

采用传感器融合网络的单光子激光雷达成像方法

蒋筱朵, 赵晓琛, 冒添逸, 何伟基*, 陈 钱

(南京理工大学 电子工程与光电技术学院, 江苏 南京 210094)

摘要: 激光雷达系统采用主动照明的方式, 激光发射脉冲周期信号至目标场景, 激光脉冲经目标表面漫反射, 由单光子雪崩二极管 (Single-Photon Avalanche Diode, SPAD) 探测器记录回波光子的到达时间, 获得场景的深度信息。然而在探测过程中, 测量结果往往会遭到环境光的干扰。传感器融合是进行单光子成像的有效方法之一。最近提出的基于 LiDAR 和强度相机融合的数据驱动方法大多采用扫描式激光雷达, 深度获取速度慢。SPAD 阵列的出现打破了帧率的限制。SPAD 阵列允许同时收集多个回波光子的, 加速了信息采集, 但分辨率较低, 在探测过程中还会受到环境光的干扰, 因此需要通过算法打破 SPAD 阵列的固有限制, 从噪声中分离深度信息。针对分辨率为 32×32 pixel 的 SPAD 阵列探测器, 提出了一种卷积神经网络结构, 旨在强度图的引导下, 将低分辨率 TCSPC 直方图映射至高分辨率深度图。该网络采用多尺度方法提取输入特征, 并基于注意力模型融合深度数据和强度数据。另外, 设计了一个损失函数组合, 适用于处理 TCSPC 直方图数据的网络。在采集数据上进行了验证, 提出方法能成功将深度数据的空间分辨率提升 4 倍, 并在质量和数据指标上都优于其他算法。

关键词: 激光雷达; 单光子成像方法; 传感器融合; SPAD 阵列; 卷积神经网络

中图分类号: O439 **文献标志码:** A **DOI:** 10.3788/IRLA20210871

Single-photon LiDAR imaging method based on sensor fusion network

Jiang Xiaoduo, Zhao Xiaochen, Mao Tianyi, He Weiji*, Chen Qian

(School of Electronic and Optical Engineering, Nanjing University of Technology and Science, Nanjing 210094, China)

Abstract: LiDAR systems with active illumination obtain depth information of the scene using Single-Photon Avalanche Diode (SPAD) detectors to record the arrival time of reflected photons from the laser pulse. However, there is ambient light that interferes measurements during the detection period. Sensor fusion is one of the effective methods for single-photon imaging. Recently, many data-driven methods based on intensity-LiDAR fusion have achieved gratifying results, but most of them use the scanning LiDAR which has a slow depth acquisition speed. The advent of the SPAD array can overcome the limitation of frame rates. The SPAD array allows the collection of multiple returned photons at the same time, which accelerates the information collection process. However, the spatial resolution of SPAD array detectors is typically low, and the detection process is also interfered by the ambient light. Therefore, it is necessary to break the inherent limitation of the SPAD array through an algorithm to separate the depth information from the noise. In this paper, for the SPAD array detector with the array size of 32×32 pixel, a convolutional neural network was proposed, which could reconstruct high-

收稿日期: 2021-11-23; 修订日期: 2021-12-28

基金项目: 国家自然科学基金 (61875088, 62005128)

作者简介: 蒋筱朵, 女, 硕士生, 主要从事单光子激光雷达成像技术与应用方面的研究。

导师简介: 陈钱, 男, 副校长, 教授, 博士生导师, 博士, 主要从事光电成像与信息处理等方面的研究。

通讯作者: 何伟基, 男, 研究员, 博士生导师, 主要从事光电成像技术及应用方面的研究。

resolution clean TCSPC histogram under the guidance of the intensity image. A multi-scale approach was adopted to extract input features, and the fusion of depth data and intensity data was further processed based on the attention mechanism in the network. In addition, a loss function combination suitable for the TCSPC histogram data processing network was designed, where the overall distribution of photons and the ordinal relationship between time bins in the temporal dimension could be simultaneously considered. The method proposed in this paper can successfully increase the depth spatial resolution by 4 times, and the efficacy of proposed method is verified on realistic data, which is superior to state-of-the-art methods qualitatively and quantitatively.

