组成，$SSIM$ 用于比较源图像 I_t 与重建图像 I_s^w 之间的结构相似性，α 设置为 0.85。

$$\mathcal{L}_p(I_t,I_s^w)=\alpha\frac{1-SSIM(I_t,I_s^w)}{2}+(1-\alpha)\left\|I_t,I_s^w\right\|_1 \tag{4}$$

为了生成平滑的深度图，同时保留图像不连续区域的锐利边缘，加入边缘感知平滑损失函数，如公式（5）所示：

$$\mathcal{L}_s=|\partial_x d_t^*|e^{-|\partial_x I_t|}+|\partial_y d_t^*|e^{-|\partial_y I_t|} \tag{5}$$

∂_x，∂_y 为源图像的空间梯度，d_t^* 为平均归一化视差。总损失通过上述两个损失函数以一定比例组合得到，如公式（6）所示，λ 设置为 0.001。

$$\mathcal{L}_{\text{loss}}=L_p+\lambda L_s \tag{6}$$

2 实验过程及结果

在本节中，将展示实验结果来证明提出的优化方法的可行性。通过各种编码端和解码端选项的比较，并根据准确性和延迟指标对其进行分析。

2.1 实验设置

采用 KITTI[21] 数据集，所提出的网络使用 PyTorch 实现，并在批量大小为 12 的单个 NVIDIA GeForce RTX 2080 Ti 上进行训练。1×10^{-4} 采用的初始学习率，训练轮数设置为 30 轮。为了提高训练的稳健性，采用了以下数据增强策略作为预处理步骤：以 50％的几率对训练数据集进行水平翻转、亮度调整（±0.2）、饱和度调整（±0.2）、对比度调整（±0.2）、色调抖动（±0.1），这些调整以随机顺序应用于训练集。

2.2 评估指标

为了便于评估本文方法与其他在 KITTI 数据集上进行训练的相关工作性能，采用了文献［2］中常用的评价指标对算法性能进行分析。其中，d_i^* 和 d_i 分别为像素的预测深度值和真实深度值，N 为图像的像素总数。评价指标定义如下：

1）绝对相对误差（Abs Rel，absolute relative error）：

$$Abs\ Rel=\frac{1}{N}\sum_{i=1}^{N}\frac{\left\|d_i^*-d_i\right\|}{d_i} \tag{7}$$

2）平方相对误差（Sq Rel，squared relative error）：

$$Sq\ Rel=\frac{1}{N}\sum_{i=1}^{N}\frac{\left\|d_i^*-d_i\right\|^2}{d_i} \tag{8}$$

3）均方根误差（RMSE，root mean squared error）：

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i^*-d_i)^2} \tag{9}$$

4）均方根对数误差（RMSE log，root mean squared logarithmic error）：

$$RMSE\ \log=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log d_i^*-\log d_i)^2} \tag{10}$$

5）准确率（Accuracy）：

准确率为满足如下条件的 d_i^* 的百分比：$\delta=max\left(\frac{d_i^*}{d_i},\frac{d_i}{d_i^*}\right)<thr$，$thr=1.25,1.25^2,1.25^3$。

2.3 实验结果与分析

2.3.1 不同编码端对比实验

本节基于 DepthNet 编码端，对比了 6 种不同的编码端网络在 KITTI 数据集上测试的数据结果，输入单目视频帧分辨率为 640×192，如表 3 所示。

表 3 不同编码端对比实验

编码端	*Parameters*/M	*Abs Rel* ↓	*RMSE* ↓	$\delta_1<1.25$ ↑
Swin TransformerV2	88	0.137	5.271	0.837
EfficientNetV2	24	0.121	4.991	0.862
Resnet-18	11.2	0.119	4.894	0.872
FasterNet	3.9	0.128	5.068	0.858
EGE-UNet	0.2	0.141	5.372	0.835
本文方法	2.8	0.117	4.929	0.870

从表 3 中可以看出，改进后编码端网络相较于原 Resnet-18 网络，参数量减少了 75%，精度指标并未下降，优于其他编码端网络。

2.3.2　不同解码端对比实验

本节基于 DepthNet 解码端，对比分析了两种不同的图像插值方法对 DepthNet 网络的影响，以及改进后解码端与原解码端对比实验结果，如表 4 和表 5 所示。

