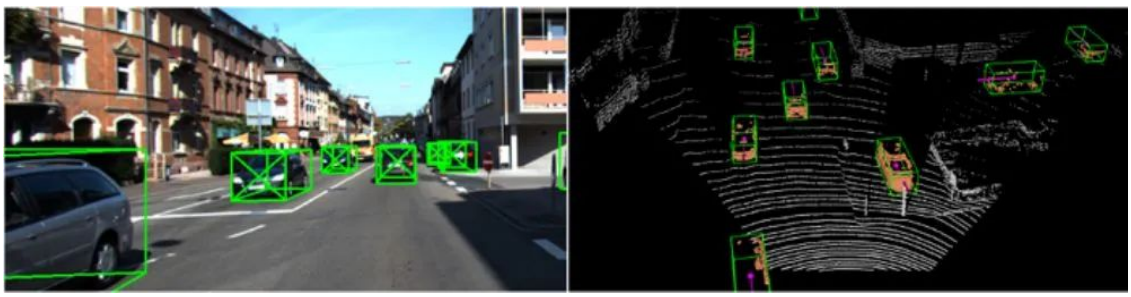
# 浅析自动驾驶传感器融合：激光雷达+摄像头

计算机视觉life 2022-02-05 11:24



## / 导读 /



自动驾驶感知技术所采用的传感器主要包括摄像头，激光雷达和毫米波雷达。这些传感器各有优缺点，也互为补充，**因此如何高效的融合多传感器数据，也就自然的成为了感知算法研究的热点之一。**本篇文章介绍如何在感知任务中融合激光雷达和摄像头，重点是目前主流的基于深度学习的融合算法。

**摄像头产生的数据是2D图像，对于物体的形状和类别的感知精度较高。**深度学习技术的成功起源于计算机视觉任务，很多成功的算法也是基于对图像数据的处理，因此目前基于图像的感知技术已经相对成熟。图像数据的缺点在于受外界光照条件的影响较大，很难适用于所有的天气条件。对于单目系统来说，获取场景和物体的深度（距离）信息也比较困难。双目系统可以解决深度信息获取的问题，但是计算量很大。激光雷达在一定程度上弥补了摄像头的缺点，可以精确的感知物体的距离，但是限制在于成本较高，车规要求难以满足，因此在量产方面比较困难。同时，激光雷达生成的3D点云比较稀疏（比如垂直扫描线只有64或128）。对于远距离物体或者小物体来说，反射点的数量会非常少。

如下图所示，图像数据和点云存在着巨大的差别。首先是视角不同，图像数据是真实世界通过透视投影得到的二维表示，而三维点云则包含了真实世界欧式坐标系中的三维信息，可以投影到多种视图。其次是数据结构不同，图像数据是规则的，有序的，稠密的，而点云数据是不规则的，无序的，稀疏的。在空间分辨率方面，图像数据也比点云数据高很多。



图片来源于参考文献[1]

**自动驾驶感知系统中有两个典型的任务：物体检测和语义分割。**深度学习技术的兴起首先来自视觉领域，基于图像数据的物体检测和语义分割已经被广泛和充分的研究，也有很多非常全面的综述文章，这里就不赘述了。另一方面，随着车载激光雷达的不断普及以及一些大规模数据库的发布，点云数据处理的研究这几年来发展也非常迅速。本专栏之前的两篇文章分别介绍了点云物体检测和语义分割的发展情况，感兴趣的朋友可以参考。下面以物体检测任务为主来介绍不同的融合方法。语义分割的融合方法可以由物体检测扩展得到，就不做单独介绍了。

