• 计算机工程与应用 •

基于结构化损失的单目深度估计算法研究

霍智勇，乔 璐*

(南京邮电大学通信与信息工程学院 南京 210023)

【摘要】为了提高单目图像深度估计的精度，针对图像中几何形状无法准确预测以及边缘模糊的问题，该文提出了一种基于多尺度结构相似度和梯度匹配的单目深度估计算法，利用多尺度结构相似度损失和尺度不变梯度匹配损失组成联合结构化损失，对相对深度点对进行排序来实现单目深度估计，实现了对图像中几何形状的准确预测，减小了边缘模糊，提高了深度预测精度。在 Ibims、NYUDv2、DIODE、Sintel 4 个不同类型的数据集进行了数值实验和主观评测，结果表明该算法降低了深度预测误差，有效提高了预测的准确性，并具有一定的泛化性能。

关 键 词 卷积网络； 深度估计； 梯度匹配损失； 单目图像； 多尺度结构相似度损失； 排序损失

中图分类号 TP391；TP183 **文献标志码** A doi:10.12178/1001-0548.2020386

Research on Monocular Depth Estimation Algorithm Based on Structured Loss

HUO Zhiyong and QIAO Lu*

(College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023)

Abstract This paper proposes a monocular depth estimation algorithm based on multi-scale structure similarity and gradient matching for improving the accuracy of monocular image depth estimation and solving the problems of inaccurate prediction of geometric shapes and blurred edges in the image. In this algorithm, a joint structured loss is formed by using multi-scale structure similarity degree loss and scale-invariant gradient matching loss. The relative depth points are sorted to achieve monocular depth estimation, which realizes accurate prediction of geometric shapes in the image, reduces edge blur, and improves depth prediction accuracy. Numerical experiments and subjective evaluations are performed on four different types of data sets: Ibims, NYUDv2, DIODE, and Sintel. The results show that the algorithm significantly reduces the depth prediction error, effectively improves the accuracy of the prediction, and has a certain generalization performance.

Key words convolutional network; depth estimation; gradient matching loss; monocular image; multi-scale structural similarity loss; ranking loss

从单目图像中获取深度信息是理解场景几何关系的重要方法，也是三维重建[1]和视点合成[2-3]的关键性技术。传统的基于光流或运动恢复结构(structure from motion, SfM)[4]的算法可以获取单目运动图像序列或单目视频的深度信息，却无法预测单帧静止图像的深度。近年来，利用深度学习的方法预测单目静止图像的深度图成为研究热点。文献 [5] 首次提出采用卷积神经网络进行单目深度估计，运用神经网络获取全局粗略深度图以及改善局部细节。文献 [6] 提出了一种包含残差网络模块的全卷积网络对单目图像和深度图之间的模糊映射进行建模的方法，为了提高输出分辨率，再提出了特征上采样的学习方法以及引入反向 Huber 损失进行优化。文献 [7] 对未作标记的单目图像序列，采用无监督的方式实现对单目深度估计网络和相机姿态估计网络的训练。文献 [8] 将卷积神经网络与连续条件随机场相结合，估计单目图像深度。文献 [9] 提出了采用相对深度进行深度预测的方法，即对输入图像中由人工标注的相对深度注释点对之间的相对关系进行排序估计。之后，文献 [10] 又通过采用质量评价网络识别出基于 SfM 方法获得的高质量重构图像，作为监督视图以获取估计深度。文献 [11] 对由双目图像获得的 GT(ground-truth) 深度图和由深度卷积网络生成的预测深度图进行随机

收稿日期：2020－10－14；修回日期：2021－06－11

作者简介：霍智勇 (1976－)，男，博士，教授，主要从事模式识别与计算机视觉等方面的研究.

