Modeling and Simulation 建模与仿真, 2023, 12(5), 4642-4653
Published Online September 2023 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2023.125423

Hans汉斯

面向盲人避障的单目深度估计方法

杨　双，张荣芬*，刘宇红，刘　源，程娜娜，刘昕斐

贵州大学大数据与信息工程学院，贵州　贵阳

收稿日期：2023年7月19日；录用日期：2023年9月12日；发布日期：2023年9月19日

摘　要

盲人作为弱势群体，他们的衣食住行值得被人们关注，其中出行问题是造成盲人群体远离社会的重要原因。本文提出了一种基于DenseNet改进的深度估计算法，以解决盲人出行时无法感知周围障碍物的问题。首先，以DenseNet作为编码器的编解码过程中，信息丢失会造成深度估计不准确，为了减少这种问题在编码器与解码器的跳跃连接中引入RHAG残差混合注意力组，加强模型对细节特征的识别能力，提升模型恢复深度信息的准确性；然后，在解码出深度图后采用AdaBins后处理模块，对深度图进行优化，以更好地恢复出RGB场景的深度信息；最后通过ACB非对称卷积替换DenseNet中DenseBlock的卷积，通过增强卷积骨架，提升模型特征提取能力。实验结果表明，本文改进的算法与原网络相比，精度提升了约3.04%，均方根误差降低了约3.39%。与目前先进的深度估计网络MonoDepth相比，精度提升了约2.2%，绝对相对误差降低了约1.3%。本文算法在通过单张RGB图进行深度估计时能获取到更准确的深度信息，优于对比算法，且满足边缘计算设备的要求，具有一定的实用价值。

关键词

单目视觉，避障系统，深度估计，混合注意力组

A Monocular Depth Estimation Method for Blind Obstacle Avoidance

Shuang Yang, Rongfen Zhang*, Yuhong Liu, Yuan Liu, Nana Cheng, Xinfei Liu

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: Jul. 19th, 2023; accepted: Sep. 12th, 2023; published: Sep. 19th, 2023

Abstract

Blind people, as a vulnerable group, deserve attention for their clothing, food, housing, and trans-

*通讯作者。

文章引用: 杨双, 张荣芬, 刘宇红, 刘源, 程娜娜, 刘昕斐. 面向盲人避障的单目深度估计方法[J]. 建模与仿真, 2023, 12(5): 4642-4653. DOI: 10.12677/mos.2023.125423

portation, among which transportation issues are an important reason for the blind group to stay away from society. This article proposes an improved depth estimation algorithm based on DenseNet to solve the problem of blind people being unable to perceive surrounding obstacles when traveling. Firstly, information loss during the encoding and decoding process can lead to inaccurate depth estimation. In order to reduce this problem, RHAG residual mixed attention groups are introduced in the skip connection between the encoder and decoder to enhance the model's ability to recognize detailed features and improve the accuracy of the model in restoring depth information; then, the AdaBins post-processing module is used to optimize the depth map to better recover the depth information of RGB scenes; finally, ACB asymmetric convolution is used to replace the convolution of DenseBlock in DenseNet, enhancing the model's feature extraction ability by enhancing the convolution skeleton. The experimental results show that the accuracy of the improved algorithm is improved by 3.04% and the root-mean-square deviation is reduced by 3.39% compared with the original network. Compared with the current advanced depth estimation network MonoDepth, the accuracy has been improved by about 2.2% and the absolute relative error has been reduced by about 1.3%. The algorithm in this paper can obtain more accurate depth information when depth estimation is carried out through a single RGB image, which is superior to the comparison algorithm and meets the requirements of edge computing devices, and has certain practical value.

Keywords

Monocular Vision, Obstacle Avoidance System, Depth Estimation, Hybrid Attention Group

1. 引言

据统计，中国有 1730 万盲人，位列世界第一，全球范围内也有约 22 亿人患有视力障碍的疾病。出行是盲人融入社会需要面对的首要问题，而我国关于盲人的基础设施建设并不完善，使得盲人的出行受到很大限制，因此很多盲人选择封闭自己，不愿与外界联系，这就是盲人数量如此之多，我们却很难在日常生活中看到他们身影的原因。由于传统辅助手段如盲道、导盲犬等不能很好地解决盲人出行问题，基于计算机视觉的辅助方法迎来新的挑战和机遇，其中深度估计技术是解决盲人出行中避障问题的关键[1] [2]。

近年来，随着深度学习的飞速发展，计算机视觉技术也越发成熟，被更多地应用于实际生活中，如VR 虚拟现实、自动送货机器人、无人机自主飞行等[3] [4]。这些应用得以实现的基础之一便是深度估计，即估计相机获取的图片中每个像素点与相机之间的实际距离。传统深度估计的方法大致可分为两种，一是通过激光雷达获取精准的深度信息；二是通过双目摄像头获取同一物体不同视角的图片，进行立体匹配计算出深度信息。相对来说，激光雷达获取到的深度信息更为精确，但结构精密、造价昂贵，普通盲人家庭难以承受。双目摄像头虽然成本上降低了很多，但其需要复杂的相机标定，两个相机的视差也需要精确的匹配，考虑到盲人在生活中可能碰到的各种突发情况，其维护成本太高，也不利于推广。相比之下，单目深度估计成本更低，实际使用中更方便，所以近年来基于深度学习的单目深度估计技术被更多人关注[5]。