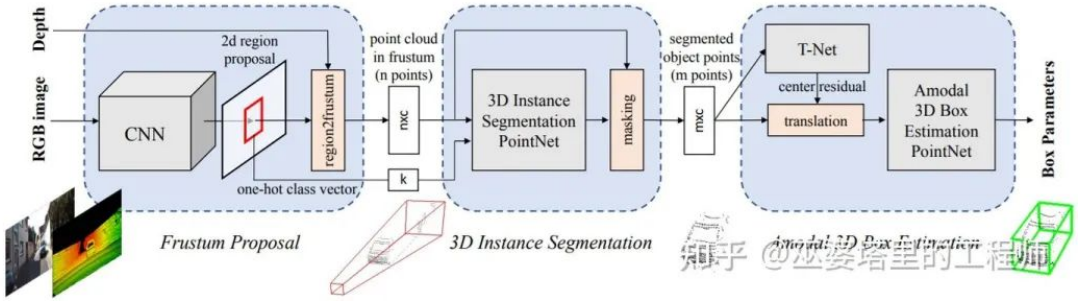
## 不同的融合策略

**物体检测的策略分为：决策层融合，决策+特征层融合，以及特征层融合。**在决策层融合中，图像和点云分别得到物体检测结果（BoundingBox），转换到统一坐标系后再进行合并。这种策略中用到的大都是一些传统的方法，比如IoU计算，卡尔曼滤波等，与深度学习关系不大，本文就不做介绍了。下面重点来讲讲后两种融合策略。

### 决策+特征层融合

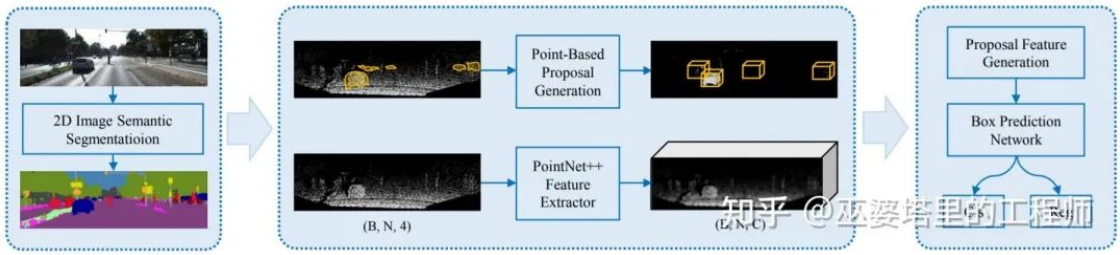
这种策略的主要思路是将先通过一种数据生成物体的候选框（Proposal）。如果采用图像数据，那么生成的就是2D候选框，如果采用点云数据，那么生成的就是3D候选框。然后将候选框与另外一种数据相结合来生成最终的物体检测结果（也可以再重复利用生成候选框的数据）。这个结合的过程就是将候选框和数据统一到相同的坐标系下，可以是3D点云坐标（比如F-PointNet），也可以是2D图像坐标（比如IPOD）。

F-PointNet[2]由图像数据生成2D物体候选框，然后将这些候选框投影到3D空间。每个2D候选框在3D空间对应一个视椎体（Frustum），并将落到视椎体中所有点合并起来作为该候选框的特征。视椎体中的点可能来自前景的遮挡物体或者背景物体，所以需要进行3D实例分割来去除这些干扰，只保留物体上的点，用来进行后续的物体框估计（类似PointNet中的处理方式）。这种基于视椎的方法，其缺点在于每个视椎中只能处理一个要检测的物体，这对于拥挤的场景和小目标（比如行人）来说是不能满足要求的。