*通信作者：乔璐，E-mail：13851603316@163.com

采样，从而训练出相对深度预测网络模型。上述提到的相对深度方法均采用排序损失，仅针对输入图像中的全局相对深度信息进行训练，忽略了图像中的几何信息以及局部边缘信息，在几何形状以及深度不连续处不能获得准确的预测结果。因此，本文提出了一种基于多尺度结构相似度和梯度匹配的联合损失函数，对输入的单目图像获得更准确的深度预测，深度不连续处也更加清晰。

1　算法概述

1.1　网络架构

本文在训练中采用了基于文献 [12] 的多尺度编码器–解码器神经网络架构，其网络架构如图 1 所示。编码器部分是在 ResNet50 网络基础上，删除了 ResNet50 网络的最后一个池化层、全连接层以及 softmax 层，使编码器更好地应用于密集的每像素预测任务；解码器部分采用多尺度融合模块，每个融合模块由两个残差卷积块和一个双线性上采样层组成；在解码器的最后添加一个自适应输出模块，该模块由两个卷积层和一个双线性上采样层组成。

输入图像通过编码器网络生成一系列具有不同语义的特征图，根据特征图的分辨率将编码器分为 4 个不同的构建模块。由于 ResNet 包含步长为 2 的卷积序列和池化操作，因此增大了卷积的接受域以捕获更多的上下文信息，但同时降低了输出特征图的分辨率。在解码器部分，考虑到若直接使用简单的上采样和反卷积会生成粗略的预测图像，若使用空洞卷积生成的深度图会带有棋盘伪影，所以为了获取准确的预测深度图，本文采用多尺度特征融合模块。

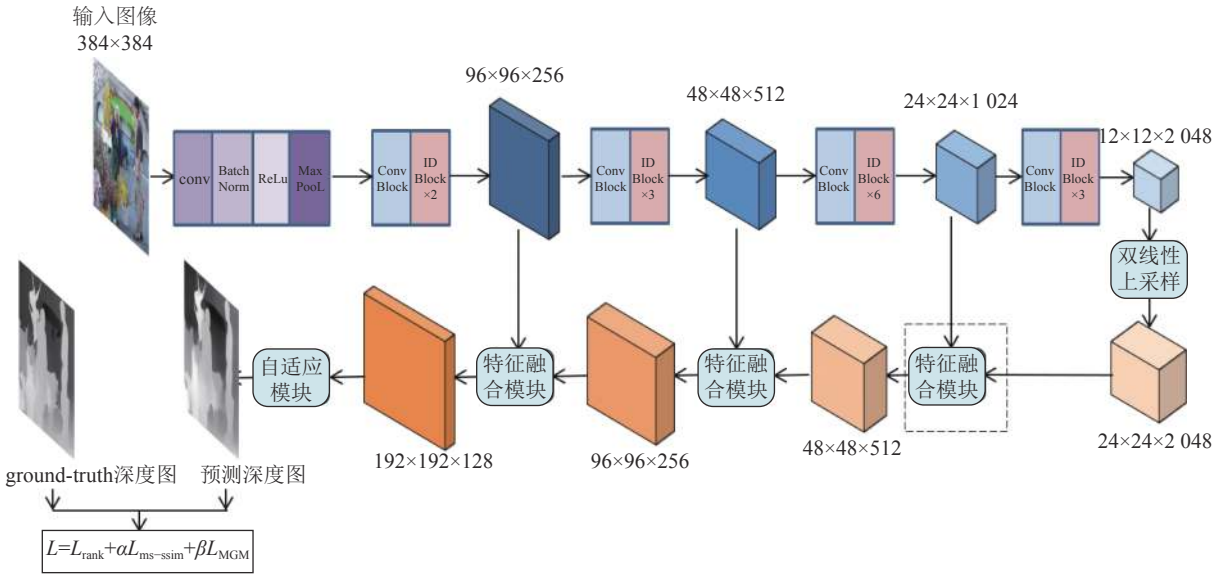


图 1　深度估计网络架构

解码器中特征融合部分的前向传播过程为：首先对 ResNet50 生成的最后一组特征图进行上采样；然后将编码器部分获取到的特征图与上层融合特征图通过多尺度特征融合模块得到下层融合特征图，具体如图 2 所示：对由编码器获取到的特征图使用一个残差卷积块，再将其与上层融合特征图进行合并，最后将合并的结果再通过一个残差卷积块以及上采样，以生成与下一个输入块的分辨率相同的特征图。为了生成最终的深度预测结果，将通过 3 个特征融合模块后的得到的特征结果输入到自适应输出模块，此模块包括两个 3×3 卷积层和一个双线性上采样层，得到最终的深度预测图像。