Key words: LiDAR; single-photon imaging method; sensor fusion; SPAD array; convolutional neural network

0 引言

从感知场景中推断出正确的深度信息对许多应用来说是至关重要的,例如自动驾驶、虚拟现实、增强现实和机器人技术。激光雷达是深度成像中的领先技术,目前,大多数激光雷达系统采取单点/扫描的方式,使用共轴对齐的激光二极管和单光子探测器,由激光器发射激光,探测器时间标注经场景反射回来的到达光子。扫描式激光雷达系统虽然能够获取较为准确的深度信息,但采集速度慢。然而,越来越多的应用要求对场景进行快速获取,在此需求之下,单光子雪崩二极管(Single-Photon Avalanche Diode, SPAD)阵列应运而生。通过同时采集多个回波光光子,SPAD阵列可以提供准确快速的场景深度信息。

近年来,许多研究团队着力发展 SPAD 阵列^[1-2]。目前,激光雷达的分辨率较低,尤其是 SPAD 阵列,因此,深度重建也是一个热门的研究方向^[3-6],也有许多研究依赖于神经网络方法^[7-8]。仅从激光雷达系统中获取的信息进行深度重建效果是有限的,多维信息融合^[9]是解决这一问题的方法之一。Lindell 等人结合常规高分辨率相机和线列 SPAD,使用多尺度深度卷积网络,提出了一种用于效率 3D 成像的数据驱动方法^[10];在此基础上,Sun 等人引入了单目深度估计算法,能从强度信息中得到更可靠的相对距离信息^[11];Ruget 等人使用了相同的 SPAD 阵列传感器,基于神经网络,利用强度图和多个从直方图中提取出的特征引导深度上采样^[12]。

在人类视觉系统中,大脑会自动忽略场景中低价值的信息,为了模仿这一行为,注意力模型被提出。在神经网络中,注意力模型能够硬性选择输入的某些

部分,或者给输入的不同部分分配不同的权重,目前在各个领域被广泛使用^[13]。最近的工作将注意力模型应用于三维点云数据上,但解决的都是分类问题,文中将深度图像重构视为回归问题,将注意力模型嵌入处理时间相关单光子计数(Time-Correlated Single-Photon Counting, TCSPC)直方图数据的神经网络中,证明注意力模型在三维数据回归问题中的有效性。

为了打破 SPAD 阵列的固有图像分辨率限制和去除探测器探测过程中的噪声光子,论文基于传感器融合策略提出了一种卷积神经网络结构,引入多尺度特征提取和注意力机制模块,提高了融合质量。此外,设计了一个针对 TCSPC 直方图的损失函数,不仅关注光子在时间维度上的总体分布,还考虑各个时间仓间光子的序数关系。文中提出的方法可以将深度数据的空间分辨率提升 4 倍,并在仿真数据和真实采集数据上都取得了比其他算法更好的质量效果和量化指标。

1 数据获取

1.1 实验设置

实验设置如图 1 所示,可分为照明模块和感知模块。照明模块包括一个 635 nm 的皮秒脉冲激光器和一个散射片。激光器发射 20 MHz 重复频率的激光脉冲,激光经散射片被发散,以覆盖 SPAD 阵列的探测范围。感知模块包括一个 SPAD 探测器和常规相机。SPAD 探测器阵列的空间分辨率为 32×32 pixel,每一个探测器可独立运行于 TCSPC 模式,时间分辨率为 55 ps,并感知由目标反射回来的光子以获取激光脉冲的时间信息。在一个曝光周期内,每个像素仅探测第一个到达的光子,待下一个曝光周期到达后,

重置数字时间转换器 (Time-to-Digital Converter, TDC), 该过程即为一个探测帧。通过叠加多个探测帧的返回光子数, 每个像素都包含一个关于返回光子的时间直方图, 其中总光子数代表了激光脉冲在相应空间位置上的照射强度, 直方图形状表示了返回光子的到达时间。

