

# 多重注意力非局部特征融合的序列影像 - 激光点云深度补全

张晨<sup>1\*</sup>, 张飞<sup>2</sup>, 陈驰<sup>1</sup>, 杨必胜<sup>1</sup>

<sup>1</sup>武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉

<sup>2</sup>上海维智卓新信息科技有限公司, 上海

收稿日期: 2022年3月30日; 录用日期: 2022年4月21日; 发布日期: 2022年4月27日

## 摘要

有效且准确的深度信息能够精确感知场景的几何结构。目前主要采用的是激光雷达传感器。但由于其扫描线数有限, 深度信息的稠密度非常低。由此提出深度补全任务(Depth Completion), 基于给定的稀疏深度值来估计稠密深度信息。本文针对先前深度补全存在的边界模糊和混合深度的问题, 提出一种多重注意力非局部特征融合的序列影像 - 激光点云深度补全模型。该模型通过融合多序列影像和激光点云, 实现了多模态数据的优势互补, 通过网络学习到更丰富的特征。实验表明该方法可以有效降低误差, 提高深度补全的效果, 极大地提高例如目标识别、目标跟踪、路径规划等任务的精度。

## 关键词

深度补全, 注意力机制, 自校准卷积, 非局部特征, 点云

# Depth Completion between Single Image and Sparse LiDAR Data with Multiple Attention and Nonlocal Feature

Chen Zhang<sup>1\*</sup>, Fei Zhang<sup>2</sup>, Chi Chen<sup>1</sup>, Bisheng Yang<sup>1</sup>

<sup>1</sup>State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan Hubei

<sup>2</sup>Shanghai Figure Interesting Information Technology Co., Ltd., Shanghai

Received: Mar. 30<sup>th</sup>, 2022; accepted: Apr. 21<sup>st</sup>, 2022; published: Apr. 27<sup>th</sup>, 2022

\*通讯作者。

文章引用: 张晨, 张飞, 陈驰, 杨必胜. 多重注意力非局部特征融合的序列影像 - 激光点云深度补全[J]. 测绘科学技术, 2022, 10(2): 111-120. DOI: 10.12677/gst.2022.102011

## Abstract

Effective and accurate depth information can accurately perceive the geometry of the scene. At present, LiDAR sensors are mainly used. But due to the limited number of scanning lines, the density of depth information is very low. Therefore, a depth completion task is proposed to estimate the dense depth information based on the given sparse depth map. In order to solve the problems of boundary blur and mixed depth in previous depth completion tasks, this paper proposes a depth completion network between single image and sparse LiDAR data with multiple attention and nonlocal feature. The fusion of image and LiDAR data realizes the complementary advantages of multimodal data, and richer features can be learned through the network. Experiments show that this network can effectively reduce the error, improving the effect of depth completion, and greatly improve the accuracy of tasks such as target recognition, target tracking, path planning and so on.

## Keywords

Depth Completion, Attention Model, Self-Calibrated Convolutions, Nonlocal Feature, Point Cloud

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,随着自动驾驶[1]、增强现实[2]、虚拟现实[3]、3D 场景重建[4]等计算机视觉应用的快速发展,深度信息的感知和获取已成为一个重要问题。为了获得可靠的深度预测,通常利用来自各种传感器的多模态信息,例如 RGB 相机、雷达、激光雷达(LiDAR)和超声波传感器。深度传感器(如 LiDAR 传感器)可以高频率地采集准确的深度信息。然而,由于扫描线数量有限,采集的深度信息往往非常稀疏。为了克服这些局限性,已有大量的工作是基于给定的稀疏深度值来估计稠密深度信息,称为深度补全。

早期的深度补全算法[5][6]仅选择稀疏深度值作为输入。由于缺乏场景信息和输入稀疏程度不确定的问题,生成的稠密深度图存在边缘模糊和输入数据高度稀疏的情况下无法完成补全等问题。由于 RGB 图像包含丰富的颜色信息和场景信息,最近的研究将 RGB 图像融合进深度补全算法,以获取更多的特征信息,进而估计精确的稠密深度图。

融合 RGB 图像和稀疏深度图的深度补全算法[7]选择 RGB 图像和稀疏深度图作为输入,并使用深度卷积神经网络(Convolutional Neural Network, CNN)直接预测出稠密深度图。与早期的深度补全算法相比,深度补全结果有了显著改善。然而,它们仍然会在接近深度边界的地方生成模糊的深度图。在基于亲和力的空间特征传播网络提出后,上述问题得以解决[8]。通过学习局部邻域像素的亲性和,并迭代地细化深度预测,最终的稠密深度图变得更加准确。但是,最初的空间传播网络[9]设计在固定的局部邻域内传播。固定的局部邻域通常有不相关的信息,在特征传播后,会导致前景和背景对象的深度值混合,特别是深度边界附近。因此,在深度补全任务中仍存在混合深度问题。