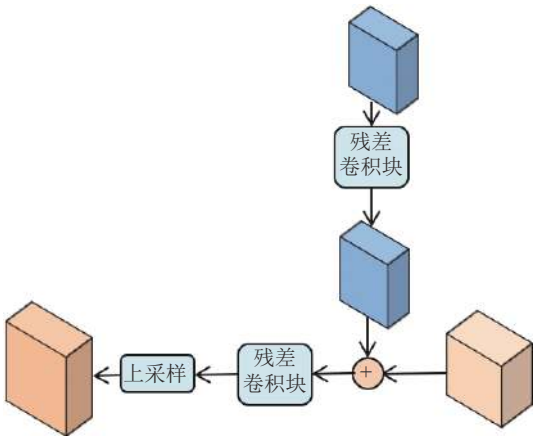


图 2　特征融合模块

1.2 联合损失函数

本文提出的联合损失函数为：

$$L = L_{\text{rank}} + \alpha L_{\text{ms-ssim}} + \beta L_{\text{MGM}} \tag{1}$$

式中，第一项损失L_{rank}为排序损失，用来训练图像中的相对深度以及惩罚预测深度图中像素对之间错误的排序关系；α、β为平衡因子。

对于每张输入图像I，随机采样N个相对深度点对(i, j)，其中i和j分别代表点对中第一个和第二个点的位置，总排序损失L_{rank}可表示为：

$$L_{\text{rank}} = \frac{1}{N}\sum_{i=1}^{N}\phi(p_i, p_j) \tag{2}$$

深度点对(i, j)在相应预测深度图像上的深度值为(p_i, p_j)，用$\phi(p_i, p_j)$表示预测深度中成对排序损失：

$$\phi(p_i, p_j) = \begin{cases} \log(1+\exp(-l_{ij}(p_i - p_j))) & l_{ij} \neq 0 \\ (p_i - p_j)^2 & l_{ij} = 0 \end{cases} \tag{3}$$

式中，l_{ij}为排序标签。为了获取每对点对之间的排序标签，首先从 GT 深度图中获取深度值(p_i^*, p_j^*)，然后获得 GT 点对深度排序标签，有：

$$l_{ij} = \begin{cases} +1 & p_i^*/p_j^* >= 1+\tau \\ -1 & p_i^*/p_j^* <= \dfrac{1}{1+\tau} \\ 0 & \text{其他} \end{cases} \tag{4}$$

式中，τ为阈值。

联合损失函数中第二项 ($L_{\text{ms-ssim}}$) 为多尺度结构相似度损失[13]，是一种方便的融合不同分辨率图像细节的损失。该损失在结合图像分辨率和查看条件的变化方面比单尺度相似度损失提供了更大的灵活性。多尺度结构相似度损失用于预测输入图像中的几何形状，从而提高深度估计的准确度：

$$L_{\text{ms-ssim}} = 1 - \text{MSSSIM} = 1 - [l_M(p, p^*)]^{\alpha_M}\prod_{j=1}^{M}\left[c_j(p, p^*)\right]^{\beta_j}\left[s_j(p, p^*)\right]^{\gamma_j} \tag{5}$$

式中，$c_j(p, p^*)$、$s_j(p, p^*)$分别表示在尺度为j时，预测深度与 GT 深度在对比度和结构上的比较；$l_M(p, p^*)$表示仅在最高尺度M时在亮度上的对比；参数α_M、β_j、γ_j用于调整不同成分的相对重要性。为了简化参数选择，在尺度j的情况下，设置$\alpha_j = \beta_j = \gamma_j$。

联合损失函数中第三项 (L_{MGM}) 为尺度不变梯度匹配损失，用于改善仅使用排序损失带来的边缘模糊问题，实现与 GT 中的不连续处相一致以及梯度平滑，将梯度匹配项定义为[14]：

$$L_{\text{MGM}} = \frac{1}{M}\sum_{s}\sum_{i}\left(\left|\nabla_x R_i^s\right| + \left|\nabla_y R_i^s\right|\right) \tag{6}$$

