基于数据增强的半监督单目深度估计框架

赵宏伟，周伟民

（吉林大学 软件学院，长春 130012）

摘 要：为解决监督学习在单目深度估计中需要大量标签数据的问题，提出了一种基于教师-学生模型的半监督深度估计框架AugDepth。其通过对数据进行扰动，训练模型学习扰动前、后的深度一致性。首先，采用平滑随机强度增强方法从连续域中采样强度，随机选择多个操作以增加数据随机性，并混合强弱增强输出，防止过度扰动。然后，考虑到不同无标签样本的训练难度不同，在通过Cutout提高模型对全局信息推理的前提下，根据对无标签样本的置信度，自适应地调整Cutout策略，以提高模型的泛化和学习能力。在KITTI和NYU-Depth数据集上的实验结果表明：AugDepth能够显著提高半监督深度估计的准确性，并在有标签数据稀缺的情况下表现出良好的鲁棒性。

关键词：计算机应用；半监督学习；数据增强；单目图像；深度估计
中图分类号：TP391 **文献标志码**：A **文章编号**：1671-5497(2025)06-2082-07
DOI：10.13229/j.cnki.jdxbgxb.20230964

Semi-supervised monocular depth estimation framework based on data augmentation

ZHAO Hong-wei，ZHOU Wei-min
（*College of Software*，*Jilin University*，*Changchun* 130012，*China*）

Abstract：To address the problem of requiring a large amount of labeled data for supervised learning in monocular depth estimation，a semi-supervised depth estimation framework AugDepth was proposed based on a teacher-student model. It operates by perturbing the data and training the model to learn depth consistency before and after the perturbation. Firstly，the smooth random intensity enhancement method was used to sample the intensity from the continuous domain. Multiple operations were randomly selected to increase the randomness of the data，and the output was enhanced by mixing the strength and weakness to prevent excessive disturbance. Then，considering the varying training difficulties of different unlabeled samples，while improving the model′s inference of global information through Cutout，the Cutout strategy is adaptively adjusted based on the confidence level of unlabeled samples to enhance the model′s generalization and learning abilities. The experimental results on the KITTI and NYU Deeph datasets show that AugDepth can significantly improve the accuracy of semi supervised depth estimation and exhibit good robustness in situations where labeled data is scarce.

收稿日期：2023-09-10.
基金项目：吉林省省级科技创新专项项目(20190302026GX)；吉林省自然科学基金项目(20200201037JC).
作者简介：赵宏伟(1962-)，男，教授，博士. 研究方向：嵌入式人工智能. E-mail：zhaohw@jlu.edu.cn

Key words: computer application; semi-supervised learning; data agumentation; monocular image; depth estimation

0　引　言

单目深度估计是一种利用深度学习技术从单张图像中恢复出每个像素深度值的方法，它在许多领域有着广泛的应用，例如三维重建、虚拟现实和自动驾驶等。近年来，基于有监督单目深度估计的研究已经获得了显著发展[1-3]。

由于真实深度数据的获取成本高，使基于真实深度图训练监督深度估计模型面临严峻的挑战。而半监督单目深度估计方法可以有效避免以上问题，它们利用有限的有标签深度数据和大量的无标签图像数据提高深度估计的准确性。目前已有一些半监督方法被提出，例如 Ji 等[4]利用对抗学习框架，从少量图像深度对和大量无标签图像中评估深度，通过生成器和判别器的竞争提高模型的准确率。另外，Cho 等[5]和 Guo 等[6]提出了两种基于立体图像对进行预训练的匹配网络半监督方法，它们利用教师网络生成深度伪标签，并通过知识蒸馏框架指导学生网络从无标签数据中学习深度信息。尽管这些框架都利用无标签数据降低了对数据的依赖，提升了深度估计的性能，但其大都需要引入额外的网络或者训练过程，从而增加了模型的复杂性和计算成本。