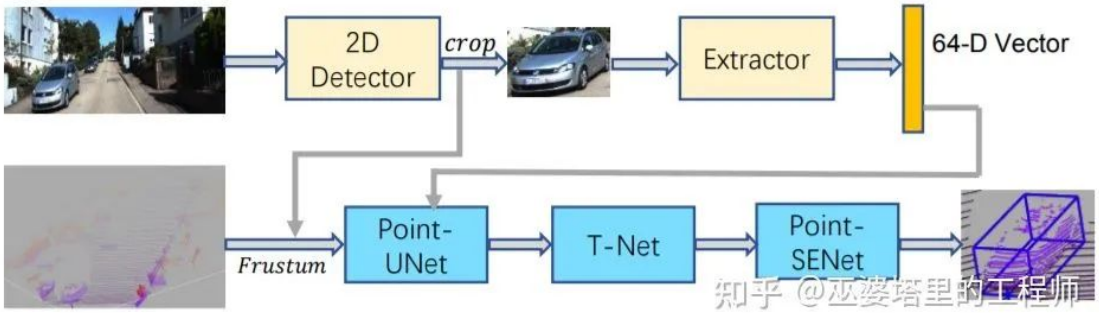

F-PointNet网络结构图

针对视椎的上述问题，IPOD[3]提出采用2D语义分割来替换2D物体检测。首先，图像上的语义分割结果被用来去除点云中的背景点，这是通过将点云投影到2D图像空间来完成的。接下来，在每个前景点处生成候选物体框，并采用NMS去除重叠的候选框，最后每帧点云大约保留500个候选框。同时，PointNet++网格被用来进行点特征提取。有了候选框和点特征，最后一步采用一个小规模的PointNet++来预测类别和准确的物体框（当然这里也可以用别的网络，比如MLP）。IPOD在语义分割的基础上生成了稠密的候选物体框，因此在含有大量物体和互相遮挡的场景中效果比较好。



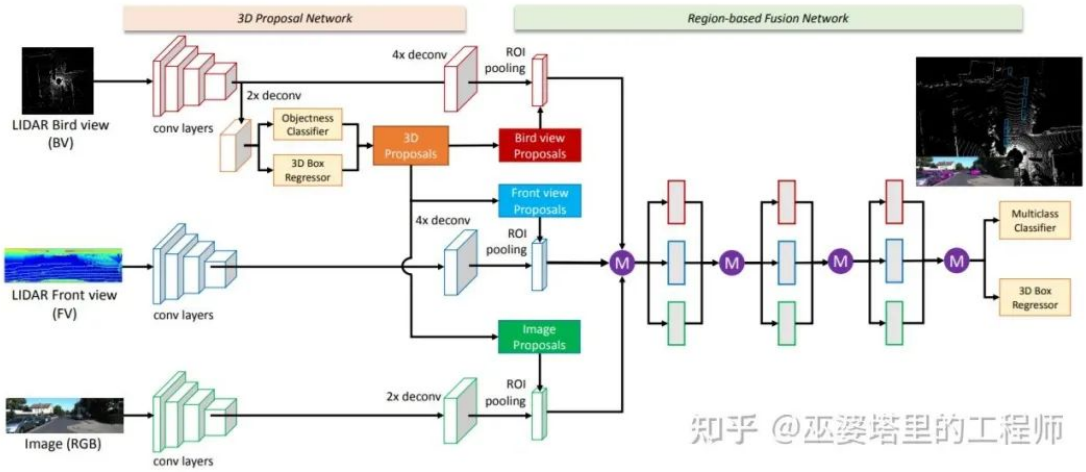

IPOD网络结构图

上面两个方法分别通过2D图像上的物体检测和语义分割结果来生成候选框，然后只在点云数据上进行后续的处理。SIFRNet[4]提出在视椎体上融合点云和图像特征，以增强视椎体所包含的信息量，用来进一步提高物体框预测的质量。