最早在 2014 年，Eigen 等人[6]提出了在深度估计中使用卷积神经网络(Convolutional Neural Network,

CNN)，利用全局粗尺度网络粗略估计场景深度，再用局部精细尺度网络优化局部信息。2015 年，Liu 等人[7]提出了深度卷积神经场模型对单张 RGB 图像的深度进行预测。2016 年，Laina 等人[8]首次将残差网络(ResNet)应用到深度估计领域，将 ResNet 作为编码器提取信息丰富的特征图，再由解码器逐步恢复场景深度信息。DenseNet [9]是以 ResNet 为基础的模型，受到 ResNet 的启发，它也被应用于深度估计领域。2017 年，Jung 等人[10]将生成式对抗网络引入深度估计领域，生成器根据场景图片生成深度图，判别器判别生成的深度图是否为真实深度。2021 年 RanFtl 等人[11]首次尝试使用自然语言处理领域的 Transformer [12]结构替换卷积神经网络提取特征，通过融合局部与全局注意力能够获得较准确的深度信息。2022 年，Yuan 等人[13]优化了条件随机场(Conditional Random Field, CRF)，通过将输入分割成多个窗口，构建全连接的 CRF，捕捉像素之间的更多关系，在深度估计问题上取得了良好的效果。

目前，为了满足高精度的要求，设计的网络结构越来越复杂，计算量较大，难以部署到用于辅助盲人出行的边缘设备上。轻量化网络对于恢复物体细节特征，精确获取深度信息上又略显吃力。为了平衡两者，以满足盲人出行辅助设备所需的实时性和准确性，本文提出了一种基于 DenseNet 的单目深度估计算法，主要贡献包括一下几点：

① 建立编码器和解码器之间的跳跃连接，并引入 RHAG (Residual Hybrid Attention Group)残差混合注意力组，激活更多像素，提高解码过程中深度图边缘定位的准确性，有效提升了模型的深度估计精度。

② 在解码器之后增加 AdaBins 模块对输出特征图进行全局统计分析，通过全局信息优化输出的深度图，以更好地恢复 RGB 图像中的深度信息。

③ 将原网络中的卷积核替换为 ABC (Asymmetric Convolution Block)非对称卷积块，增强卷积核骨架，提高模型对特征的提取能力，以此提升深度估计精度。

2. DenseNet-169 网络介绍

DenseNet-169 网络主要包括 4 个 DenseBlock 和 3 个 Transition 层[14]，如图 1 中所示。

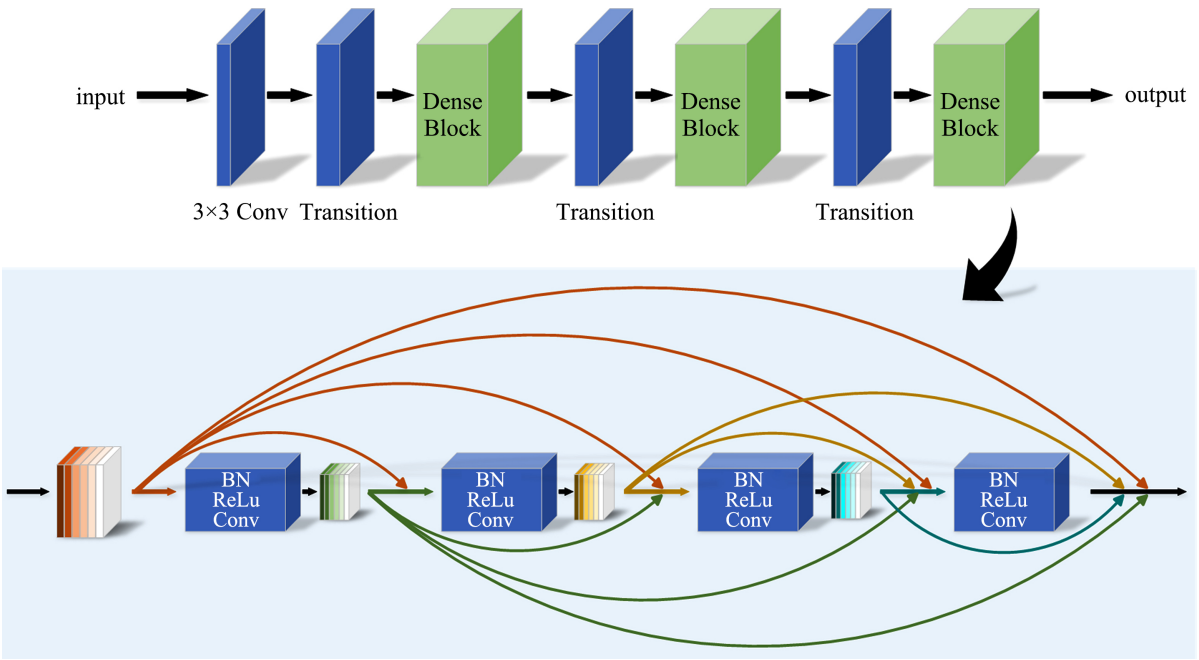


Figure 1. DenseNet-169 network structure
图 1. DenseNet-169 网络结构图

DenseBlock 包含 5 层 layer，其中每一层由 1 × 1 卷积和 3 × 3 卷积组成，输入来自于前面所有层的输出，它的输出也将直接连接到后面层的输入，以此建立低层次和高层次之间的密集连接，实现特征重用，即对不同 layer 的特征进行总体性的再探索，最大化前后层之间的信息交流，一定程度上解决了梯度消失的问题，提升了效率。值得说明的是，在卷积之前会先进行归一化操作，使得每一层在接收前面所有层的输出特征时，能够解决在特征重用时碰到不同层的输出特征的数值差异问题。