为了解决当前半监督深度估计框架的复杂性和有监督框架对有标签数据量依赖的问题。本文设计了一种简单而有效的半监督框架 AugDepth，它依靠数据扰动来增强半监督单目深度估计的性能。主要工作包括：①打破了采用越来越复杂的设计提高模型性能的趋势，设计了一种简单有效的双分支教师学生半监督框架 AugDepth，降低了监督学习对有标签数据量的依赖。②对常用的数据增强方法[7-10]进行改进，首先是改进了传统的随机增强，从连续的空间中采样扰动程度，使其更加适合半监督深度估计任务，同时为了防止过度扰动，混合了强弱增强输出。其次对无标签样本执行 Cutout 以提升模型性能，迫使模型综合利用场景全局信息进行推理，而不仅依赖局部重复特征，同时不同无标签样本的训练难易程度存在差异，为平衡统一增模型对不同样本的训练过程，防止容易样本的过拟合以及难样本的损失，模型根据对无标签样本的置信度，自适应地调整 Cutout 策略，以提高模型的泛化和学习能力。③在 KITTI 和 NYU-Depth 数据集上的实验结果表明：本文框架优于现有的半监督框架，且在标签数据稀少的情况下，AugDepth 算法明显优于监督学习算法。

1　本文方法

1.1　AugDepth 方法概述

本文构建了一个半监督学习框架，采用半监督学习中广泛使用的一致性正则化方法[11]训练深度估计网络。如图 1 所示，学生模型在有标签的数据上进行监督学习以优化网络参数；在无标签数据上，该框架通过数据增强，并计算教师模型与学生模型的输出一致性损失来提高模型泛化能力。其中教师模型的参数根据学生模型权重的指数滑动平均进行更新。具体更新公式如下：

$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s \tag{1}$$

式中：θ_s 和 θ_t 分别为学生模型和教师模型的参数；动量参数 α 设为 0.999。

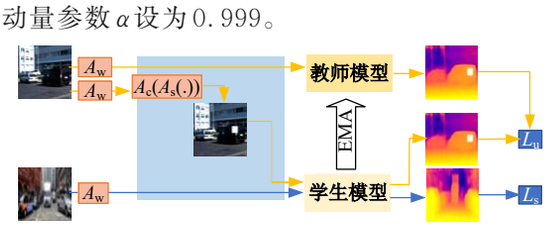

图 1　AugDepth 框架图

Fig. 1　Overall architecture of AugDepth framewo

本文通过同时最小化有监督损失 L_s 和无监督一致性损失 L_u 来训练学生模型。具体而言，学生模型的训练损失函数 L 定义为：

$$L = L_s + L_u \tag{2}$$

与其他深度估计方法一致，本文采用了 Eigen 等[1]提出的平方根损失函数作为有监督损失 L_s，它保证了预测场景中深度值的尺度一致，如下所示：

$$L_s = \alpha\sqrt{\frac{1}{N}\sum_i\left(\log\left(\frac{y_i}{g_i}\right)\right)^2 - \frac{\lambda}{N^2}\left(\sum_i\log\left(\frac{y_i}{g_i}\right)\right)^2} \tag{3}$$

式中：y_i 为预测值；g_i 为真实深度值；N 为有效像素的总数；平衡因子 λ 设为 0.85，α 设为 10。

AugDepth 通过级联的两阶段数据增强操作 $T = A_c(A_s(\cdot))$ 实现无标签数据的充分利用，它对同一无标签输入 u_i 生成不同的输出，从而导致教师网络和学生网络之间的预测不一致性。学生模型和教师模型的预测结果分别如下：

$$y_s = f(A_w(u_i;\ \theta_t)) \tag{4}$$

$$y_t = f(T(A_w(u_i)); \theta_s) \tag{5}$$

式中:A_w为弱几何增强,本文设置为随机水平翻转。

因此,无监督损失L_u如下:

$$L_u = \alpha \sqrt{\frac{1}{N}\sum_i \left(\log\left(\frac{y_s}{y_t}\right)\right)^2 - \frac{\lambda}{N^2}\left(\sum_i \log\left(\frac{y_s}{y_t}\right)\right)^2} \tag{6}$$

AugDepth没有引入额外模块,它通过一致性正则化和数据增强实现简单高效的半监督深度学习。一致性正则化强制网络对扰动输入保持稳定预测,以有效利用无标签数据,而无需复杂的对抗或蒸馏训练过程。这种简单高效的框架结构降低了计算和实现的复杂度。