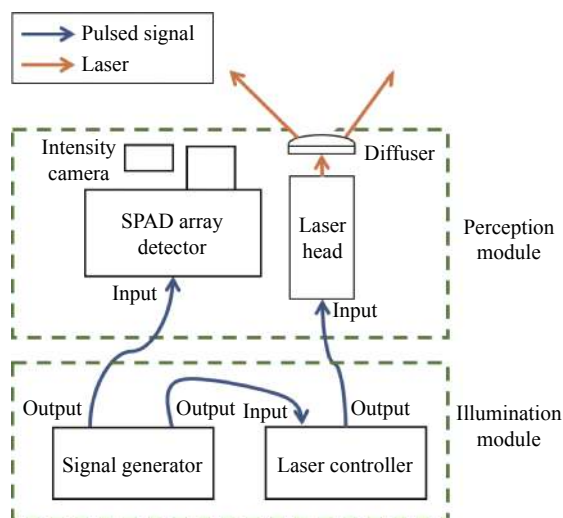


图 1 实验设置图

Fig.1 Experiment setup

SPAD 探测器能够探测到的光子主要分为三类: 在目标表面漫反射返回的信号光子、环境中的背景光子和探测器自身产生的暗电流噪声, 其中, 主要噪声光子来源为背景光子和暗计数。由于在探测过程中, 这两种噪声均处于强度稳定状态, 因此可以将其分布视为泊松分布。并且背景光子和暗计数相互独立, 所以二者的叠加噪声仍遵循泊松分布。

1.2 预处理

对于 SPAD 阵列探测器及漫散射照明的配准, 在约 1 m 位置处放置一块白色板, 调整激光发射二极管的位置, 以保证在探测器探测中心能够观察到激光点, 再在激光器前方放置散射片。对于高分辨率相机获取的强度图像和 SPAD 阵列探测器获取的深度图像配准, 通过探测棋盘格特征, 再应用映射转换以对齐两个图像。

与 Ruget 等人的工作^[12]相似, 对 SPAD 探测器获取的原始数据进行了预处理, 以提高数据质量, 减少网络处理数据量。但与该工作中使用多种方法提取

深度数据中的各尺度特征不同, 文中方法无需从原始数据中提取多个重要特征图, 仅将原始数据减掉在相同条件下、但没有激光传播所获取的环境光和暗计数率之和, 然后在时间维度上截取包含信号光子的 200 个时间仓, 最后在空间维度上应用最近邻法, 将深度数据上采样 4 倍, 得到网络的输入数据。

2 基于注意力模块的传感器融合网络

基于注意力模块的多阶段融合网络采用 U-Net 网络结构。网络的目标是由高分辨率干净强度图引导, 将上采样后的有噪 TCSPC 直方图映射成精细的高分辨率去噪直方图。损失函数为 Kullback-Leibler (KL) 散度、序数回归损失和全变差空间正则项的权重和, 训练过程的目标为找到滤波器的权重和偏差使损失函数趋于收敛。

2.1 网络结构

基于注意力模块的多尺度融合网络结构 (如图 2 所示) 由特征提取和融合重构两个模块组成, 其中特征提取模块用于提取 TCSPC 直方图数据和强度数据中的多尺度特征, 使网络可以学习不同尺度的丰富层次特征, 更好地适应精细和大规模尺度的上采样; 融合重构模块引入了注意力模型^[14], 使网络关注融合数据中特征更多的区域, 逐渐上采样至输入分辨率。损失函数由 KL 散度、序数回归损失和全变差正则项组成, 在网络输出 TCSPC 直方图上施加 KL 散度和序数回归损失以分别关注光子数在时间维度上的总体分布和每个时间仓之间的序数关系。为了保持重构结果的边缘, 在经 3 D-2 D 映射后得到的 2 D 深度图上施加全变差正则项。