综上,深度补全任务存在边缘模糊和混合深度的问题,因此本文基于残差网络提出融合多重注意力的自校准卷积(Self-Calibrated Convolutions, SC-Conv)深度补全网络,并利用非局部特征传播网络迭代地优化初始深度预测,得到更精确的稠密深度图。

## 2. 方法

本文提出了一种多重注意力非局部特征融合的序列影像-激光点云深度补全方法, 该方法的主体流程如图 1。首先, 将 RGB 图像和稀疏深度图一起输入融合多重注意力的自校准卷积深度补全网络, 通过编码器-解码器结构分别获取颜色特征和深度特征, 输出预测的初始深度图以及对应的置信度, 同时输出非局部邻域以及亲缘关系。接着, 将上述输出一起输入非局部特征空间传播网络对得到的预测初始深度图进行优化, 根据预测的非局部邻域和亲缘关系, 通过置信度和非局部空间传播过程迭代地优化改进初始深度预测, 最后得到最终深度图。融合多重注意力的自校准卷积深度补全网络的具体流程如图 2, 非局部特征空间传播网络补全深度优化的具体流程如图 7。

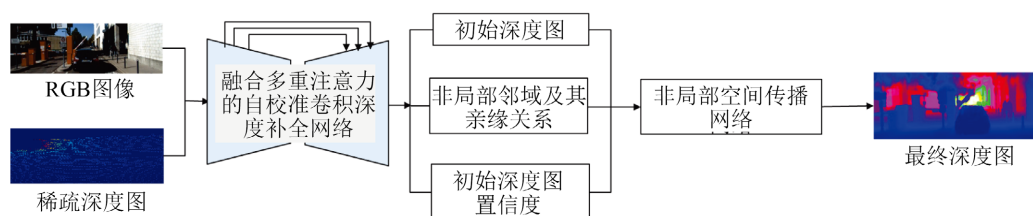


Figure 1. Framework of the proposed method

图 1. 本文方法的流程图

### 2.1. 融合多重注意力的自校准卷积深度补全网络

如图 2 所示, 本文设计了一种融合多重注意力的自校准卷积深度补全网络。将 RGB 图像和稀疏深度图先分别进行卷积操作, 再将融合后的图像送入编码部分。在编码阶段, 以残差网络 ResNET50 为基础, 将每个残差模块的  $3 \times 3$  高层卷积组替换为自校准卷积, 得到自校准残差模块(SC-block), 构造自校准残差网络(SCNet), 提取更丰富的特征信息。在 SCNet 网络先通过 64 个大小为  $7 \times 7$  的卷积核, 步长为 2, 后连接一个批归一化层(BN)和激活函数层(ReLU)。之后依次通过 E2—E5 四个编码层, 每个卷积后都连接一个 BN 层和 ReLU 层。E2 包含 3 个 SC-block, E3 包含 4 个 SC-block, E4 包含 6 个 SC-block, E5 包含 3 个 SC-block。E2 中每个 SC-block 的卷积核步长均为 1, 在经过 E2 处理后特征图大小不变, 输出特征图维度与卷积核个数一致。E3、E4、E5 中每个 SC-block 的卷积核步长均为 2, 依次经过 E3、E4、E5 处理后, 特征图大小依次变为初始特征图的  $1/2$ 、 $1/4$ 、 $1/8$ 。SC-block 的基本结构如图 3 所示。

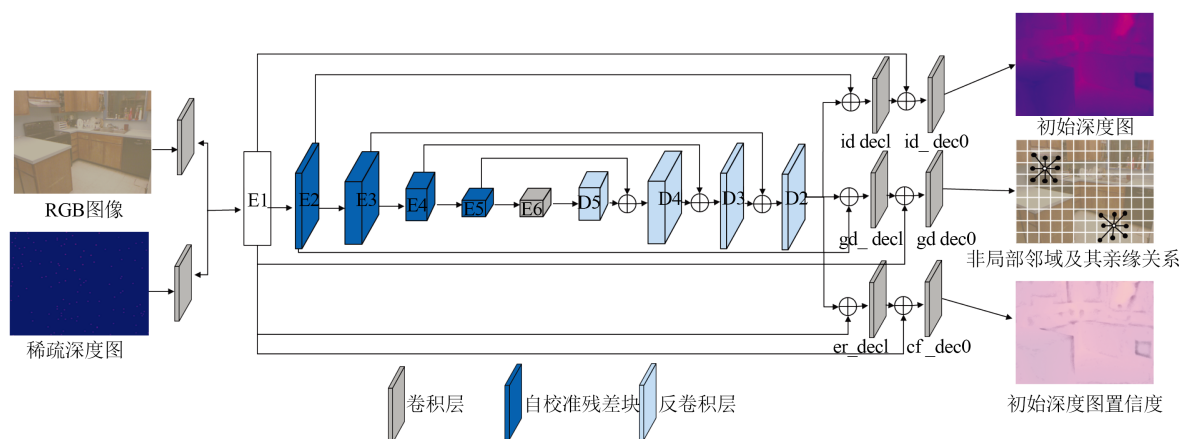


Figure 2. Self-calibration convolution depth completion network with multiple attention