Transition 层通过 1 × 1 卷积和平均池化操作衔接两个 DenseBlock，整合上一个 DenseBlock 获得的特征进行下采样，降低特征图的大小，起到了压缩模型、提高计算效率的作用。

密集连接的方式必然会面临增大的参数量，DenseNet 很好的处理了这个问题，不仅将每一层设计的很窄，还在相邻 DenseBlock 之间用 Transition 层连接，这种结构降低了网络的冗余度，减少了运算时的参数量，同时，特征重还可以起到抗过拟合的效果。

3. 算法改进策略

本文整体网络结构如图 2 所示，编码器使用的是 DenseNet-169，通过建立前后层的密集连接，利用低层次和高层次特征在通道上的连接实现特征重用，最后将提取的特征作为编码器的输入。解码器部分采用双线性插值的方法上采样逐步恢复图像的细节特征，并建立编码器与解码器之间的跳跃连接，融合低层像素信息和高层语义信息，使恢复的深度图边缘定位更准确，最后输出特征图。在编码器与解码器的跳连接中引入 RHAG 残差混合注意力组激活更多的输入像素以实现更好的单目深度估计。由于难以恢复在编码过程中丢失的如特征分辨率等信息，所以经过解码器后，将输出的特征图交由 AdaBins 模块进行信息的全局处理，更好的恢复出原图像中的深度信息。同时为了获取到更好的特征表达，使用 ACB 非对称卷积块对 DenseBlock 中的 3 × 3 卷积核骨架进行增强，以提高深度估计的准确率。

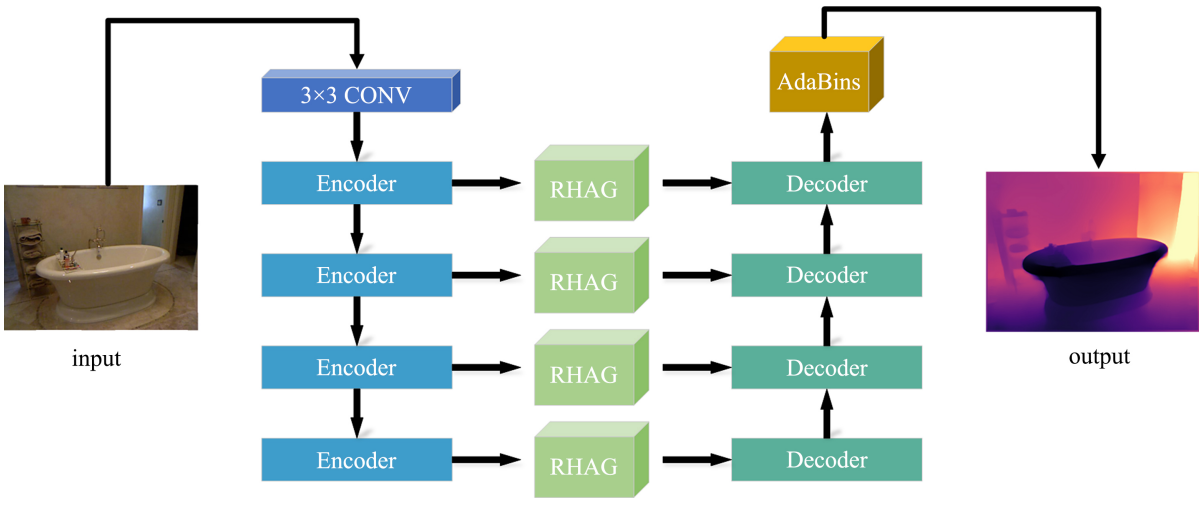


Figure 2. Overall network architecture
图 2. 整体网络架构

最后，本实验中采用平均绝对误差(L1_Loss)作为损失函数，计算预测深度与真实深度之间的误差，并求其平均值：

$$L = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \hat{x}_i\right| \tag{1}$$

式中 n 表示像素点个数，x_i 表示像素的真实值，$\hat{x}_i$ 表示预测值。

3.1. RHAG 残差混合注意力组

表示像素点个数，在计算机视觉领域，注意力机制的出现，使得神经网络能像人类视觉系统一样在面对大量的图像信息时，能聚焦于图像中的有用信息，避免将计算资源浪费在无关紧要的信息上，从而提升模型的整体性能。对于本实验中的编解码网络结构，由于编码阶段连续的卷积和下采样，使得解码后恢复的深度图边缘定位不准确，导致预测的深度信息不准确，所以在编解码的跳连接中引入 RHAG 残差混合注意力组。