1.2 平滑强度随机增强

传统的随机增强RandAugment是从预定义的增强池中随机选择固定数目的操作,并在离散强度集合中选取一定程度的增强应用于输入图像。这种增强策略是为了适应下游任务的要求,而不是针对半监督学习设计的。半监督学习中数据扰动的目标是从同一输入生成两个不同的图像,以提高模型对输入变化的稳定预测能力。同时,过度的扰动增强会破坏数据分布并且让半监督学习的表现变差[12]。为解决这些问题,设计了一种平滑随机强度增强A_s来扰动无标签数据。具体而言,为实现有效的强增强,本文从连续域中随机采样增强程度,并从表1所示增强池中随机选择不超过k种操作。

表1 增强池

Table 1 Enhancement tank

增强操作	增强策略细节
映射	返回原始图像
均衡化	均衡化图像的直方图
高斯模糊	用高斯核模糊图像
对比度	调整图像的对比度到[0.05,0.95]
锐度	调整图像的锐度到[0.05,0.95]
颜色	将图像的颜色平衡增强到[0.05,0.95]
亮度	调整图像的亮度到[0.05,0.95]
海报化	将每个像素减少到[4,8]位
太阳化	反转图像中高于[1,256]阈值的像素

以上操作增加了数据随机性,更适合半监督学习场景。另外,为避免强增强导致过度的扰动,将强增强输出和弱增强输出进行混合,以平衡增强效果。形式上,无标签实例u_i的平滑随机强度增强输出可以表示为:

$$A_s(u_i) = \gamma_i A_b(A_w(u_i)) + (1-\gamma_i)A_w(u_i) \tag{7}$$

式中:A_b为平滑强度随机增强的强增强部分;A_w表示弱几何增强,本文设置为随机水平翻转;γ_i为强度平衡因子超参数。

混合强弱增强输出可以平衡增强效果并维持数据原始分布,避免过度扰动输入,降低了模型面临数据分布偏移的风险。相比传统数据增强(例如几何变换和颜色变换)的固定模式,具有更高的灵活性和适应性,且更能适应半监督学习的特点。

1.3 自适应Cutout数据增强

在单目深度估计任务中,相同的局部图像特征可能对应不同的深度范围,仅根据局部信息进行预测通常不可靠。因此,本文采用了Cutout数据增强方法,其通过在图像中随机遮挡一定面积的连续区域,迫使模型不能过分依赖局部重复特征,而需综合利用全局内容(如物体间遮挡关系、场景线索与阴影等上下文信息)进行推,以获得更准确的深度预测。然而,不加区别地对所有无标签样本应用Cutout数据增强,会对难以训练的样本造成过度扰动,降低模型的训练效果。为解决此问题,本文设计了一种自适应Cutout数据增强A_c,它可以动态调节每个批次中无标签样本执行Cutout的概率。具体而言,对模型预测置信度高的样本,以较大概率施加Cutout,增强模型鲁棒性;而对置信度较低的样本,以较小概率施加Cutout,防止容易样本的过拟合以及难样本的损失,维持模型的学习能力。

本文采用了Poggi等[13]提出的不确定性估计策略。对于输入图像I和其水平翻转的副本I',用教师模型分别对两张图生成深度图d和d',再将d'进行水平翻转得到d'',不确定性图u_c被定义为两个深度图之间的差异,公式如下:

$$u_c = |d - d''| \tag{8}$$

将不确定性图转换成置信分数s_i:

$$s_i = 1 - \frac{p_i - p_{\min}}{p_{\max} - p_{\min}} \tag{9}$$

式中:p_i为当前批次中第i个不确定性图的平均值;$p_{\min}$、$p_{\max}$分别为当前批次中的不确定值的最小值、最大值。

如图2所示,对每个样本,模型会输出一个预测置信度分数s_i,用于衡量当前模型对该样本的预测置信水平。同时对每个样本随机生成一个0到1之间的数c_i。如果$c_i > s_i$,则对该样本执行Cutout数据增强;否则不做处理。通过这种自适应调控数据增强概率的机制,可以针对不同难易程度的样本进行个性化处理,有助于模型综合利