图 2. 融合多重注意力的自校准卷积深度补全网络

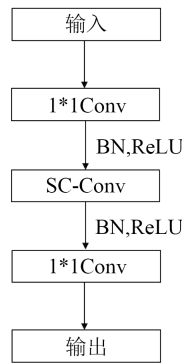


Figure 3. Basic structure of SC-block  
图 3. 自校准残差模块基本结构

在解码阶段之前，先进行了一次卷积 E6，步长为 2，输入通道数与输出通道数一致，特征图大小变为初始特征图的 1/16。在解码部分，先与对应编码阶段得到的特征进行特征融合，再通过一个转置卷积来进行上采样，依次经过 D5-D2 四个解码层，特征图大小依次变为初始特征图的 1/8、1/4、1/2，最后在 D2 回到初始特征图的大小。之后对初始深度图、非局部邻域以及亲缘关系和初始深度图的置信度分别进行推算。以初始深度图的推算为例，将 D2 得到的特征图与 E2 得到的特征图融合，进行步长为 1 的卷积操作，得到的特征图再与 E1 初始融合图像特征进行融合，进行步长为 1 的操作，最后得到通道数为 1 的初始特征图。

### 2.1.1. 自校准卷积方法

自校准卷积方法[10]流程如图 4 所示。其中  $C$  为输入层的通道数， $H$  为输入层的高， $W$  为输入层的宽，先将输入特征图分成两个大小相同的子特征图，分别送进自校准卷积空间和简单卷积空间。在自校准卷积空间中，先对子特征图  $X_1$  进行平均池化下采样，通过 Conv1 卷积，再进行上采样，经过 Sigmoid 函数激活后通过矩阵乘法对 Conv2 卷积提取的特征进行校准，再通过 Conv3 卷积得到输出特征  $Y_1$ 。在简单卷积空间中，对子特征图  $X_2$  经过 Conv4 卷积提取得到特征  $Y_2$ 。最后对两个卷积空间输出特征  $Y_1$ 、 $Y_2$  进行拼接操作，得到最终输出特征  $Y$ ，具体计算公式见公式(1)。

$$Y = \text{Concat}\left\{\text{Conv3}\left(\text{Conv2}(X_1) \cdot \sigma\left(X_1 + \text{Up}\left(\text{Conv1}\left(\text{Down}(X_1)\right)\right)\right)\right); \text{Conv4}(X_2)\right\} \quad (1)$$

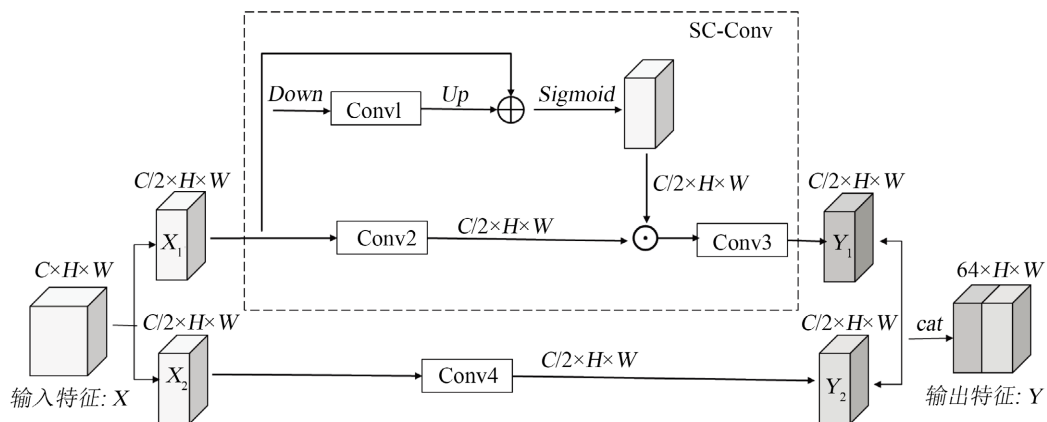


Figure 4. Algorithm flow of SC-Conv [10]  
图 4. 自校准卷积方法流程[10]