其中, I 代表强度数据; D 代表 TCSPC 直方图数据; 下标 $0, \dots, L$ 表示不同尺度的特征, L 越大, 特征图分辨率越低; 上标“ \wedge ”为处理后的特征。网络输入为分辨率 $H \times W \times N_{time_bins}$ 的有噪 TCSPC 直方图数据 D_0 和分辨率为 $H \times W$ 的强度数据 I_0 , 通过特征提取器产生强度图和 TCSPC 直方图数据的各尺度特征, 使用经 2 D-3 D 映射算子的 3 D 强度图像作为引导, 进行数据融合重构, 最终输出 $H \times W \times N_{time_bins}$ 的去噪后的 TCSPC 直方图数据 \hat{D} , 经过 soft argmax 算子得到 2 D 深度图 \hat{D}' 。

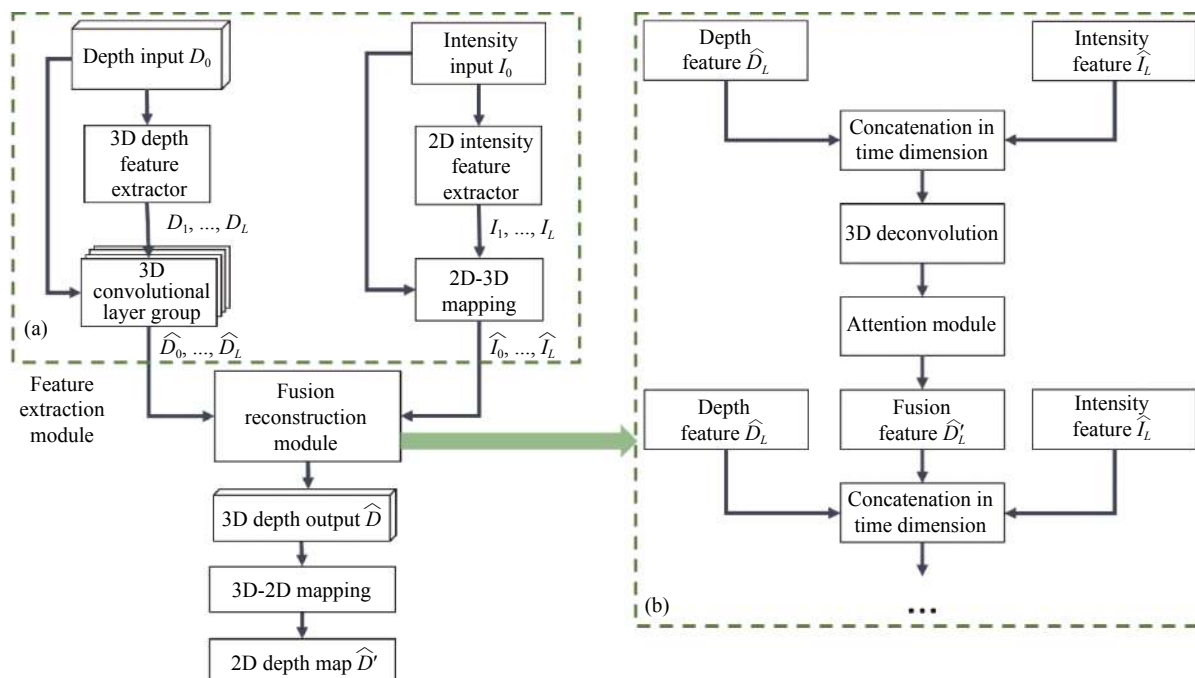


图 2 基于注意力模块的多尺度融合网络结构示意图

Fig.2 Schematic diagram of multi-scale fusion network structure based on attention module

2.2 损失函数

为了训练注意力融合网络,定义总损失函数为三个主要贡献的权重相加:在网络输出 TCSPC 直方图 \hat{D} 和干净 TCSPC 直方图 D 之间约束了 D_{KL} 和 $Loss_{OR}$, 以及用于 \hat{D} 的全变差项。总的训练损失函数为:

$$Loss_{total} = D_{KL} + \lambda_1 Loss_{OR} + \lambda_2 TV \quad (1)$$

对于网络输出 TCSPC 直方图 \hat{D} 和干净 TCSPC 直方图 D ,施加了两个约束项, D_{KL} 关注网络重构后的 TCSPC 直方图数据 \hat{D} 和地面真实 TCSPC 直方图数据 D 之间的概率分布差异,同时添加 $Loss_{OR}$ 以考虑 TCSPC 直方图时间仓之间的序数关系。