表 4　不同图像插值方法对比实验

方法	*Abs Rel* ↓	*RMSE* ↓	$\delta_1<1.25$ ↑
Bilinear	0.122	5.025	0.865
Nearest	0.119	4.894	0.872

从表 4 中可以看出，最近邻插值上采样方法在评价指标上均优于双线性插值上采样方法，因此解码端采用最近邻插值进行上采样。

表 5　改进后解码端与原解码端对比实验

解码端	*Parameters*/M	*Abs Rel* ↓	*RMSE* ↓	$\delta_1<1.25$ ↑
原解码端	3.2	0.119	4.894	0.872
本文方法	0.21	0.120	4.910	0.865

从表 5 中可以看出，改进后解码端参数量减少了 93%，评价指标有所下降，但相对于网络的复杂度，精度指标在可以接受的范围内。

2.3.3　边缘设备推理时间对比实验

本节对比了在 Jetson TX2 边缘设备上 Monodepth2 算法与本文方法在不同输入单目视频帧分辨率下的推理时间，其中 ARM、ALL、MAXQ、MAXN 为 TX2 的 4 种功耗模式，如表 6 所示。

表 6　网络推理时间对比实验　　ms

方法	分辨率	ARM	ALL	MAXQ	MAXN
Monodepth2	640×192	45	47	59	42
	1 024×320	100	103	132	86
本文方法	640×192	20	31	32	26
	1 024×320	40	45	54	42

从表 6 中可以看出，提出的方法在 640×192 分辨率下，网络的推理速度达到了 50 FPS，相较于 Monodepth2 的 22 FPS，推理时间减少了 56%；在 1 024×320 分辨率下，网络的推理速度达到了 25 FPS，相较于 Monodepth2 的 10 FPS，推理时间减少了 60%，满足了当前在边缘设备上进行实时单目深度估计的要求。

2.3.4　不同方法对比实验结果

为了对本文提出的方法进行量化评估，本节对比了几种无监督学习单目深度估计算法的实验结果，如表 7 所示。

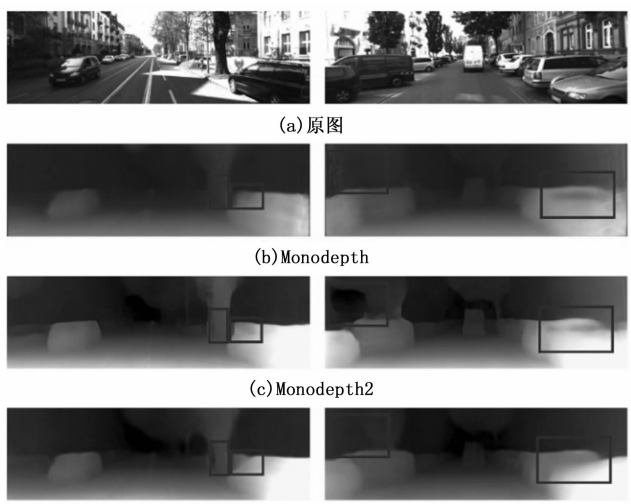
表 7　不同方法对比实验

方法	*Abs Rel* ↓	*Sq Rel* ↓	*RMSE* ↓	*RMSE* log ↓	$\delta_1<1.25$ ↑	$\delta_1<1.25^2$ ↑	$\delta_1<1.25^3$ ↑
Monodepth	0.124	1.076	5.311	0.219	0.847	0.942	0.973
PyD-Net	0.153	1.363	6.030	0.252	0.789	0.918	0.963
Monodepth2	0.119	0.908	4.894	0.195	0.872	0.958	0.981
文献[10]	0.183	1.595	6.709	0.270	0.734	0.902	0.957
文献[18]	0.147	1.317	5.826	0.229	0.815	0.935	0.971
文献[22]	0.133	1.126	5.515	0.231	0.826	0.934	0.969
本文方法	0.119	0.918	4.915	0.194	0.868	0.958	0.982

从表 7 中可以看出，与 Monodepth2 相比，提出的方法仅在 *Sq Rel*、*RMSE* 和评价指标上略有下降，但与整体网络架构的复杂度相比，略有下降的指标可以忽略不计。与轻量级单目深度估计 PyD-Net 和文献[18]方法相比，提出的方法精度分别比其高 7.9%和 5.3%。提出的方法在各个评价指标上均优于 Monodepth、文献[10]和[22]其余无监督学习方法。