### 2.1.2. 多重注意力机制设计

多重注意力机制设计如图 5 所示，通道注意力模块和空间注意力模块采用顺序连接的方式加在自校准残差模块之间。给定一个中间特征图，沿着通道和空间两个维度依次推断出注意力权重，分别在通道维度和空间维度上学习特征，然后与原特征图相乘来对特征进行自适应调整。

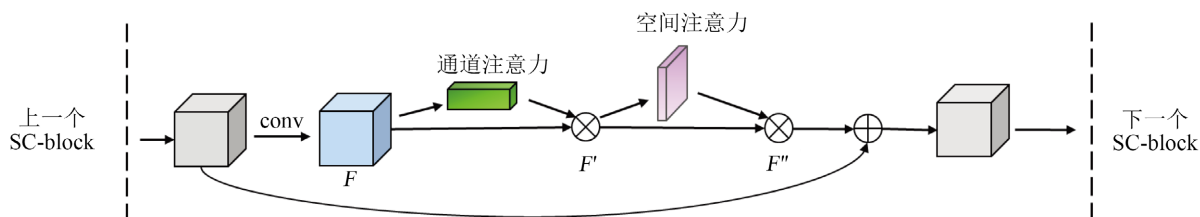


Figure 5. The design of multiple attention mechanism

图 5. 多重注意力机制设计

多重注意力机制具体实现流程如下。首先计算通道注意力，将输入特征分别进行空间维度的最大池化和平均池化得到两个  $1 \times 1 \times C$  大小的通道描述。再分别送入一个两层的神经网络。第一层神经元个数为  $C/r$ ，激活函数为 ReLU，第二层神经元个数为  $C$ 。这个两层的神经网络是共享的。将得到的两个特征相加后经过 Sigmoid 激活函数得到权重系数，再与输入特征相乘即可得到通道注意力。接着将通道注意力作为空间注意力的输入，分别进行通道维度的最大池化和平均池化，得到两个  $H \times W \times 1$  大小的通道描述。将这两个描述按照通道拼接后，经过  $7 \times 7$  的卷积层和 Sigmoid 激活函数，得到权重系数，再与输入的通道注意力相乘即可得到最终的特征。通道注意力和空间注意力[11]的方法流程如图 6 所示。

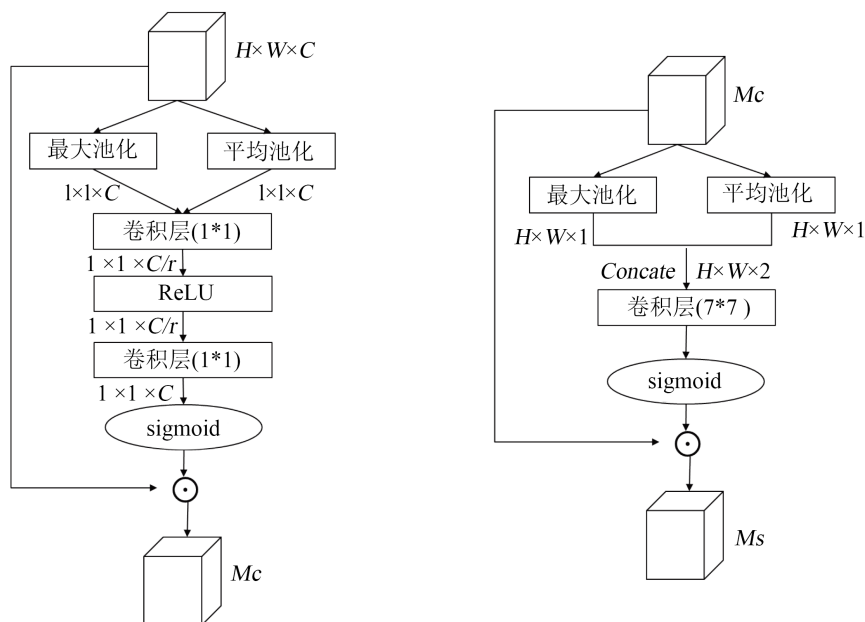


Figure 6. Channel attention (left) and spatial attention (right) [11]

图 6. 通道注意力(左)和空间注意力(右) [11]