2.2.1 KL 散度

KL 散度,又称相对熵,用来度量网络输出直方图 \hat{D} 和干净直方图 D 之间概率分布的相似性:

$$D_{KL}(D||\hat{D}) = \sum_n \left[D(n) \log \frac{D(n)}{\hat{D}(n)} \right] \quad (2)$$

式中: n 代表时间仓索引。概率分布越相近, KL 散度值越小。

2.2.2 序数回归损失

由于 KL 散度独立计算各个时间仓的分布,因此引入了序数回归损失,考虑每个时间仓之间的序数关系:

$$Loss_{OR}(D, \hat{D}) = \frac{-1}{W \times H} \sum_w \sum_h \psi(D, \hat{D})$$

$$\psi(D, \hat{D}) = \sum_{n=1}^l \log(1 - cumsum(\hat{D}(n))) + \sum_{n=l+1}^K \log(cumsum(\hat{D}(n))) \quad (3)$$

式中: W 和 H 为图像宽度与高度; l 为地面真实检测速率峰值的时间仓索引;“cumsum”代表累计求和。

2.2.3 全变差正则项

引入 TV 空间正则项的目的在于去除伪影,保留边缘等重要细节,同时防止模型过拟合:

$$TV(\hat{D}) = \sum_{w=1}^{W-1} \sum_{h=1}^{H-1} |\hat{D}_{w+1,h} - \hat{D}_{w,h}| + \sum_{w=1}^{W-1} \sum_{h=1}^{H-1} |\hat{D}_{w,h+1} - \hat{D}_{w,h}| \quad (4)$$

$$\hat{D} = \text{softargmax}(\hat{D}) = \sum_n n \cdot \hat{D}(n)$$

考虑到网络的主要目的在于去除原始 TCSPC 直方图数据中含有的噪声光子,会给予 KL 散度较大的权重。在保证具有良好的去噪能力基础上,再提升重构图像的精细程度,即序数回归损失和全变差正则项。由于全变差正则项作用于 2D 图像上,而网络主要处理数据为 3D 的 TCSPC 直方图数据,因此给予全变差正则项较小权重。综上所述,在训练过程中,

设置 KL 散度的权重为 1, 序数回归损失的权重设置为 $\lambda_1 = 0.5$, 全变差正则项的权重设置为 $\lambda_2 = 10^{-4}$ 。

3 实验与分析

该节共进行了三个实验: 比较实验与其他深度重构方法对比, 证明文中方法的有效性和优越性; 消融实验验证所提出的网络结构和损失函数的合理性和必要性; 上采样实验验证预上采样法的重要性。所有网络使用 NYU V2 数据集^[15] 进行训练, 量化结果采用根均方差 (Root Mean Square Error, RMSE) 指标, 单位

为 m。

实验测试数据由 1.1 节提到的实验装置采集, 从 SPAD 探测器中得到的 TCSPC 直方图经预处理后输入至训练好的网络。地面真实数据为 SPAD 探测器输出的原始数据经过预处理, 再使用中值滤波去除探测器暗计数, 最后给图像反射率设定阈值, 将像素反射率小于阈值的像素设定为背景得到。

3.1 比较实验

比较实验的量化结果如表 1 所示, 文中提出方法在各场景中都获得了最佳结果。

表 1 实验量化结果比较

Tab.1 Comparison of experimental quantitative results

	Preprocessed	MLE	He et al. ^[16]	Lindell et al. ^[10]	Proposed
"N" and "J"	0.8489	0.4823	0.3970	0.3253	0.3068
Multi objects	0.7204	0.4510	0.6129	0.2432	0.1958