式中，M表示 GT 深度图的像素值；R_i^s表示在不同尺度s下预测深度值p和 GT 深度值p^*之间的差值，s设置为 4 个尺度。本文的训练实验中设置α=β =0.5。

2 数值仿真结果

2.1 实验设置

本文基于深度学习框架 Pytorch，计算用 CPU 为 NVIDIA GTX 1080ti，操作系统为 Centos 7.0。实验过程中，训练网络参数采用随机梯度下降 (stochastic gradient descent, SGD) 优化算法。

本文在训练深度预测网络时用高分辨率网络双目图像 (HR-WSI) 数据集，这是从网上收集的高分辨率双目图像的多样化集合。此数据集使用 FlowNet2.0 生成的视差图作为数据集中 ground-truth 部分，并且使用前后向流一致性屏蔽图像中的异常值。此外，通过预训练的网络计算高质量的天空分割掩模，并将天空区域的视差设置为最小观测值。通过手工剔除不良 GT 数据后，此数据集包括 20 378 张图像用于训练，400 张图像用于验证。

为了适应深度估计网络的输入，将图片尺寸随意裁剪成为 384×384，并且对裁剪后的图片进行归一化处理。网络训练时的批大小选为 4，训练周期设为 80。训练时编码器部分的学习率设置为10^{-5}，解码器部分的初始学习率为10^{-4}。

2.2 数值实验

为了测试深度估计模型的准确性与泛化能力，本文选择了 4 种数据集进行测试：Ibims[15]、NYUDv2[16]、DIODE[17]、Sintel[18]。以下简要概述这 4 种数据集。

2.2.1 Ibims 数据集

Ibims 为一组室内数据集，包含高分辨率和低分辨率的各种室内场景的 100 组 RGB-D 图像对。由数字单镜头反射相机和高精度激光扫描仪组成的定制采集，用于采集各种室内场景的高分辨率图像和高度精确的深度图。与相关的 RGB-D 数据集相比，Ibims 数据集具有噪声非常低、ground-truth 深度清晰、无遮挡以及范围广等优点。

2.2.2 NYU Depth 数据集

NYU Depth 数据集由来自各种室内场景的视频序列组成，这些视频序列由来自 Microsoft Kinect

的 RGB 和 Depth 摄像机记录。其中包含 1 449 个密集标记的 RGB 和深度图像对，该数据集中被标记的数据集为视频数据的子集，并带有密集的多类标记。标记数据集是原始 NYU Depth 数据集的子集，由成对的 RGB 帧和深度帧组成，并为每个图像标注了密集标签。

2.2.3　DIODE 数据集

DIODE 数据集为一组包含各种高分辨率彩色图像的数据集，具有准确、密集、远距离深度测量的特点，是第一个包含由一个传感器组获得的室内和室外场景 RGBD 图像的公共数据集。

2.2.4　Sintel 数据集

由于 ground-truth 光流很难以自然运动在真实场景中进行测量，所以光流数据集在尺寸大小、复杂性和多样性方面受到限制，使得光流算法难以在实际数据上进行训练和测试。文献 [18] 引入了一个新的光流数据集，该数据集来自开源 3D 动画短片 Sintel，具有在流行的 Middlebury 流评估中不具备的长序列、大运动、镜面反射、运动模糊、散焦模糊和大气影响等重要特征。由于生成电影的图形数据是开源的，因此能够在复杂度不同的情况下渲染场景，以评估现有流算法失败的地方。

4 种数据集在数值比较实验中采用了排序误差[19]来评价深度预测的准确性，深度边界误差 (depth boundary error, DBE)[15]来评价预测深度图的边缘准确性。

1) 深度预测的准确性对比

本文算法结果同 ReDWeb、Youtube3D、HR-WSI 深度估计方法的结果进行了对比试验；同时，为了研究在不同损失函数下预测深度图的准确性，实验也对 Ours_MS-SSIM、Ours_MGM 和 Ours_ALL 方法进行了数值比较。其中，Ours_MS-SSIM 采用 HR-WSI 作为训练集，在排序损失上再添加一项多尺度结构相似度损失；Ours_MGM 采用排序损失与多尺度尺度不变梯度匹配损失；Ours_ALL 采用本文提出的基于结构化的联合损失，将排序损失、多尺度结构相似度损失以及梯度匹配损失相结合作为损失函数。