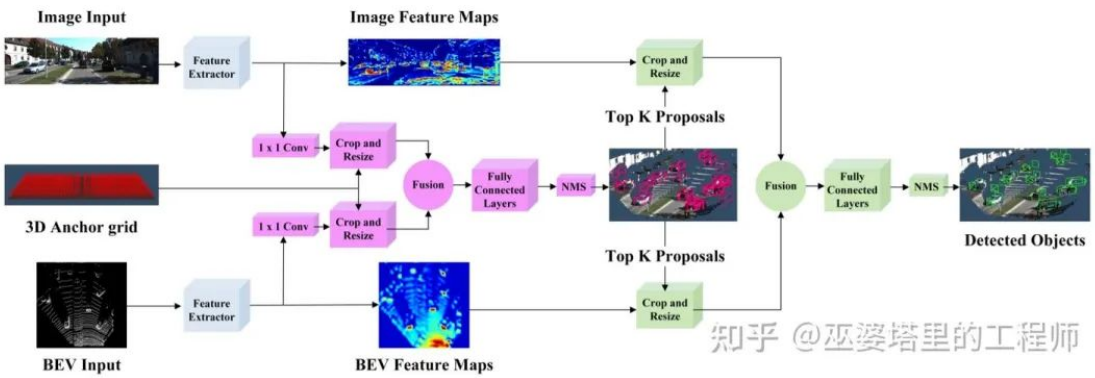

SIFRNet网络结构图

近年来，随着3D物体检测技术的快速发展，物体候选框的选取也从逐渐从2D向3D转变。MV3D[5]是基于3D候选框的代表性工作。首先，它将3D点云映射到BEV视图，并基于此视图生成3D物体候选框。然后，将这些3D候选框映射到点云的前视图以及图像视图，并将相应的特征进行融合。特征融合是以候选框为基础，并通过ROI pooling来完成的。




MV3D网络结构图

AVOD[6]的思路也是在3D候选框的基础上融合图像和点云特征。但是原始候选框的生成并不是通过点云处理得到，而是通过先验知识在BEV视图下均匀采样生成的（间隔0.5米，大小为各个物体类的均值）。点云数据用来辅助去除空的候选框，这样最终每帧数据会产生8万到10万个候选框。这些候选框通过融合的图像和点云特征进行进一步筛选后，作为最终的候选再送入第二阶段的检测器。因此，也可以认为AVOD的候选框是同时在图像和点云上得到的。

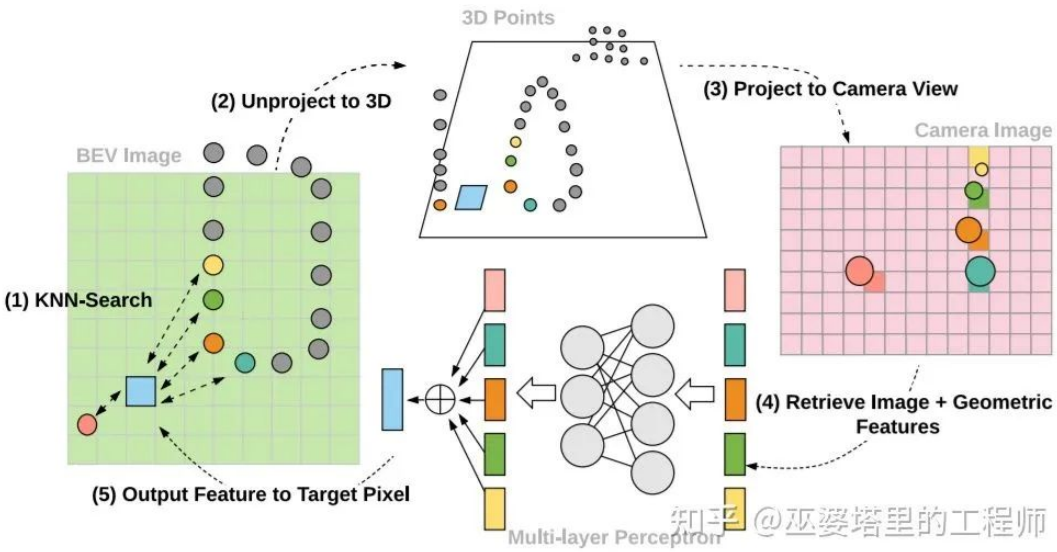



AVOD网络结构图

## 特征层融合

决策+特征层融合的特点是以物体候选框为中心来融合不同的特征，融合的过程中一般会用到ROI pooling（比如双线性插值），而这个操作会导致空间细节特征的丢失。另外一种思路是特征层融合，也就是直接融合多种特征。比如说将点云映射到图像空间，作为带有深度信息的额外通道与图像的RGB通道进行合并。这种思路简单直接，对于2D物体检测来说效果不错。但是融合的过程丢失了很多3D空间信息，因此对于3D物体检测来说效果并不好。由于3D物体检测领域的迅速发展，特征层融合也更倾向于在3D坐标下完成，这样可以为3D物体检测提供更多信息。