2.3.5　模型可视化结果

为了验证提出的网络架构生成的深度图质量并未变差，定性对比了 Monodepth、Monodepth2 以及提出方法生成的深度图，如图 5 所示。



(a)原图　(b)Monodepth　(c)Monodepth2　(d)本文方法

图 5　模型可视化结果

从图 5 中可以看出，Monodepth 方法生成的深度图质量最差，提出的方法与 Monodepth2 方法生成的深度图相比，在树的形状、指示牌以及车辆的轮廓等物体细节信息上优于 Monodepth2 方法。

为了定性地评估模型的泛化性，基于校园场景采集了以下图像数据，如图 6 所示。图像中人体和标定板的距离是从 2～10 m 之间渐变的，从图中可以看出，草坪、树木、车辆以及人体都可以很好地被感知到。

图 6 模型泛化性可视化结果

3 结束语

针对目前无监督单目深度估计网络模型结构复杂、参数量和计算量大、部署在边缘设备上导致推理时间长、不具有实时性等问题，提出了一种优化的编解码网络架构。该架构具有低复杂性和低延迟的编解码网络，使得编码端参数量减少了 75%，计算量减少了 92%，解码端参数量减少了 93%。在 NVIDIA Jetson TX2 上，提出的网络模型的推理时间达到了 50 FPS，满足了实时性的要求。同时在 KITTI 数据集上进行了广泛的实验，表明了模型的精确度并未下降。提出的方法由于轻量级和实时性的优点，可以与目标检测或者图像分割领域相结合，从而获取图像中物体的空间位置信息，该并行任务值得在未来进一步研究。

参考文献：

[1] MAO S, ZHANG N, LIU L, et al. Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks [J]. IEEE Transactions on Vehicular Technology, 2021, 70 (10): 10820 - 10831.

[2] EIGEN D, PUHRSCH C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network [J]. Advances in Neural Information Processing Systems, 2014, 27: 2366 - 2374.

[3] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks [C] //2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 239 - 248.

[4] LEE J H, HAN M K, KO D W, et al. From big to small: Multi-scale local planar guidance for monocular depth estimation. arXiv 2019 [J]. Arxiv Preprint Arxiv: 1907. 10326, 1907.

[5] BHAT S F, ALHASHIM I, WONKA P. Adabins: Depth estimation using adaptive bins [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4009 - 4018.

[6] GODARD C, MAC AODHA O, BROSTOW G J. Unsupervised monocular depth estimation with left-right consistency [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 270 - 279.

[7] ZHOU C, ZHANG H, SHEN X, et al. Unsupervised learning of stereo matching [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017: 1567 - 1575.

[8] SMOLYANSKIY N, KAMENEV A, BIRCHFIELD S. On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018: 1007 - 1015.

[9] SKINNER K A, ZHANG J, OLSON E A, et al. Uwstereonet: unsupervised learning for depth estimation and color correction of underwater stereo imagery [C] //2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 7947 - 7954.

[10] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1851 - 1858.

[11] BIAN J, LI Z, WANG N, et al. Unsupervised scale-consistent depth and ego-motion learning from monocular video [J]. Advances in Neural Information Processing Systems, 2019, 32: 35 - 45.

[12] GODARD C, MAC AODHA O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3828 - 3838.

[13] RANJAN A, JAMPANI V, BALLES L, et al. Competitive collaboration: joint unsupervised learning of depth, camera motion, optical flow and motion segmentation [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12240 - 12249.

[14] ZHANG Y, XU S, WU B, et al. Unsupervised multi-view constrained convolutional network for accurate depth estimation [J]. IEEE Transactions on Image Processing, 2020, 29: 7019 - 7031.

[15] POGGI M, ALEOTTI F, TOSI F, et al. Towards real-

time unsupervised monocular depth estimation on cpu [C] //2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 5848 - 5854.

[16] WOFK D, MA F, YANG T J, et al. Fastdepth: fast monocular depth estimation on embedded systems [C] // 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 6101 - 6108.

[17] ELKERDAWY S, ZHANG H, RAY N. Lightweight monocular depth estimation model by joint end-to-end filter pruning [C] //2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 4290 - 4294.