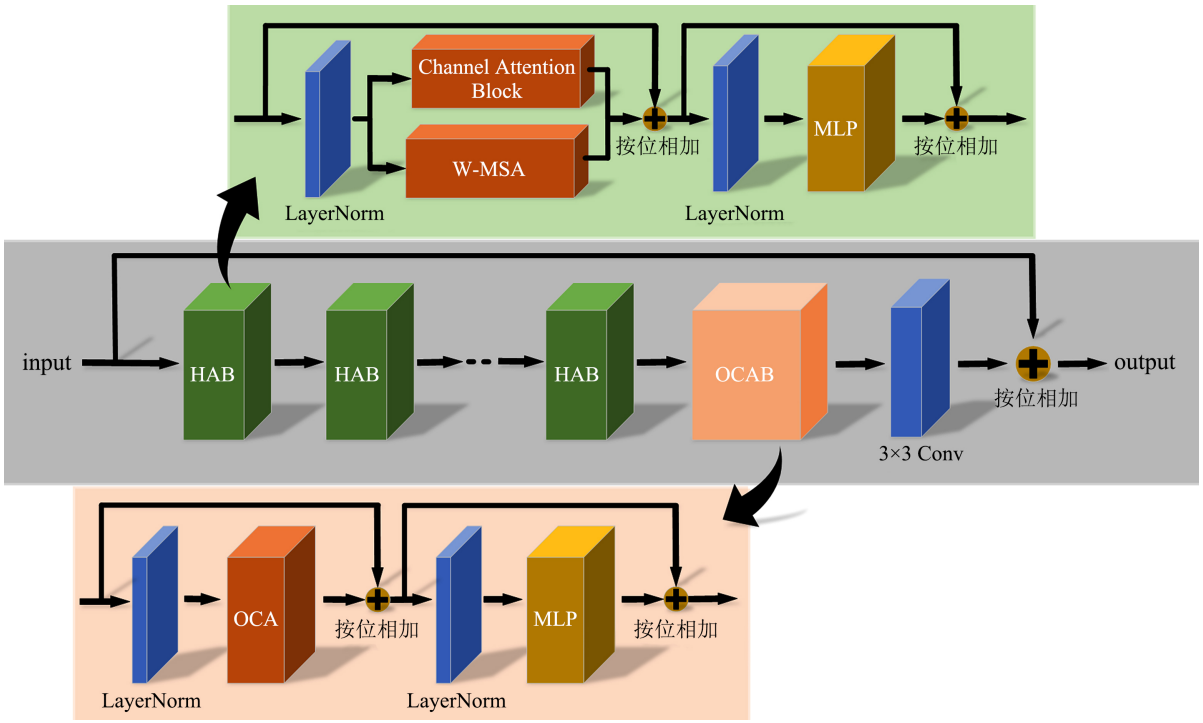


Figure 3. The structure of RHAG
图 3. RHAG 残差混合注意力组

如图 3 所示，RHAG 包含多个 HAB (Hybrid Attention Block)混合注意力块，一个 OCAB (Overlapping Cross-Attention Block)重叠交叉注意力块和一个带有残差连接的 3 × 3 卷积层[15]。HAB 模块由窗口自注意力机制和通道注意力机制组成，通过结合不同类型的注意力机制来激活更多的像素，同时利用局部和全局信息，实现更好的重建效果。在 HAB 中，首先对输入特征进行归一化处理，然后窗口注意力机制将特征图划分成若干个局部窗口，并在每个窗口内计算自注意力，可以通过此操作捕获到局部区域的关联信息。接下来，通过通道注意力机制引入全局信息，利用全局信息对特征进行加权，从而激活更多像素。最后将窗口注意力机制和通道注意力机制的加权和输出，并通过残差连接与输入特征相加。OCAB 模块通过引入重叠交叉注意力层，利用窗口内部的像素信息进行查询，建立窗口之间的交叉连接，以增强网络特征表达的能力，提高网络深度估计任务的性能。

3.2. AdaBins 模块

单目深度估计中，由于场景的复杂性以及深度分布的无规律性，例如：客厅、卫生间等图像中包含的深度信息大部分分布在一个较小的范围内，而走廊这种场景的图像中包含的深度信息可能从网络支持

的最小值到最大值。这使得深度估计的准确性很难达到理想的状态。因此本文引入了 AdaBins 模块，通过对解码器输出的特征图进行全局统计分析，并通过后处理构建块对输出进行优化，进而分析和修整深度的分布值。AdaBins 模块的结构如图 4 所示。

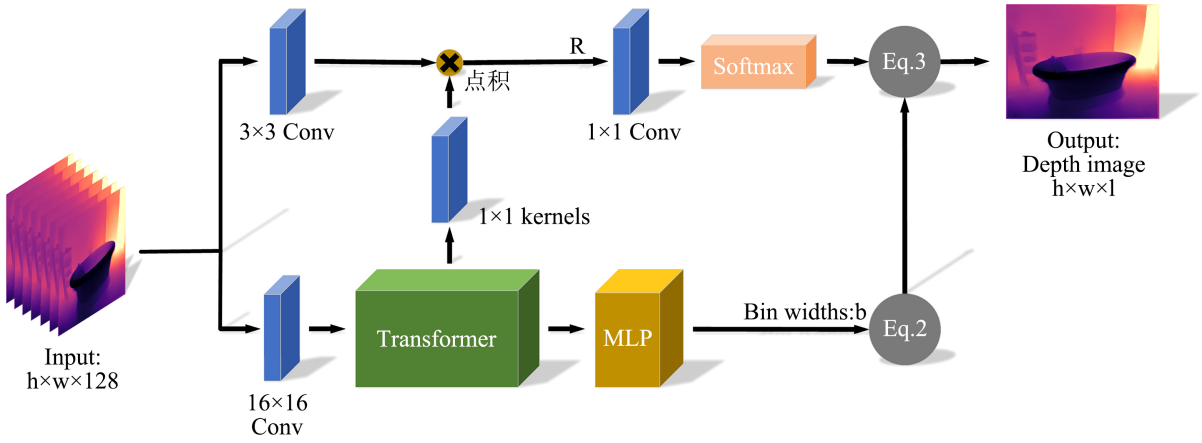


Figure 4. The structure of AdaBins
图 4. AdaBins 结构图

其中 mini-Vit Block 为 Transformer 结构，结合局部结构信息和全局分布信息，估计深度范围更可能出现的层段。解码特征张量输入到 mini-Vit Block，经过 p × p 大小的嵌入式卷积(Embedding Conv)核重构成一个空间平坦的张量，并作为 Transformer 的输入[16]。之后通过 MLP 层归一化后深度信息被划分为若干个深度区间相同的单元 b_i，每个单元的中心深度值 $c(b_i)$ 计算公式为：

$$c(b_i)=y_{\min}+(y_{\max}-y_{\min})\left(\frac{b_i}{2}+\sum_{j=1}^{i-1}b_j\right) \tag{2}$$