在 4 个数据集下排序误差的数值结果如表 1 所示。算法对数据集中的每张图像随机采样 50 000 对相对深度点对来计算排序误差，排序误差的表达式为：

$$\varepsilon_{\text{ord}}=\frac{\sum_i \omega_i\left(l_i \neq l_{i,\tau}^{*}(p)\right)}{\sum_i \omega_i} \tag{7}$$

式中，ω_i设置为 1，并且使用式 (2) 获得l_i和$l_{i,\tau}^{*}(p)$之间的排序标签。

表 1　4 种数据集下排序误差数值比较

%

方法	数据集			
	Ibims	NYUDv2	DIODE	Sintel
ReDWeb	25.55	21.10	37.94	22.09
Youtube3D	22.81	19.03	35.86	21.05
HR-WSI	22.46	18.68	35.89	21.20
Ours_MS-SSIM	21.68	18.58	**34.64**	21.16
Ours_MGM	22.12	18.82	35.55	**20.64**
Ours_ALL	**21.56**	**18.43**	34.94	20.76

根据表 1 中的实验结果可以看出，在 4 种测试集下本文算法的排序误差均低于前 3 种方法。①由于 Ibims、NYUDv2 仅包含室内场景数据集，场景中多为刚性物体，深度预测时对场景中物体的几何形状和边缘要求都很高，由表中二、三列可以看出，本文采用的基于多尺度结构相似度和梯度匹配的算法得到的排序误差最小，能够更准确地预测几何形状以及深度不连续处，从而预测的准确性最高；②数据集 DIODE 主要包含以建筑物为主的室外静态场景，所以在预测深度时更关注这些建筑物的几何结构，所以表 1 的第四列中使用 Ours_MS-SSIM 方法，在排序损失上仅添加结构相似度损失，更准确地预测图像中的几何形状，从而得到最好的数值结果；③Sintel 为 3D 动画短片视频帧数据集，这些视频帧的前景大多为非刚性的运动的人物，对几何形状要求不高，更着重于深度点对的排序准确性和深度不连续处的一致性，所以第五列中仅在排序损失上添加多尺度梯度匹配损失的 Ours_MGM 方法获得了最低的排序损失结果。

2) 深度图边缘准确性对比

为了评价预测深度图的边缘准确性，本文采用了深度边界误差 (DBE)[15] 作为度量标准，通过比较预测深度图与 GT 深度图中的边缘，检查预测的深度图是否能够以准确的方式表示所有相关的深度不连续性，在测试 Ibims 数据集上分别计算了准确误差$\varepsilon_{\text{DBE}}^{\text{acc}}$以及考虑缺失边缘的完整误差$\varepsilon_{\text{DBE}}^{\text{comp}}$。两种误差公式分别为：

$$\varepsilon_{\mathrm{DBE}}^{\mathrm{acc}}(Y)=\frac{1}{\sum_i\sum_j y_{\mathrm{bin};i,j}}\sum_i\sum_j e^*_{i,j}y_{\mathrm{bin};i,j} \tag{8}$$

$$\varepsilon_{\mathrm{DBE}}^{\mathrm{comp}}(Y)=\frac{1}{\sum_i\sum_j y^*_{\mathrm{bin};i,j}}\sum_i\sum_j e_{i,j}y^*_{\mathrm{bin};i,j} \tag{9}$$

式中，y_{bin}表示使用结构化边缘提取出的边缘；y^*_{bin}表示通过二值边缘图像的截短倒角距离获取的 GT 边缘。

根据表 2 中的实验结果可以看出，Ours_MGM 和 Ours_ALL 两种方法得出的$\varepsilon_{\mathrm{DBE}}^{\mathrm{acc}}$以及$\varepsilon_{\mathrm{DBE}}^{\mathrm{comp}}$明显小于其他方法。主要是由于这两种方法均在损失函数中添加了一项尺度不变梯度匹配损失，提高了与 GT 中不连续性处的一致性，从而改善仅使用排序损失带来的边缘模糊问题。