## 2.2. 非局部特征空间传播网络

在现有的空间传播网络中，亲和力描述了像素之间的相关性，并为基于相似性的传播提供指导。在

这种情况下，空间中的每个像素都被同等对待，而不考虑其可靠性。然而，在深度完成任务中，不同的像素应该根据其可靠性进行加权。例如，来自不可靠像素(例如，噪声像素和深度边界上的像素)的信息不应传播到相邻像素，无论不可靠像素与相邻像素的亲合力如何。所以本文的非局部特征空间传播网络见图 7，首先将初始深度图的置信度纳入非局部邻域的亲合力归一化计算，建立可学习的亲合力归一化模块，以减少传播过程中来自不可靠深度的干扰。然后，将归一化后的亲合力与非局部邻域亲合力矩阵送进非局部空间传播模块，对预测的初始深度图进行迭代细化。非局部邻域定义见公式(2)。其中， $I$  和  $D$  分别时 RGB 图像和稀疏深度图， $(m,n)$  分别表示参考像素的坐标。

$$N_{m,n}^{NL} = \{x_{m+p,n+q} \mid (p,q) \in f_{\phi}(I,D,m,n), p,q \in \mathbb{R}\} \quad (2)$$

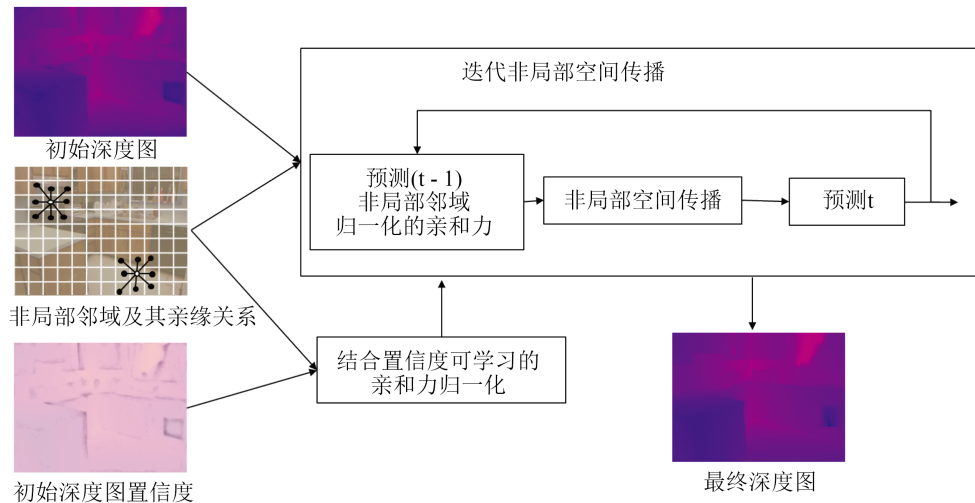


Figure 7. Non-Local spatial propagation network  
图 7. 非局部特征空间传播网络

### 2.3. 损失函数

为准确计算预测的稠密深度图，在训练过程中，本文采用绝对值误差  $l_1$  和均方误差  $l_2$  作为损失函数，计算预测的稠密深度图和实际深度真值之间的误差。损失函数的数学形式定义为公式(3)。

$$L_{recon}(D^{gt}, D^{pred}) = \frac{1}{n} \sum_{\nu} |D^{gt} - D^{pred}|^{\rho} \quad (3)$$

其中， $D^{gt}$  表示深度真值图， $D^{pred}$  表示预测的深度图， $\nu$  表示深度真值图中所有合法像素的集合， $n$  表示所有合法像素的个数。当  $\rho=1$  时， $l_1$  是绝对值误差；当  $\rho=2$  时， $l_2$  是绝对值误差。

## 3. 实验结果与分析

### 3.1. 实验环境与数据

本文实验在 Ubuntu16.04 操作系统下进行，计算机配置为 4 块 NVIDIA GeForce RTX 2080 Ti 显卡。训练和测试采 PyTorch1.6 深度学习框架进行。采用了  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  的 ADAM 优化器，初始学习率为 0.001。

为了证明本文提出的深度补全网络的有效性，在 KITTI 和 NYU Depth v2 两个公开数据集分别进行实验。其中，KITTI 为室外场景数据集，NYU Depth v2 为室内场景数据集。KITTI 数据集有超过 93,000

张原始深度图, 本文选择 33 个场景序列用于训练, 另外 28 个场景序列用于测试。忽略没有 LiDAR 投影的区域, 最后得到训练图像 89,200 张, 测试图像 1000 张。采用中心裁剪得到大小为  $1216 \times 240$  的图像。NYU Depth V2 数据集包含 464 个室内场景的 RGB 图像和深度图。从原始数据集中抽取了 45,205 万张图作为训练集, 654 张图作为测试集。每个图像都被缩小到  $320 \times 240$ , 然后采用  $304 \times 228$  的大小进行中心裁剪。

实验选择对模型进行了 25 次训练, 在前 10 次训练之后, 学习速率平均每 5 次训练衰减 0.2。实验将批量大小设置为 12, 非局部邻域大小设置为  $8 \times 8$ , 根据经验将非局部空间传播迭代次数设置为 18 次。