对比实验将文中方法与 MLE、He 等人^[16] 提出的方法和 Lindell 等人^[10] 提出的方法进行比较。相比于传统方法 MLE 和 He 等人^[16] 提出的方法, 使用神经网络, 即 Lindell 等人^[10] 提出的方法和文中提出的方法, 能够学习输入和输出之间非线性的复杂映射, 灵活适应不同的成像场景。MLE 不采用传感器融合策略, 这种方法给定了概率模型, 不能处理掉探测器的暗计数及探测过程中产生的异常值, 重构结果仍存在大量噪声。He 等人^[16] 提出的方法通过引导图找出物

体的边缘, 滤波器在图像平滑的地方进行均值滤波, 而在边缘处不进行滤波, 或者进行轻微的滤波, 从而达到保留物体边缘的目的。这种方法不能滤除物体边缘处的噪声。Lindell 等人^[10] 提出的数据驱动方法采用传感器融合策略和多尺度方法, 但仅在最大尺度的深度特征图上融合强度特征, 没有充分利用强度信息, 会造成严重的深度缺失现象。如图 3 所示, MLE 方法并不能完全除去噪声, 而 He 等人^[16] 的方法过于平滑物体边缘。Lindell 等人^[10] 的卷积神经网络方法

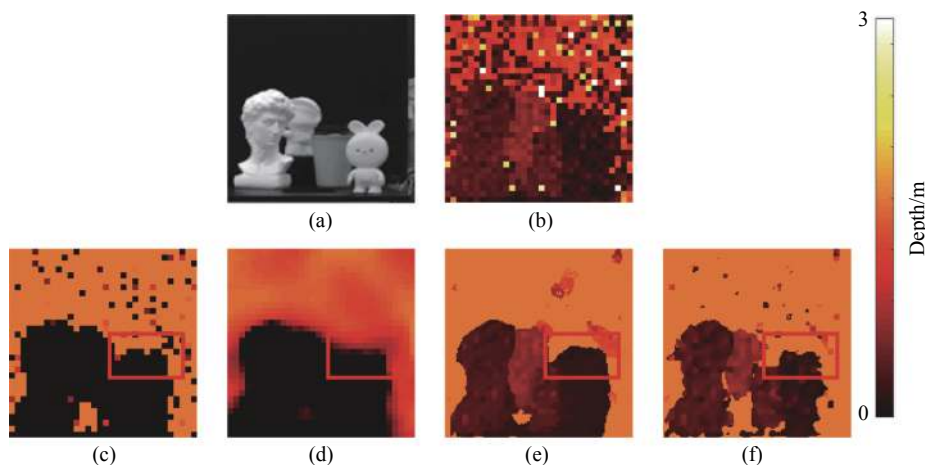


图 3 (a) 网络输入强度图; (b) 网络输入深度图; (c) MLE 方法处理结果; (d) He et al. ^[16] 方法处理结果; (e) Lindell et al. ^[10] 方法处理结果; (f) 文中方法处理结果

Fig.3 (a) The input intensity map; (b) The input depth map; (c) The result processed by MLE method; (d) The result processed by He et al. ^[16]; (e) The result processed by Lindell et al. ^[10]; (f) The result processed by proposed method in this paper

能重构出场景,但是会造成部分边缘深度缺失的现象,特别是对于较远物体和深度值较少区域。文中提出方法能够可靠恢复场景深度信息,并且对远处物体和细小物体也具有重构鲁棒性。

3.2 消融实验

不同网络结构的消融研究比较了不含注意力模块的网络和没有强度引导的网络,实验结果如图 4(c)、(d)所示。不含注意力模块的网络对特征图的每一部分都给予相同的关注,而没有强度引导的网络无法提取更精确的边缘等细节特征,重构性能均不佳。文中提出的网络使用强度引导并引入注意力机制,可以从强度图中学习细节特征,也可以关注融合数据中特征更丰富的区域,能够去除绝大部分的噪声,目标边缘清晰。