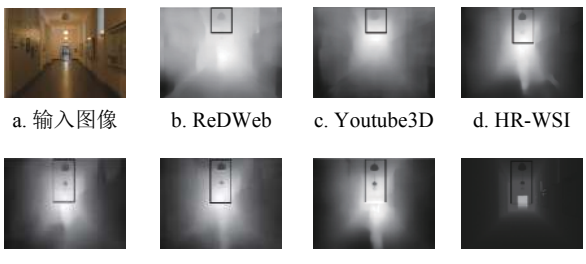
表 2　Ibims 数据集下深度边界误差 (DBE) 数值比较

方法	DBE	
	准确误差($\varepsilon_{\mathrm{DBE}}^{\mathrm{acc}}$)	完整误差($\varepsilon_{\mathrm{DBE}}^{\mathrm{comp}}$)
ReDWeb	2.640	7.379
Youtube3D	9.899	9.992
HR-WSI	2.413	6.995
Ours_MS-SSIM	2.311	7.065
Ours_MGM	**1.944**	6.834
Ours_ALL	2.007	**6.690**

2.3　主观评价

图 3～图 6 通过主观比较更直观地体现出不同单目深度估计方法下预测深度图的结果。

图 3 是在数据集 Ibims 上的对比，输入图像显示走廊上方吊灯数为 3 盏，图 3b、图 3c 只能预测出一盏灯的深度，图 3d 也只能隐约预测出第二盏灯，而图 3e、图 3g 可以在预测图像上呈现 3 盏灯的深度。由于 Ours_MS-SSIM、Ours_ALL 方法的损失函数中均包含多尺度结构相似度损失，所以可以更准确地预测图像中的几何形状。

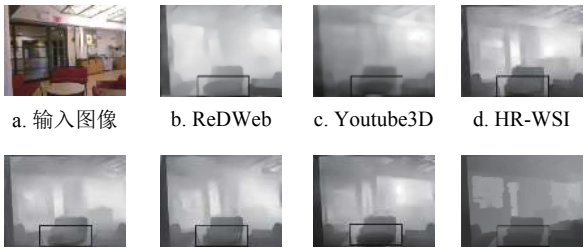


a. 输入图像　b. ReDWeb　c. Youtube3D　d. HR-WSI

e. Ours_MS-SSIM　f. Ours_MGM　g. Ours_ALL　h. GT

图 3　Ibims 测试集上的主观比较

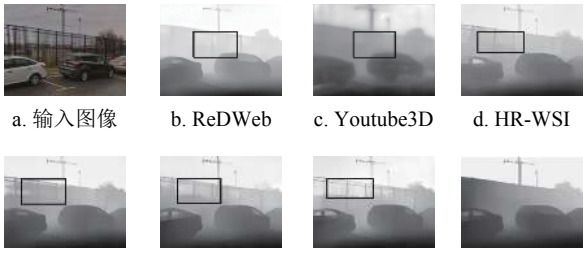
图 4 是在室内场景数据集 NYUDv2 上的对比，相比于图 4b、图 4c、图 4d，使用本文算法预测出的图 4g 中，靠近沙发的桌子边缘以及沙发自身边缘都更加清晰。同样的，图 5 为 DIODE 测试集上的比较，图 5g 中栏杆的清晰程度优于其他仅使用排序误差的深度预测方法。由图 4 和图 5 可以得出：本文算法在训练函数中添加一项梯度匹配损失，可以改善图像边缘的清晰程度，使得深度不连续处与 GT 图像更加具有一致性。