### 3.2. 精度评价指标

为了定量的评价深度补全精度, 实验选用深度补全任务统一的评价指标, 将网络模型预测深度图和深度真值图对比, 分别用均方根误差(RMSE), 平均绝对误差(MAE), 逆均方根误差(iRMSE), 逆平均绝对误差(iMAE), 平均绝对误差(REL)和平均  $\log_{10}$  误差三个指标来评价深度补全值的精度, 它们的计算公式如下。

$$\text{RMSE}(\text{mm}): \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i^{gt} - d_i^{pred})^2} \quad (4)$$

$$\text{MAE}(\text{mm}): \frac{1}{N} \sum_{i=1}^N |d_i^{gt} - d_i^{pred}| \quad (5)$$

$$\text{iRMSE}(1/\text{km}): \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{1}{d_i^{gt}} - \frac{1}{d_i^{pred}} \right)^2} \quad (6)$$

$$\text{iMAE}(1/\text{km}): \frac{1}{N} \sum_{i=1}^N \left| \frac{1}{d_i^{gt}} - \frac{1}{d_i^{pred}} \right| \quad (7)$$

$$\text{REL}: \frac{1}{N} \sum_{i=1}^N \left( \frac{|d_i^{gt} - d_i^{pred}|}{d_i^{gt}} \right) \quad (8)$$

$$\text{平均 } \log_{10} \text{ 误差}: \frac{1}{N} \sum_{i=1}^N |d_i^{gt} - d_i^{pred}| \quad (9)$$

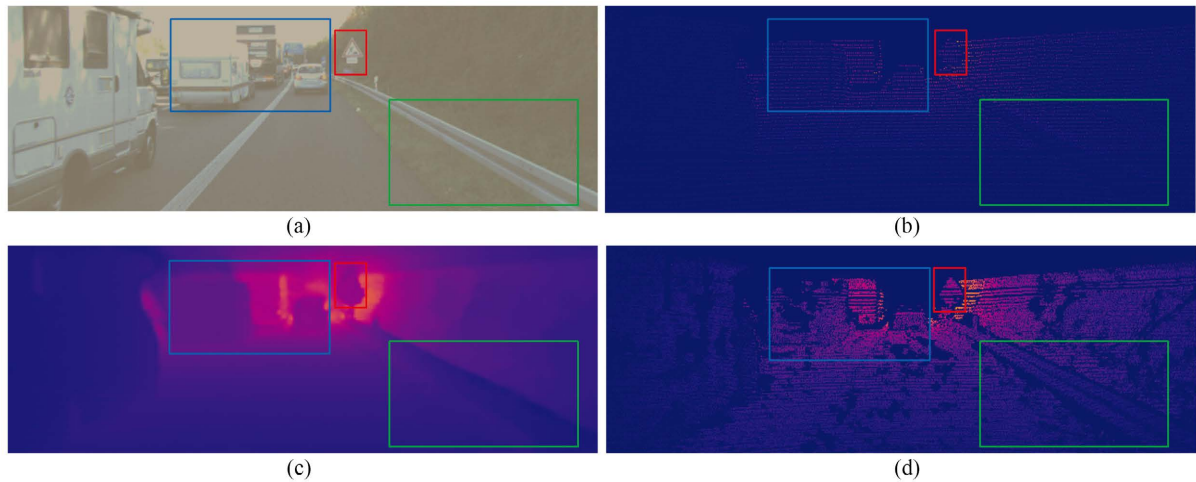
$$\delta_{\tau}: \max \left( \frac{d_i^{gt}}{d_i^{pred}}, \frac{d_i^{pred}}{d_i^{gt}} \right) < \tau \quad (10)$$

其中,  $d_i^{gt}$  和  $d_i^{pred}$  代表第  $i$  处像素的实际深度值和模型预测深度值。  $\delta$  为准确率, threshold 为准确率的阈值(分别取 1.25、1.252 和 1.253)。