不同损失函数的消融研究在训练过程中使用不考虑序数回归损失的损失函数和不考虑 KL 散度的损失函数,如图 4(e)、(f)所示。不考虑序数回归损失的损失函数训练的网络无法重构出物体完整的边缘,原因在于 KL 散度关注的为 TCSPC 直方图上光子的总体分布,仅滤除了和信号光子差异明显的背景光子,无法去除目标边缘受到回波光子微弱影响的背景像素;而使用没有 KL 散度的损失函数训练的网络进行重构,物体内部存在深度缺失,边缘呈锯齿状,这是由于序数回归损失考虑的是局部的时间仓间的序数关系,而不考虑整个时间维度上的光子数分布。文中设计的损失函数结合了 KL 散度和序数回归损失,并赋予了不同权重,不仅关注时间维度上光子的总体分布,也考虑每个时间仓间的序数回归关系。使用文中设计的损失函数训练得到的网络重构结果不仅具有目标轮廓,并且像素呈连续性。

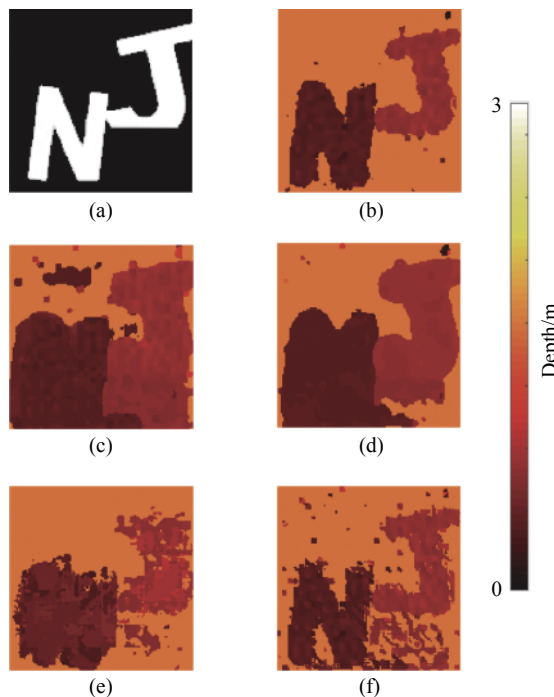


图 4 (a) 网络输入强度图; (b) 文中提出方法处理结果图; (c) 不含注意力模块的网络处理结果图; (d) 无强度引导的网络处理结果; (e) 使用不含序数回归损失的损失函数的网络处理结果; (f) 使用不含 KL 散度的损失函数的网络处理结果

Fig.4 (a) Network input intensity map; (b) The result of the method proposed in the paper; (c) The result processed by the network without attention module; (d) The result processed by the network without intensity guidance; (e) The result processed by the network of loss function without ordinal regression loss; (f) The result processed by the network without KL divergence

表 2 为消融实验的量化结果,可以看出使用注意力模块和强度引导的网络结构,在训练过程中同时采用 KL 散度和序数回归损失进行约束,即文中提出方法,能够获得最佳的量化结果。

表 2 消融实验量化结果

Tab.2 Ablation experimental quantitative results

	Without attention	Without intensity	KL + TV	OR + TV	Proposed
"N" and "J"	0.7204	0.4510	0.6129	0.2432	0.1958

3.3 上采样实验

文中使用预上采样法,即在输入网络前将 SPAD 阵列原始数据的空间分辨率从 32×32 pixel 提升至 128×128 pixel。上采样的表现形式之一为稀疏

点云更加密集。对比低分辨率点云图、使用后上采样法产生的点云图(先将数据输入网络处理,再进行上采样)和预上采样法产生的点云图,结果如图 5 所示,预上采样法提高了深度数据携带的信息量,使网络可



图 5 (a) 没有上采样的点云图; (b) 使用后上采样法的点云图; (c) 使用预上采样法的点云图

Fig.5 (a) Point cloud without upsampling; (b) Point cloud with post-upsampling method; (c) Point cloud with pre-upsampling method