用全局信息和局部细节，进而提高深度预测的准确性和模型泛化能力。与传统的Cutout相比，这种机制提高了数据增强的针对性，是一种更有效的策略。

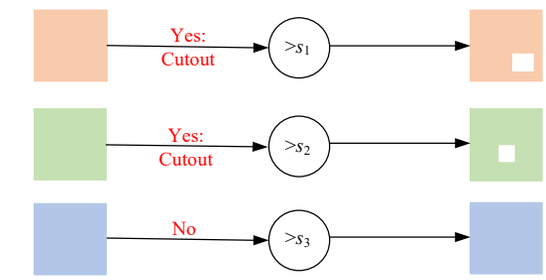

图 2　自适应的 Cutout 增强

Fig. 2　Adaptive cutout augmentation

2　实验结果及分析

2.1　数据集描述及实验设置

为了证明本文模型的有效性，选择在数据集KITTI和NYU-Depth上进行实验验证。KITTI数据集是由相机拍摄的RGB图像和激光雷达扫描获得的深度图组成的室外场景，其RGB图像在训练时分辨率调整为640×192，该数据集表示的距离是0～80 m。NYU-Depth数据集则是由相机拍摄的RGB图像和深度相机采集的深度图组成的室内场景，其训练时的分辨率为576×448，该数据集表示的距离是0～10 m。同时，针对模拟不同程度的标签数据缺失情况，本文遵循Baek等[14]提出的方案。从原始数据集中随机抽取23 158、10 000、1 000和100张图像作为有标签训练数据，剩余图像作为无标签数据在数据集上进行实验。在KITTI数据集的652张Eigen测试集和NYU-Depth数据集的654张官方测试集上评估模型性能，并与其他方法进行比较。

本文教师模型和学生模型都采用了LapDepth的网络结构，其编码器为在ILSVRC数据集上预训练的ResNet 50，解码器为LapDepth中提出的LapDecoder，其权重采用随机初始化方案。训练轮次设为40，批处理大小为12，使用Adam优化器更新模型参数，其中Adam优化器的参数设置β_1为0.9，β_2为0.999，初始学习率设置为0.000 1，使用多项式学习率调度器，最终学习率为0.000 01。

最后，本文遵循Eigen等[1]工作的标准评价协议，来评估AugDepth的有效性。采用以下几种误差指标来衡量深度预测的准确性和误差：绝对相对误差(AbsRel)，平方相对误差(SqRel)，均方根误差(RMSE)，均方根对数误差(RMSElog)，以及在阈值(<1.25)下的准确率(δ_1)。

2.2　深度估计结果

本节探究了标签数量对有监督单目深度估计的影响以及AugDepth框架在标签稀缺时维持深度预测质量的有效性。此外，还与多种监督方式的深度估计方法进行了比较，其中各个模型中的训练参数与其原始文献保持一致。

2.2.1　AugDepth的鲁棒性

本文在KITTI数据集上首先将AugDepth与有监督基准模型LapDepth进行比较实验。图3

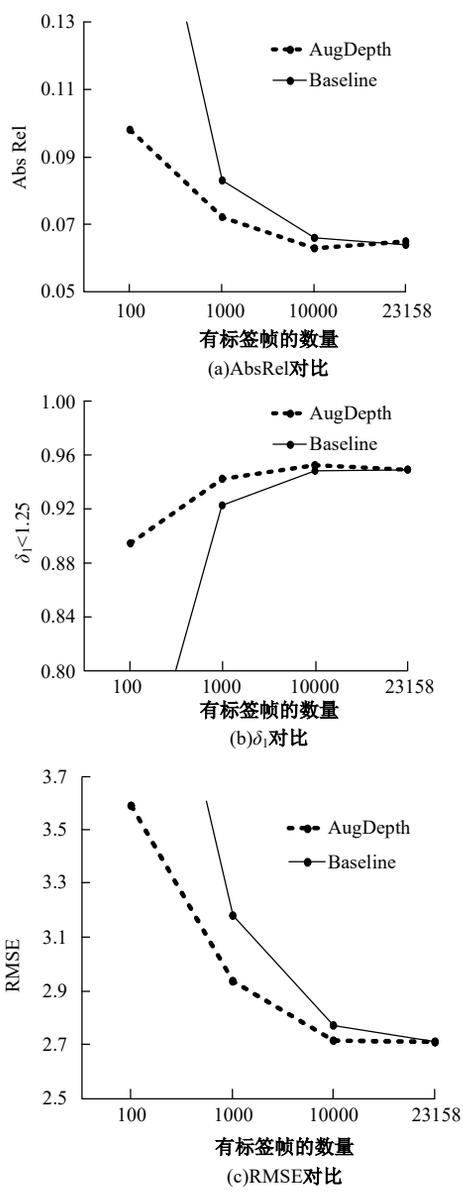

图 3　不同标签数量的定量结果

Fig. 3　Quantitative results for different number of labels

展示了使用不同数量监督训练帧下的AbsRel、RMSE和δ_1结果。

从图3可以看出,本文提出的半监督模型AugDepth在任意数量有标签数据上均优于LapDepth。随着标签数据量的进一步减少,LapDepth的性能出现了显著下降。而本文提出的AugDepth却能够有效利用无标签数据进行深度估计,从而提升模型的性能。相比基准模型,AugDepth具有更强的鲁棒性与更好的泛化能力。