### 3.3. 精度评价指标

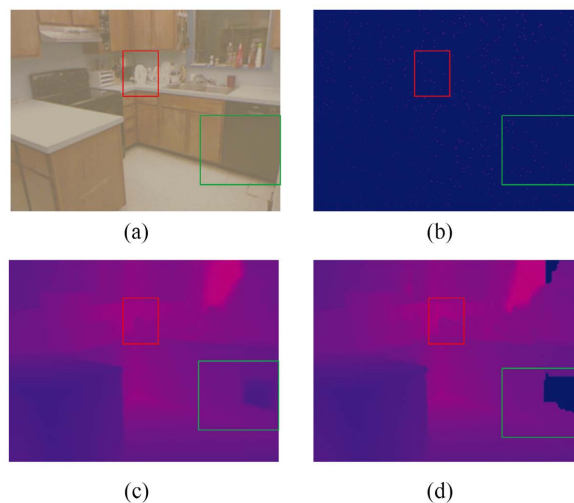
如图 8 选择了一张室外场景的高速区域数据做定性分析, 图中所示区域包括了高速场景中常见的物体: 车辆、指示牌、护栏等。图中(a)是原始彩色图像, (b)是稀疏深度图, (c)是实验结果图, (d)是深度真值图。发现, 与深度真值图相比, 实验结果能准确描绘出物体的形状、大小和相对位置。绿色框内是道路护栏, 是高速场景典型的线型物体, 通过本文的补全方法得到了准确的形状信息, 同时与周围环境有明显的边缘线。蓝色框里是常见的高速道路几辆车交错的场景, 通过本文方法可以准确地反映出车辆之间的前后、左右位置关系, 并且在彩色图像中彼此之间重合的边缘线, 在深度补全结果图中未发生深度混合, 车辆间的深度信息清晰可见。红色框中是高速场景中最重要的物体之一, 指示牌。发现经本文方法补全后, 指示牌的边缘信息更清晰, 但出现了过度外扩的问题, 可能是因为距离相机较远, 指示牌区

域初始点云数据过于稀疏，且在深度真值图上，点云数量也较为稀疏，导致在补全过程中，指示牌与附近深度信息产生了一定程度上的混合，但最终得到的结果基本也能表达指示牌形状轮廓。



**Figure 8.** Experiment results of KITTI  
**图 8.** KITTI 实验结果

如图 9 选择了一张室内场景的厨房区域数据做定性分析，图中所示区域既有明显且干净的物体边缘，也有微小且混乱的物体边缘，此外也有前景和后景的明显区分，适合做定性分析。图中(a)是原始彩色图像，(b)是稀疏深度图，(c)是实验结果图，(d)是深度真值图。发现，实验用的稀疏深度图深度信息密度非常低，原始深度图根本无法分辨物体的轮廓，通过实验算法得到的补全深度图可以较为清晰的分辨出较大物体的轮廓，比如桌子、柜子和洗手台等。对于桌子上摆放的小型物体，比如瓶子、盘子等，对比真实深度图，还是可以分辨出一些细节的轮廓。尤其是红色框中圈出的物体，实验结果图比真实深度值图更接近真实的轮廓形状。另外，在获取深度信息的时候由于物体表面反射，容易造成深度信息缺失，形成深度空洞，例如图中绿色框圈出的位置，由于玻璃反射导致该区域在深度真值图中没有深度信息。从实验结果图发现，实验算法可以在一定程度上补全由于反射造成的深度空洞。



**Figure 9.** Experiment results of NYU Depth v2  
**图 9.** NYU Depth v2 实验结果



### 3.4. 定量结果分析

**Table 1.** Experiment results of KITTI

**表 1.** KITTI 实验结果

方法	误差(越低越好)			
	RMSE (mm)	MAE	iRMSE	iMAE
CSPN [8]	1019.64	279.46	2.93	1.15
DeepLiDAR [12]	758.38	226.50	2.56	1.15
DepthNormal [13]	777.05	235.17	2.42	1.13
本文方法	741.89	221.19	2.34	1.10

**Table 2.** Experiment results of NYU Depth v2

**表 2.** NYU Depth v2 实验结果

方法	误差(越低越好)		准确率(越高越好)		
	RMSE	REL	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
CSPN [8]	0.117	0.016	99.2	99.9	100.0
DeepLiDAR [12]	0.115	0.022	99.3	99.9	100.0
DepthNormal [13]	0.112	0.018	99.5	99.9	100.0
本文方法	0.093	0.012	99.5	99.9	100.0

表 1 和表 2 分别显示了 KITTI 数据集 NYU Depth V2 数据集的定量评价。通过结果对比可以得知本文算法结果更优。与其余主流方法的结果相比，本文模型仍然保持了较高的准确率。在 KITTI 数据集上与 DeepLiDAR 的结果相比，均方根误差和平均绝对误差分别降低了 4.5% 和 6.0%。在 NYU Depth V2 数据集上与 DepthNormal 的结果相比，均方根误差和平均相对误差分别降低了 17.0% 和 33.3%。