式中：$y_{\max}$ 和 $y_{\min}$ 分别指数据集中真实深度的最大值和最小值。

之后将输出的向量通过一个 1 × 1 卷积核和 3 × 3 的卷积核来获取范围注意力图 R (Range attention maps)。之后，范围注意力图 R 经过 1 × 1 卷积核得到 N 通道，并进行 softmax 激活，将得到深度单元中心在每个像素处的概率[17]。最后，将 softmax 得到的概率和深度单元中心 $c(b_i)$ 线性组合得到每个像素的最终深度图：

$$\hat{y}=\sum_{k=1}^{M}c(b_k)E_k \tag{3}$$

式中：E_k 指第 k 通道，其中每个像素的数值为该图像的深度值落在第 k 个区间的概率。

3.3. ACB 非对称卷积块

ACB 非对称卷积块可以替代模型中的普通方形 d × d 卷积核，在推理阶段不需要额外的计算成本的前提下，提升模型的准确率。ACB 由(d × d)、(1 × d)、(d × 1)的三个卷积核组成，三个平行层相加可以对方形卷积核的骨架进行增强。

对于输入特征图 I，先进行 $H^{(1)}$ 和 I 卷积，$H^{(2)}$ 和 I 卷积后，再对结果进行相加得到的结果，与 $H^{(1)}$ 和 $H^{(2)}$ 的逐点相加后再和 I 进行卷积得到的结果是一致的[18]。如式(4)所示，这也解释了为什么 ACB 非对称卷积块能在获取更好的特征表达的同时，又不增加网络推理阶段的任何计算量。

$$I*H^{(1)}+I*H^{(2)}=I*\left(H^{(1)}\oplus H^{(2)}\right) \tag{4}$$

以图 5 中 3 × 3 卷积核为例，可以使用 3 × 3、1 × 3 和 3 × 1 的三个卷积核组成的 ACB 来替换 3 × 3 卷积核，将三个卷积层的计算结果进行融合得到最终输出。在推理阶段，强化后的网络结构与原网络是完全一样的，只是强化后的网络参数采用了特征提取能力更强的参数即融合后的卷积核参数，因此在推理阶段不会增加计算量。

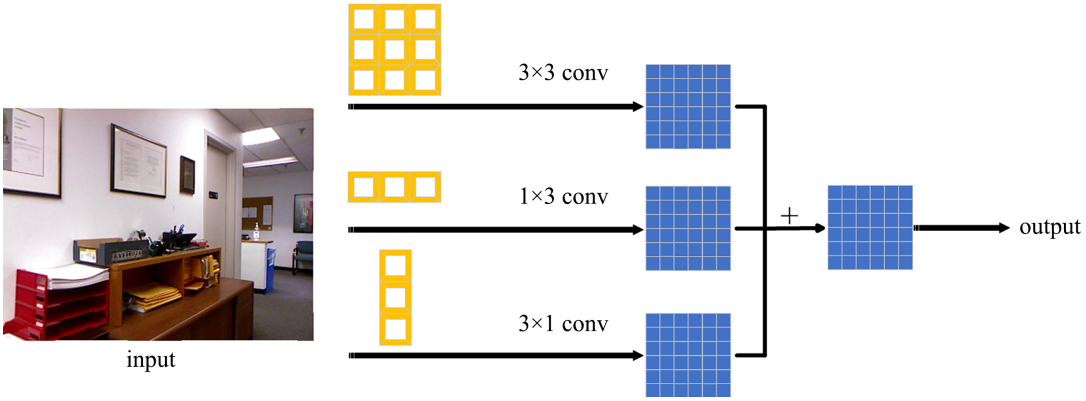


Figure 5. Asymmetric convolution block
图 5. ACB 非对称卷积块

4. 实验结果与分析

4.1. 数据集

实验数据集使用的是目前最广泛最常用的公开数据集 NYU Depth V2 [19]，同样采用该数据集进行评估测试。这个数据集是由微软 Kinect 的 RGB 和 Depth 摄像机记录的视频序列组成，深度范围为 0.5~10 m，共有 407,024 帧 RGB-D 图像对，其中有 1449 张标注的 RGB 图像和深度图，来自 3 个城市，包含 464 个场景，分辨率均为 640 × 480。其中 795 幅 RGB-D 图像对用于训练，654 幅 RGB-D 图像对用于测试。

4.2. 实验环境与评价指标

本文实验基于 Ubuntu 20.04.1 操作系统，显卡使用内存大小为 24 G 的 NVIDIA GeForce GTX 3090，python 版本为 3.7，cuda 版本为 11.2，采用 Pytorch 1.7.1 框架。学习率设置为 0.0001，batch size 设置为 4，epoch 设置为 20。

单目深度估计的评价指标主要是通过计算预测深度与真实深度之间的误差和准确率。目前较为常用的评价指标包含：阈值精度(a_1,a_2,a_3)、均方根误差(RMSE)、对数均方根误差($L_{\log 10}$)、绝对相对误差(AbsRel)。具体公式如为：

阈值精度：

$$\max\left(\frac{\hat{d}_i}{d_i},\frac{d_i}{\hat{d}_i}\right)=a_i<1.25^i \tag{5}$$

均方根误差：

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{d}_i-d_i\right)^2} \tag{6}$$

对数均方根误差：

$$L_{\log 10}=\frac{1}{N}\sum_{i\in N}\left|\log_{10}\hat{d}_i-\log_{10}d_i\right| \tag{7}$$