a. 输入图像　b. ReDWeb　c. Youtube3D　d. HR-WSI

e. Ours_MS-SSIM　f. Ours_MGM　g. Ours_ALL　h. GT

图 4　NYUDv2 测试集上的主观比较

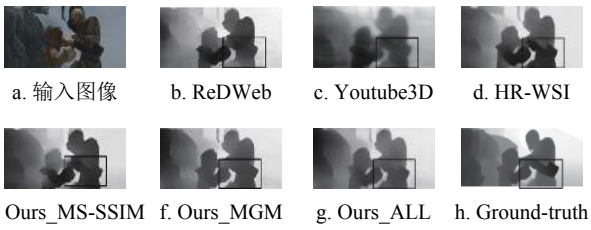


a. 输入图像　b. ReDWeb　c. Youtube3D　d. HR-WSI

e. Ours_MS-SSIM　f. Ours_MGM　g. Ours_ALL　h. Ground-truth

图 5　DIODE 测试集_上的主观比较

图 6 是在视频帧 Sintel 上的比较，左边女孩和右边男士的上身与下身的深度是一致的，表现在深度图上应该是深浅颜色几近相同，图 6f、图 6g 符合这一要求，可以看出使用本文的算法可以提高深度预测的准确性。



a. 输入图像　b. ReDWeb　c. Youtube3D　d. HR-WSI

e. Ours_MS-SSIM　f. Ours_MGM　g. Ours_ALL　h. Ground-truth

图 6　Sintel 测试集上的主观比较

3　结 束 语

为了提高单目深度估计精度，本文提出了基于多尺度结构相似度和梯度匹配的单目深度估计算

法。针对图像中几何形状无法准确预测以及边缘模糊的问题，在排序损失基础上添加了多尺度结构相似度和尺度不变梯度匹配损失，在单目深度估计过程中明显降低了排序误差和深度边界误差，有效提高了深度预测的准确性。实验对 Ibims、NYUDv2、DIODE、Sintel 4 个不同类型的数据集进行了评估，数值实验和主观评测结果表明，本文方法在定量和定性上都取得了更优的结果，并具有一定的泛化性能。

参 考 文 献

[1] VO M, NARASIMHAN S G, SHEIKH Y. Spatiotemporal bundle adjustment for dynamic 3d reconstruction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1710-1718.

[2] LU S, HANCA J, MUNTEANU A, et al. Depth-based view synthesis using pixel-level image inpainting[C]//2013 18th International Conference on Digital Signal Processing (DSP). Fira: IEEE, 2013: 1-6.

[3] ANANTRASIRICHAI N, GERAVAND M, BRAENDLER D, et al. Fast depth estimation for view synthesis[EB/OL]. [2021-01-15]. https://arxiv.org/abs/2003.06637.

[4] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4104-4113.

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

[6] 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). [S.l.]: IEEE, 2016: 239-248.

[7] 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. Honolulu, HI: IEEE, 2017: 1851-1858.

[8] LIU F, SHEN C, LIN G. Deep convolutional neural fields for depth estimation from a single image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 5162-5170.

[9] CHEN W, FU Z, YANG D, et al. Single-image depth perception in the wild[C]//Advances in Neural Information Processing Systems. [S.l.]: ACM, 2016: 730-738.

[10] CHEN W, QIAN S, DENG J. Learning single-image depth from videos using quality assessment networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 5604-5613.

[11] XIAN K, SHEN C, CAO Z, et al. Monocular relative depth perception with web stereo data supervision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 311-320.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

[13] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers. [S.l.]: IEEE, 2003: 1398-1402.

[14] LI Z, SNAVELY N. Megadepth: Learning single-view depth prediction from internet photos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2041-2050.

[15] KOCH T, LIEBEL L, FRAUNDORFER F, et al. Evaluation of cnn-based single-image depth estimation methods[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 8-14.

[16] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from RGBD images[C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 746-760.

[17] VASILJEVIC I, KOLKIN N, ZHANG S, et al. DIODE: A dense indoor and outdoor depth dataset[EB/OL]. [2021-01-15]. https://arxiv.org/abs/1908.00463.

[18] BUTLER D J, WULFF J, STANLEY G B, et al. A naturalistic open source movie for optical flow evaluation[C]//European conference on computer vision. Berlin, Heidelberg: Springer, 2012: 611-625.

[19] ZORAN D, ISOLA P, KRISHNAN D, et al. Learning ordinal relationships for mid-level vision[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 388-396.

编　辑　税　红