## 4. 结语

针对深度补全存在的边界模糊和混合深度的问题，本文提出了一种新的融合序列影像和激光点云的深度补全方法。通过融合序列影像和激光点云，获取多模态数据特征。在特征提取阶段，本文融合多重注意力机制并引入自校准卷积结构，挖掘更丰富的彩色图像特征和深度信息特征，及两者之间的内在特征互补。本文还考虑全局特征约束，消除局部不相干特征干扰，提出非局部特征空间传播网络深度补全优化方法对初始预测的深度图进行迭代，在传播过程中专注于相关的非局部邻域，以提升深度补全的效果。通过实验验证，本文方法在室内场景和室外场景中均有效对深度信息缺失区域的补全，准确识别了物体的边缘轮廓和细节特征，正确描述了物体间的相对位置信息，较为完整地表达了整个场景的深度信息。与别的深度补全方法相比，本文方法降低了深度补全的误差，更接近真实深度信息，有效提高深度补全的质量。在未来的研究中，计划在实地采集的数据集上验证训练模型的泛化性。

## 致 谢

感谢国家自然科学基金面上基金和武汉大学知卓时空智能研究基金提供资助，感谢杨老师和陈老师对论文提供指导，感谢文章里引用文献的所有者。

## 基金项目

国家自然科学基金面上基金(面向电力线路安全监测的 UAV-MMS 多模态细粒度融合感知), 42071451。

武汉大学知卓时空智能研究基金。

## 参考文献

- [1] Cui, Z., Heng, L., Yeo, Y.C., Geiger, A., Pollefeys, M. and Sattler, T. (2019) Real-Time Dense Mapping for Self-Driving Vehicles Using Fisheye Cameras. *International Conference on Robotics and Automation (ICRA)*, Montreal, 20-24 May 2019, 6087-6093. <https://doi.org/10.1109/ICRA.2019.8793884>
- [2] Song, X., Dai, Y., Zhou, D., Liu, L., Li, W., Li, H. and Yang, R. (2020) Channel Attention Based Iterative Residual Learning for Depth Map Super-Resolution. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 5630-5639. <https://doi.org/10.1109/CVPR42600.2020.00567>
- [3] Armbrüster, C., Wolter, M., Kuhlen, T., Spijkers, W. and Fimm, B. (2008) Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *Cyberpsychology Behavior*, **11**, 9-15. <https://doi.org/10.1089/cpb.2007.9935>
- [4] Park, J., Joo, K., Hu, Z., Liu, C.-K. and So Kweon, I. (2020) Non-Local Spatial Propagation Network for Depth Completion. *European Conference on Computer Vision (ECCV)*, 120-136. [https://doi.org/10.1007/978-3-030-58601-0\\_8](https://doi.org/10.1007/978-3-030-58601-0_8)
- [5] Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T. and Geiger, A. (2017) Sparsity Invariant CNNs. *International Conference on 3D Vision (3DV)*, Qingdao, 10-12 October 2017, 11-20. <https://doi.org/10.1109/3DV.2017.00012>
- [6] Chodosh, N., Wang, C. and Lucey, S. (2018) Deep Convolutional Compressed Sensing for Lidar Depth Completion. *Asian Conference on Computer Vision*, 499-513. [https://doi.org/10.1007/978-3-030-20887-5\\_31](https://doi.org/10.1007/978-3-030-20887-5_31)
- [7] Ma, F. and Karaman, S. (2018) Sparse-to-Dense: Depth Prediction from Sparse Depth Samples and a Single Image. *International Conference on Robotics and Automation (ICRA)*, Brisbane, 21-25 May 2018, 4796-4803. <https://doi.org/10.1109/ICRA.2018.8460184>
- [8] Cheng, X., Wang, P. and Yang, R. (2018) Depth Estimation via Affinity Learned with Convolutional Spatial Propagation Network. *European Conference on Computer Vision (ECCV)*, 103-119. [https://doi.org/10.1007/978-3-030-01270-0\\_7](https://doi.org/10.1007/978-3-030-01270-0_7)
- [9] Liu, S., De Mello, S., Gu, J., Zhong, G., Yang, M.-H. and Kautz, J. (2017) Learning Affinity via Spatial Propagation Networks. *Advances in Neural Information Processing Systems*, 30.
- [10] Liu, J.-J., Hou, Q., Cheng, M.-M., Wang, C. and Feng, J. (2020) Improving Convolutional Networks with Self-Calibrated Convolutions. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 10093-10102. <https://doi.org/10.1109/CVPR42600.2020.01011>
- [11] Woo, S., Park, J., Lee, J.-Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *European Conference on Computer Vision (ECCV)*, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [12] Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B. and Pollefeys, M. (2019) DeepLiDAR: Deep Surface Normal Guided Depth Prediction for Outdoor Scene from Sparse Lidar Data and Single Color Image. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 3313-3322. <https://doi.org/10.1109/CVPR.2019.00343>
- [13] Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H. and Li, H. (2019) Depth Completion from Sparse Lidar Data with Depth-Normal Constraints. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 2811-2820. <https://doi.org/10.1109/ICCV.2019.00290>