2.2.2 AugDepth的实验结果

为了验证AugDepth的有效性,本文在两个公开数据集KITTI和NYU-Depth上进行实验评估,并与当前的主流网络进行比较分析。实验结果如表2和表3所示。从表2可以看出,在KITTI数据集上,AugDepth在多个指标上均取得了相当好的性能,包括SqRel、RMSE和δ_1,这些指标反映了模型对深度预测的精度和误差。从表3可以看出,在NYU-Depth数据集上,AugDepth在AbsRel和δ_1指标上也达到了最优的水平,这两个指标反映了模型深度预测的相对误差和一致性。结果表明:AugDepth在两个数据集上都优于其他方法,在大部分指标上表现出明显的优势。实验证明了本文AugDepth方法对半监督深度估计任务的有效性。

表2 KITTI数据集上的定量结果

Table 2 Quantitative results on the Eigen split of the KITTI dataset

方法	监督方式	AbsRel	SqRel	RMSE	RMSElog	δ_1
DORN[15]	有监督	0.072	0.307	2.727	0.120	0.932
LapDepth(Resnet50)	有监督	0.064	0.259	2.828	0.102	0.949
Monodepth2[16]	自监督	0.080	0.466	3.681	0.127	0.926
FeatDepth[17]	自监督	0.079	0.666	3.922	0.163	0.925
Cho[5]	半监督	0.095	0.613	4.129	0.175	0.884
SemiDepth[18]	半监督	0.078	0.417	3.464	0.126	0.923
Baek[14]	半监督	0.071	0.316	3.049	0.111	0.941
本文	半监督	0.062	0.251	2.726	0.098	0.954

表3 NYU-Depth数据集上的定量结果

Table 3 Quantitative results on the NYU-Depth

方法	AbsRel	RMSE	δ_1
Eigen[1]	0.158	0.641	0.769
DORN	0.115	0.509	0.828
BTS	0.112	0.352	0.882
LapDepth	0.110	0.393	0.885
DPT-Hybrid[19]	0.110	0.357	0.904
Baek[14]	0.109	0.392	0.894
Ours	0.105	0.395	0.889

2.3 消融实验

为了分析本文方法各个模块的有效性和其他传统数据增强模块的优越性,在KITTI数据集上进行了实验,从有标签数据中随机抽取10 000张作为有标签训练集,剩余的作为无标签集。表4为不同模块对AbsRel和δ_1指标的影响。表5为本文提出的两种数据增强方式与传统增强方式的比较,包括RandAugment和Cutout。表4中的MT表示标准的教师学生模型框架。表4结果表明:本文提出的两个数据增强模块都可以显著提升模型性能,相比于基准模型LapDepth,本文方法在AbsRel和δ_1指标上分别取得了6.06%和0.63%的相对改进。其中,平滑强度随机增强方法可以将AbsRel指标降低1.51%,将δ_1指标提高0.21%;自适应Cutout增强方法可以将AbsRel指标降低3.03%,将δ_1指标提高0.42%;联合使用两个模块可以进一步提升模型性能,将AbsRel指标降低6.06%,将δ_1指标提高0.63%。这些结果证明本文模块设计对半监督深度估计效果有显著贡献。

表4 AugDepth的消融实验

Table 4 Ablation studies on our AugDepth

MT	A_s	A_c	AbsRel	δ_1
			0.066(supervised)	0.948(supervised)
√			0.066	0.949
√	√		0.065	0.950
√		√	0.064	0.952
√	√	√	0.062	0.954

如表5所示,在标准教师学生模型上将本文方法与传统数据增强方法进行比较,本文提出的A_c相对于传统的Cutout,在AbsRel指标上降低了1.54%,A_s相对于RandAugment在AbsRel指标