绝对相对误差：

$$\text{AbsRel}=\frac{1}{|N|}\sum_{i\in N}\frac{\left|\hat{d}_i-d_i\right|}{d_i} \tag{8}$$

其中 N 表示像素总数，d_i 表示第 i 个像素的真实深度值，$\hat{d}_i$ 表示对应的深度预测值。

4.3. 实验结果

4.3.1. 消融实验

为了验证本文提出的算法改进策略对整体性能的影响，以及整个模型的有效性，在 NYU Depth V2 数据集上进行消融实验对比分析。

从表 1 可以看出，与原网络相比，引入 ACB 卷积后，准确率提升了约 0.3%，误差下降了约 0.2%，验证了 ACB 非对称卷积块优于普通卷积。引入 RHAG 模块之后，模型阈值准确分别上升了 1.41%、0.9%、0.1%，均方根误差下降了 2.7%，对数均方根误差减少了 0.3%，绝对相对误差减少了 1.2%，说明了在编码器和解码器之间的跳跃连接引入残差混合注意力组能有效提升模型精度。引入 AdaBins 模块后，模型阈值精度提升了 1.62%、0.82%、0.2%，均方根误差下降了 1.9%，对数均方根误差下降了 0.29%，绝对相对误差降低了 1.67%，说明了在解码出特征图后，经过 AdaBins 后处理优化后能更好的恢复出场景深度信息。同时通过对各个模块的组合，模型整体性能均得到了提升，同时融入三个模块之后模型的各项指标达到最优，进一步验证了本文提出的改进策略能很好的提升模型深度估计的整体性能。

Table 1. Results of ablation experiment
表 1. 消融实验结果

Groups	Methods	a_1	a_2	a_3	RMSE	$L_{\log 10}$	AbsRel
0	原网络	0.8652	0.9712	0.9932	0.4612	0.0539	0.1298
1	+ACB	0.8684	0.9745	0.9933	0.4588	0.0545	0.1285
2	+RHAG	0.8793	0.9803	0.9945	0.4337	0.0508	0.1176
3	+AdaBins	0.8814	0.9794	0.9952	0.4419	0.0510	0.1131
4	+ACB + RHAG	0.8806	0.9810	0.9945	0.4324	0.0506	0.1167
5	+ACB + AdaBins	0.8838	0.9799	0.9948	0.4414	0.0508	0.1129
6	+RHAG + AdaBins	0.8943	0.9841	0.9962	0.4298	0.0493	0.1019
7	+ACB + RHAG + AdaBins	0.8956	0.9845	0.9965	0.4273	0.0489	0.1013

4.3.2. 不同模型对比

为了进一步验证本文提出的算法改进策略的整体性能，使用文献中不同的算法进行了深度估计对比实验，均使用相同的数据集 NYU Depth V2 进行训练，如表 2 所示，定量实验中，文献[20]提供了一种利用迁移学习以实现高分辨率深度估计的算法，与文献[20]中的评价指标进行对比，阈值精度 a_1,a_2,a_3 分别提升了 4.96%、1.1%、0.25%，均方根误差减少了 3.7%，对数均方根误差减少了 0.4%，绝对相对误差减

少了 2.17%，取得了良好的结果。对比 MonoDepth [21]，阈值精度 a_2 提升了 2.2%，绝对相对误差减少了 1.3%。从定量分析的角度证明了本文算法的有效性，通过对整体模型的改进，能通过单张 RGB 图像获取到更精确的深度图。

Table 2. Comparison of different methods on the NYU Depth V2
表 2. NYU Depth V2 数据集上不同方法对比

Methods	a_1	a_2	a_3	RMSE	$L_{\log 10}$	AbsRel
Eigen [6]	0.769	0.950	0.988	0.641	-	0.158
Laina [8]	0.811	0.953	0.988	0.573	0.055	0.127
Xu [22]	0.811	0.954	0.987	0.586	0.052	0.121
Fu [23]	0.828	0.965	0.992	0.509	0.051	0.115
Alhashim [20]	0.846	0.974	0.994	0.465	0.053	0.123
MonoDepth [21]	0.880	0.962	0.982	0.473	0.054	0.114
Ours	0.895	0.984	0.996	0.427	0.048	0.101

(a) 原图 (b) MonoDepth (c) 文献[20] (d) Ours

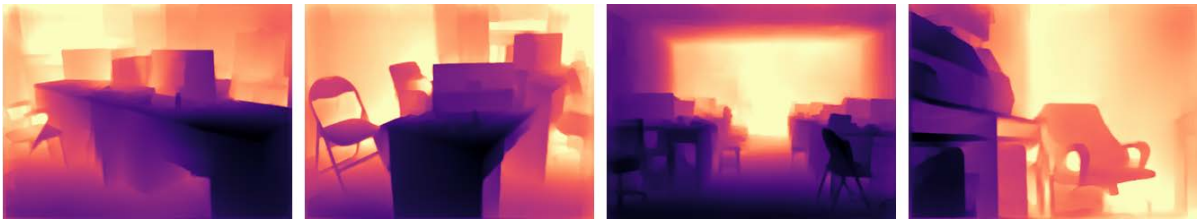
Figure 6. Comparison of depth estimation results
图 6. 深度估计结果对比