[18] LIU S, ZHAO S, ZHANG P, et al. Real-time monocular depth estimation for low-power embedded systems using deep learning [J]. Journal of Real-Time Image Processing, 2022, 19 (5): 997 - 1006.

[19] LIU J, KONG L, YANG J. Designing and searching for lightweight monocular depth network [C] //International Conference on Neural Information Processing. Cham: Springer International Publishing, 2021: 477 - 488.

[20] LEE Y, LEE S, KO J G. Monocular depth estimation for mobile device [C] //2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). IEEE, 2021: 1 - 3.

[21] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the kitti dataset [J]. The International Journal of Robotics Research, 2013, 32 (11): 1231 - 1237.

[22] WONG A, SOATTO S. Bilateral cyclic constraint and adaptive regularization for unsupervised monocular depth prediction [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5644 - 5653.

(上接第 261 页)

[9] NARA T, SUZUKI S, ANDO S. A closed-form formula for magnetic dipole localization by measurement of its magnetic field and spatial gradients [J]. IEEE Transactions on Magnetics, 2006, 42 (10): 3291 - 3293.

[10] YIN G, ZHANG Y T, FAN H B, et al. Magnetic dipole localization based on magnetic gradient tensor data at a single point [J]. J. Appl. Rem. Sens., 2014, 8 (1): 083596.

[11] SUI Y, LESLIE K, CLARK D. Multiple-order magnetic gradient tensors for localization of a magnetic dipole [J]. IEEE Magnetics Letters, 2017, 8: 1 - 5.

[12] 张樱子，刘改改，申雅丽，等. 基于归一化磁源强度的磁目标实时定位方法 [J]. 测试技术学报，2024，38 (5)：535 - 542.

[13] 李青竹，李志宁，张英堂，等. 基于二阶磁张量欧拉反褶积的磁源单点定位方法 [J]. 石油地球物理勘探，2019，54 (4)：915 - 924.

[14] LIU G, ZHANG Y, LIU W. Structural design and parameter optimization of magnetic gradient tensor measurement system [J]. Sensors, 2024, 24 (13): 4083.

[15] 赵　震，杨宾峰，王　润，等. 基于双十字形测量结构的磁信标定位方法 [J]. 传感技术学报，2021，34 (1)：70 - 74.

[16] WIEGERTR, LEE K, OESCHGER J. Improved magnetic STAR methods for real-time, point-by-point localization of unexploded ordnance and buried mines [J]. OCEANS, 2008: 1 - 7.

[17] 张樱子，邱隆清，荣亮亮，等. 超导磁梯度张量探测系统单点定位方法研究 [J]. 低温与超导，2023，51 (8)：12 - 16.

[18] CHI C, WANG D, TAO R, et al. Two-Point localization algorithm of a magnetic target based on tensor geometric invariant [J]. Sensors, 2024, 24 (7): 2224.

[19] 张朝阳，肖昌汉，高俊吉，等. 磁性物体磁偶极子模型适用性的试验研究 [J]. 应用基础与工程科学学报，2010，18 (5)：862 - 868.

[20] FAN L. A fast linear algorithm for magnetic dipole localization using total magnetic field gradient [J]. IEEE Sensors Journal, 2018, 18 (3): 1032 - 1038.

[21] 张　仑，张晓明，马喜宏，等. 基于两点磁梯度张量不变量的目标定位法 [J]. 电子设计工程，2023，31 (12)：6 - 10.

[22] HE G X, HE T J, LIAO K X, et al. Experimental and numerical analysis of non-contact magnetic detecting signal of girth welds on steel pipelines [J]. ISA transactions, 2021: 681 - 698.

[23] SUI Y Y, WANG S L, MENG H, et al. An analysis and elimination of zero drift in magnetic gradient tensor exploration system [J]. IEEE, 2011, 94: 1 - 5.

[24] AGRAWALM, MISHRAM, S S P, et al. Association rules optimization using improved PSO algorithm [C] // 2015 International Conference on Communication Networks (ICCN), 2015: 395 - 398.

[25] WBIN. A novel supply chain multi-level inventory model based on improved PSO algorithm [C] //2023 8th International Conference on Communication and Electronics Systems (ICCES), 2023: 1733 - 1737.