上同样降低了,A_c+A_s与Cutout和RandAugment的联合模块对比,AbsRel指标降低了3.12%。尽管这些传统的数据增强方法改善了模型性能,但是本文增强方法不仅实现了最佳性能,而且其随机性和自适应性使其更加适合半监督的场景。

表5　数据增强效果比较

Table 5　Comparison of data augmentation effects

方法	AbsRel	δ_1
RandAugment	0.066	0.951
Cutout	0.065	0.951
RandomAugment + Cutout	0.064	0.953
A_s	0.065	0.950
A_c	0.064	0.952
A_c+A_s	0.062	0.954

3　结束语

本文基于一致性框架和数据增强提出了AugDepth,有效利用了无标签数据来提高模型性能。与近期的半监督深度估计研究倾向于结合越来越复杂的机制不同,AugDepth不需要任何额外复杂的设计,仅通过优化数据增强方式,既保证了充分的数据扰动,又保证了不会破坏数据分布。实验结果表明:AugDepth能够显著提高半监督深度估计的准确性。同时,在有标签数据稀缺的情况下,AugDepth表现出了良好的鲁棒性。

参考文献:

[1] Eigen D, Puhrsch C, Fergus R. Depth map predictionfrom a single image using a multi-scale deep network[C]//Advances in Neural Information Processing Systems,Montreal, Canada, 2014: 2366-2374.

[2] Song M, Lim S, Kim W. Monocular depth estimation using laplacian pyramid-based depth residuals[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(11): 4381-4393.

[3] Lee J H, Han M K, Ko D W, et al. From big to small: multi-scale local planar guidance for monocular depth estimation[J/OL]. [2023-08-26]. https://arxiv.org/pdf/1907.10326

[4] Ji R, Li K, Wang Y, et al. Semi-supervised adversarial monocular depth estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2410-2422.

[5] Cho J, Min D, Kim Y, et al. A large RGB-D dataset for semi-supervised monocular depth estimation [J/OL]. [2023-08-27]. https://arxiv.org/pdf/1904.10230

[6] Guo X, Li H, Yi S, et al. Learning monocular depth by distilling cross-domain stereo networks[C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 506-523.

[7] Cubuk E D, Zoph B, Shlens J, et al. Randaugment: practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,Seattle, USA, 2020: 702-703.

[8] Zhao Z, Yang L, Long S, et al. Augmentation matters: a simple-yet-effective approach to semi-supervisedsemantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 11350-11359.

[9] Zhao Z, Long S, Pi J, et al. Instance-specific and model-adaptive supervision for semi-supervised semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023: 23705-23714.

[10] de Vries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J/OL]. [2023-08-28]. https://arxiv.org/pdf/1708.04552

[11] Tarvainen A, Valpola H. Mean teachers are better rolemodels: weight-averaged consistency targets improve semi-supervised deep learning results[C]//Advances in Neural Information Processing System,Vancouver, Canada, 2017: 1195-1204.

[12] Yuan J, Liu Y, Shen C, et al. A simple baseline for semi-supervised semantic segmentation with strong data augmentation[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021: 8209-8218.

[13] Poggi M, Aleotti F, Tosi F, et al. On the uncertainty of self-supervised monocular depth estimation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 3227-3237.

[14] Baek J, Kim G, Park S, et al. MaskingDepth: masked consistency regularization for semi-supervised monocular depth estimation[J/OL]. [2023-08-29]. https://ieeexplore.ieee.org/abstract/document/10801719

[15] Fu H, Gong M, Wang C, et al. Deep ordinal regression network for monocular depth estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos, USA, 2018: 2002-2011.

[16] Godard C, Aodha O M, Firman M, et al. Digging into self-supervised monocular depth estimation[C]// IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 2019: 3827-3837.

[17] Shu C, Yu K, Duan Z, et al. Feature-metric loss for self-supervised learning of depth and egomotion[C]// European Conference on Computer Vision, Glasgow, UK, 2020: 572-588.

[18] Amiri A J, Loo S Y, Zhang H. Semi-supervised monocular depth estimation with left-right consistency using deep neural network[C]//IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 2019: 602-607.

[19] Ranftl R, Bochkovskiy A, Koltun V. Vision transformers for dense prediction[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021: 12159-12168.