在定量对比的基础上，为了更直观的展示本文方法的优势，图 6 展示了本文方法与 MonoDepth 和文献[20]的方法预测的深度图进行对比，定性分析结果。图中第一列为原始 RGB 图，后三列依次为 MonoDepth、文献[20]和本文改进算法分别进行深度估计后得到的深度图，从图中可以看出本文算法能较好的恢复场景中的细节部分，并且在场景中物品的边缘较为明确，如图中白色方框标注的区域，与前二者对比可以明显看出本算法对于深度信息的恢复更为精确。第一行标注出来的浴缸旁边的架子，MonoDepth 及文献[20]中的算法对此部分处理的结果都较为模糊；第二行中沙发边缘及旁边的桌子，本文算法都能较好的恢复出二者的界限；第三行中对于远处的沙发，相较与其他两种方法，本文算法能较好恢复出沙发扶手的细节；第四行中标注出来的椅子及茶几，也能通过本文算法恢复出真实场景中的层次感。

4.4. 真实场景测试

如图 7 中所示，在实际场景中使用本文中训练好的网络进行深度估计，展现出良好的效果，深度图中可以看出层次感更为清晰，对于座椅、办公桌等物品能精确的恢复其边缘信息，预测结果较为准确。由真实场景测试可知，本文中模型具有良好的泛化能力，可以很好的预测盲人出行中碰到的障碍物的距离，具有一定的实用价值。

(a) 真实场景



(b) 预测深度

Figure 7. Real scenario test results
图 7. 真实场景测试结果

5. 系统设计与平台移植验证

为了进一步验证本文中改进的单目深度估计算法在实际工程中的应用，在服务器上训练完成后部署到边缘计算设备上进行实验。实验基于高性能 NVIDIA Xavier NX 边缘计算板设计了主从式的视障人群智能出行监护系统，主设备为 NX，从设备为 Air820 和香橙派 Range Pi。其中，NX 主要负责目标识别、单目深度估计以及各从设备的任务调度，Air820 包含 GPS 定位、摔倒检测、电话短信、基于 LD3320 的语音交互等功能，香橙派负责完成定点导航的功能。图 8 为整个系统的流程框图。

将本文提出的算法和双目深度估计算法移植到视障人群智能出行监护系统中进行对比实验，对于障碍物距离预测，双目深度估计在立体匹配阶段浪费了大量的算力，只能达到每秒 4 帧的检测频率，而本文中提出的算法能达到 21 帧的检测频率，满足了盲人出行时，对障碍物检测所要求的实时性和准确性。

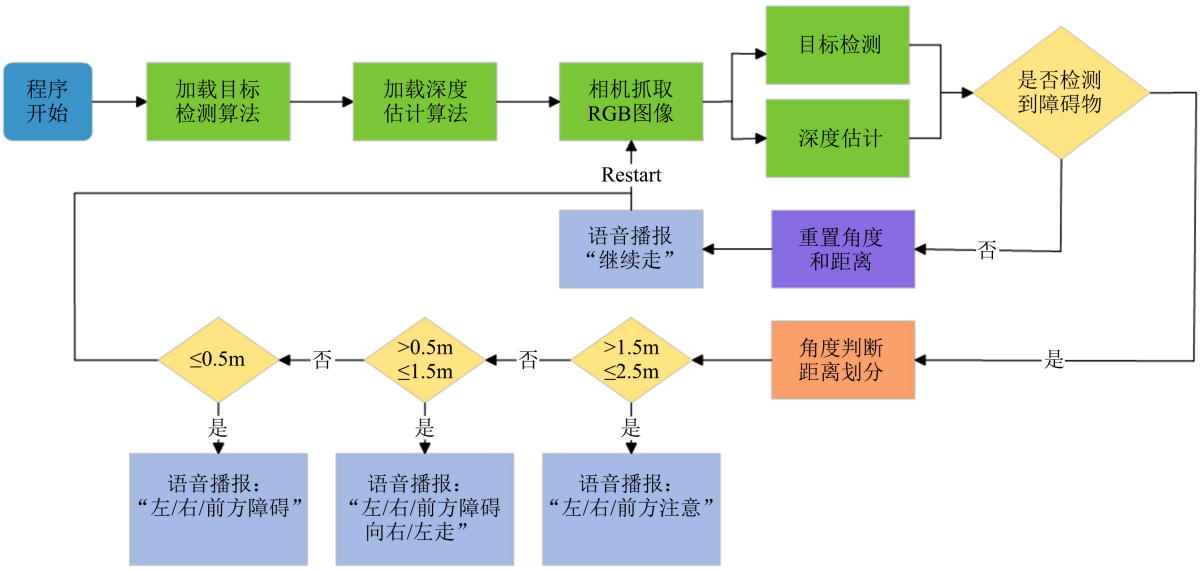


Figure 8. Flow diagram of monitoring system
图 8. 监测系统流程框图

6. 结束语

为了解决盲人出行问题，本文提出了一种基于 DenseNet 的单目深度估计改进模型，用于恢复 RGB 图片中的深度信息，检测障碍物与相机间的距离。通过在编解码之间建立跳跃连接并引入 RHAG 残差混合注意力组，结合不同的注意力机制，利用局部和全局信息，提高深度图中边缘定位的精确性，实现更好的深度恢复效果；引入 AdaBins 模块对解码器输出的特征图进行后处理，通过全局统计对深度图进行修整优化；同时通过替换卷积核以加强卷积核骨架，增强模型特征提取能力，提升深度估计精度。实验结果表明，所提出的改进算法深度预测准确性为 89.5%，绝对相对误差为 0.101，因此该算法能实现较好的单目深度估计效果。同时将算法部署到边缘检测设备 NVIDIA Xavier NX 也满足盲人出行时对障碍物距离实时监测的要求，具有一定的实用价值。后续的工作仍需在模型深度估计精度问题上继续研究，同时深入研究算法在部署到边缘设备时的运算耗时，提升运算效率，使其拥有更好的实时性，实现在盲人日常出行时快速准确的检测障碍物。

基金项目

贵州省基础研究(自然科学)项目黔科合基础-ZK [2021]重点 001 资助。

参考文献

[1] World Health Organization (2019) World Report on Vision.