以处理更多的像素, 重构结果像素具有空间联系, 边缘平滑。

4 结 论

文中针对深度重构研究方向, 介绍了一种基于传感器融合策略的卷积神经网络结构, 并结合了注意力模型, 产生了更好的融合效果。此外, 文中设计了一种损失函数, 适用于处理 TCSPC 直方图数据的算法, 同时关注时间维度上光子的总体分布和各个时间仓之间的序数回归关系。文中提出的卷积神经网络结构简单, 无需过多的预处理步骤, 并在由 SPAD 阵列探测器获取的数据上验证了深度重构的鲁棒性。在比较实验中, 文中提出的方法处理结果能够重构出边缘, 物体深度完整; 相比于其他深度重构方法, 最好可以将量化指标提高 3 倍。在消融实验中, 使用文中设计的网络结构和损失函数得到的处理结果都取得了最佳图像质量。这些实验结果均验证了文中方法具有优异的深度重构能力, 在实际应用中具有潜力。

参考文献:

[1] Henderson R K, Johnston N, Hutchings S W, et al. 5.7 A 256×256 40 nm/90 nm CMOS 3D-stacked 120 dB dynamic-range reconfigurable time-resolved spad imager[C]//ISSCC, 2019: 106–108.
 [2] Poland S P, Krstajić N, Monypenny J, et al. A high speed multifocal multiphoton fluorescence lifetime imaging

microscope for live-cell FRET imaging [J]. *Biomed Opt Express*, 2015, 6(2): 277-296.
 [3] Huang He, Li Xinrui, Song Jing, et al. A traffic image dehaze method based on adaptive transmittance estimation with multi-scale window [J]. *Chinese Optics*, 2019, 12(6): 1311-1320. (in Chinese)
 [4] Shin D, Xu F, Venkatraman D, et al. Photon-efficient imaging with a single-photon camera [J]. *Nat Commun*, 2016, 7(1): 12046.
 [5] Feng Xiaowei, Hu Haiyun, Zhuang Ruiqing, et al. Adaptive reconstruction of 3D point cloud by sparse optimization [J]. *Optics and Precision Engineering*, 2021, 29(10): 2495-2503. (in Chinese)
 [6] Rapp J, Goyal V K. A few photons among many: Unmixing signal and noise for photon-efficient active imaging [J]. *IEEE Trans Comput Imaging*, 2017, 3(3): 445-459.
 [7] Wang Chunzhe, An Junshe, Jiang Xiujie, et al. Region proposal optimization algorithm based on convolutional neural networks [J]. *Chinese Optics*, 2019, 12(6): 1348-1361. (in Chinese)
 [8] Zhou Hongqiang, Huang Lingling, Wang Yongtian, et al. Deep learning algorithm and its application in optics [J]. *Infrared and Laser Engineering*, 2019, 48(12): 1226004. (in Chinese)
 [9] Zeng Hanlin, Meng Xiangyong, Qian Weixian, et al. Image fusion algorithm based on DOG filter [J]. *Infrared and Laser Engineering*, 2020, 49(S1): 20200091. (in Chinese)
 [10] Lindell D B, O'Toole M, Wetzstein G. Single-photon 3D imaging with deep sensor fusion [J]. *ACM Trans Graph*, 2018, 37(4): 1-12.
 [11] Sun Z H, Lindell D B, Solgaard O, et al. SPADnet: Deep RGB-SPAD sensor fusion assisted by monocular depth estimation [J]. *Opt Express*, 2020, 28(10): 14948-14962.
 [12] Ruget A, McLaughlin S, Henderson R K, et al. Robust super-resolution depth imaging via a multi-feature fusion deep network [J]. *Opt Express*, 2021, 29(8): 11917-11937.
 [13] Lei Junfeng, He Rui, Xiao Jinsheng. Driving obstacles prediction network merged with spatial attention [J]. *Optics and Precision Engineering*, 2020, 28(8): 1850-1860. (in Chinese)
 [14] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//ECCV, 2018: 3-19.
 [15] Silberman N, Hoiem D, Kohli P, et al. Indoor segmentation and support inference from RGBD images[C]//ECCV, 2019: 746-760.
 [16] He K M, Sun J, Tang X O. Guided image filtering [J]. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35(6): 1397-1409.