[2] 高华, 陈秀念, 史伟云. 我国盲的患病率及主要致盲性疾病状况分析[J]. 中华眼科志, 2019, 55(8): 625-628.

[3] Khan, A.I. and Al-Habsi, S. (2020) Machine Learning in Computer Vision. *Procedia Computer Science*, **167**, 1444-1451. https://doi.org/10.1016/j.procs.2020.03.355

[4] Zhou, T., Brown, M.A., Snavely, N. and Lowe, D.G. (2017) Unsupervised Learning of Depth and Ego-Motion from Video. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 6612-6619. https://doi.org/10.1109/CVPR.2017.700

[5] Jimenez Rezende, D., Eslami, S.M., Mohamed, S., Battaglia, P.W., Jaderberg, M. and Heess, N.M. (2016) Unsupervised Learning of 3D Structure from Images. 30*th Conference on Neural Information Processing Systems* (*NIPS* 2016), Barcelona, 5-10 December 2016. https://api.semanticscholar.org/CorpusID:5395254

[6] Eigen, D. and Fergus, R. (2014) Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 2650-2658. https://doi.org/10.1109/ICCV.2015.304

[7] Liu, F., Shen, C., Lin, G. and Reid, I.D. (2015) Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 2024-2039 https://doi.org/10.1109/TPAMI.2015.2505283

[8] Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F. and Navab, N. (2016) Deeper Depth Prediction with Fully Convolutional Residual Networks. 2016 *Fourth International Conference on* 3*D Vision* (3*DV*), Stanford, 25-28 October 2016, 239-248. https://doi.org/10.1109/3DV.2016.32

[9] Huang, G., Liu, Z., Pleiss, G., van der Maaten, L. and Weinberger, K.Q. (2019) Convolutional Networks with Dense Connectivity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 8704-8716. https://doi.org/10.1109/TPAMI.2019.2918284

[10] Jung, H., Kim, Y., Min, D., Oh, C. and Sohn, K. (2017) Depth Prediction from a Single Image with Conditional Adversarial Networks. 2017 *IEEE International Conference on Image Processing* (*ICIP*), Beijing, 17-20 September 2017, 1717-1721. https://doi.org/10.1109/ICIP.2017.8296575

[11] Ranftl, R., Bochkovskiy, A. and Koltun, V. (2021) Vision Transformers for Dense Prediction. 2021 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 12159-12168. https://doi.org/10.1109/ICCV48922.2021.01196

[12] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. arXiv: 1706.03762.

[13] Yuan, W., Gu, X., Dai, Z., Zhu, S. and Tan, P. (2022) Neural Window Fully-Connected CRFs for Monocular Depth Estimation. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 3906-3915. https://doi.org/10.1109/CVPR52688.2022.00389

[14] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C.C., Qiao, Y. and Tang, X. (2018) ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: Leal-Taixé, L. and Roth, S., Eds., *ECCV* 2018: *Computer Vision—ECCV* 2018 *Workshops*, Springer, Cham, 63-79. https://doi.org/10.1007/978-3-030-11021-5_5

[15] Chen, X.Y., Wang, X.T., Zhou, J.T., Qiao, Y. and Dong, C. (2022) Activating More Pixels in Image Super-Resolution Transformer. 2023 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 22367-22377. https://doi.org/10.1109/CVPR52729.2023.02142

[16] Bhat, S., Alhashim, I. and Wonka, P. (2020) AdaBins: Depth Estimation Using Adaptive Bins. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 4008-4017.

[17] Miangoleh, S.M., Dille, S., Mai, L., Paris, S. and Aksoy, Y. (2021) Boosting Monocular Depth Estimation Models to High-Resolution via Content-Adaptive Multi-Resolution Merging. 2021 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 9680-9689. https://doi.org/10.1109/CVPR46437.2021.00956

[18] Ding, X., Guo, Y., Ding, G. and Han, J. (2019) ACNet: Strengthening the Kernel Skeletons for Powerful CNN via Asymmetric Convolution Blocks. 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 1911-1920. https://doi.org/10.1109/ICCV.2019.00200

[19] Chang, A., Dai, A., Funkhouser, T., *et al*. (2017) Matterport3D: Learning from RGB-D Data in Indoor Environments. 2017 *International Conference on* 3*D Vision* (3*DV*), Qingdao, 10-12 October 2017, 667-676. https://doi.org/10.1109/3DV.2017.00081

[20] Alhashim, I. and Wonka, P. (2018) High Quality Monocular Depth Estimation via Transfer Learning. arXiv: 1812.11941.

[21] Godard, C., Mac Aodha, O. and Brostow, G.J. (2018) Digging Into Self-Supervised Monocular Depth Estimation. 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 3827-3837. https://doi.org/10.1109/ICCV.2019.00393

[22] Xu, D., Ricci, E., Ouyang, W., Wang, X. and Sebe, N. (2017) Multi-Scale Continuous CRFs as Sequential Deep Networks for Monocular Depth Estimation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 161-169. https://doi.org/10.1109/CVPR.2017.25

[23] Fu, H., Gong, M., Wang, C., Batmanghelich, K. and Tao, D. (2018) Deep Ordinal Regression Network for Monocular Depth Estimation. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2002-2011. https://doi.org/10.1109/CVPR.2018.00214