ContFuse[7]采用连续卷积（Continuous Convolution）来融合点云和图像特征。融合过程在BEV视图下完成。对于BEV上的一个像素（网格），首先在点云数据中找到其K个最邻近的点，然后将这些3D空间中的点映射到图像空间，以此得到每个点的图像特征。同时，每个点的几何特征则是该点到相应BEV像素的XY偏移量。将图像特征和几何特征合并作为点特征，然后按照连续卷积的做法对其进行加权求和（权重依赖于XY偏移量），以得到相应BEV像素处的特征值。对BEV的每个像素进行类似处理，就得到了一个BEV特征图。这样就完成了图像特征到BEV视图的转换，之后就可以很方便的与来自点云的BEV特征进行融合。ContFuse中在多个空间分辨率下进行了上述的特征融合，以提高对不同大小物体的检测能力。

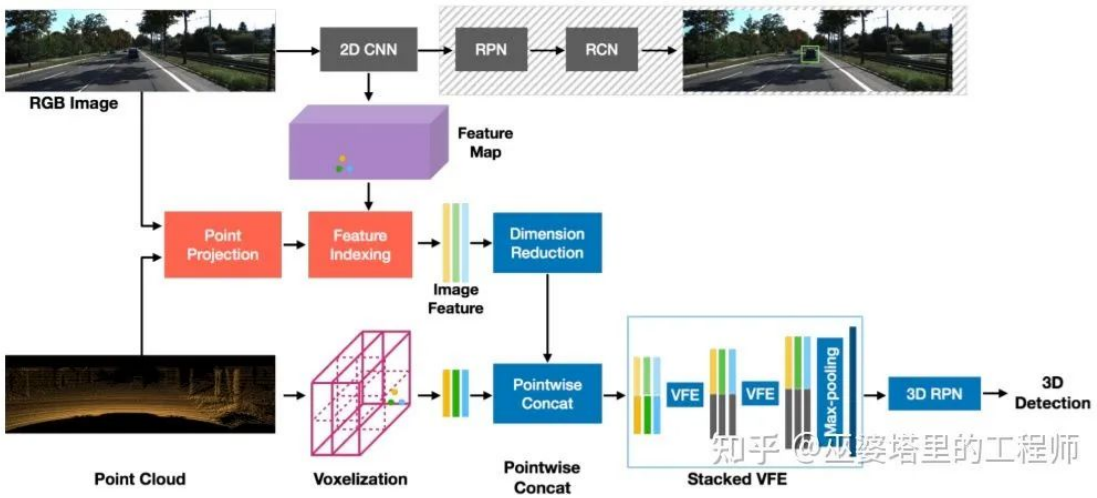



利用连续卷积将图像特征投影到BEV视图

PointPainting[8]把点云投影到图像语义分割的结果中，这与IPOD中的做法类似。但是，PointPainting没有利用语义分割的结果来分离前景点，而是直接将语义分割的信息附加到点云上。这样做的好处是，融合之后的数据还是点云（但是具有更为丰富的语义信息），可以采用任何点云物体检测网络来处理，比如PointRCNN，VoxelNet，PointPillar等等。

PointPainting的融合流程图

PointPainting中附加给点云的是2D图像的语义信息，这已经是高度抽象之后的信息，而原始的图像特征则被丢弃了。从融合的角度来看，底层特征的融合可以更大程度的保留信息，利用不同特征之间的互补性，理论上说也就更有可能提升融合的效果。MVX-Net[9]利用一个实现训练好的2D卷积网络来提取图像特征，然后通过点云和图像之间的映射关系将图像特征附加到每个点上。之后再采用VoxelNet来处理融合后的点特征。除了这种点融合策略，MVX-Net还提出了在voxel层次上融合，其主要的不同就在于将voxel而不是point投影到图像空间，因此图像特征是被附加在voxel之上。从实验结果来看，point融合比voxel融合结果略好，这也进一步说明了较低的融合层次可能会带来更好的效果。




MVX-Net中的Point融合方法

语义分割任务中的融合一般都是在特征层上进行，之前介绍的特征融合方法理论上来说可以用来进行语义分割。比如说，ContFuse在BEV网格上融合了图像和点云特征，这个特征就可以用来进行网格级别的语义分割，而PointPainting将图像特征附加到点云上，后续可以采用任何基于点云语义分割的算法来对每个点进行语义分类，甚至也可以进行实例分割和全景分割。

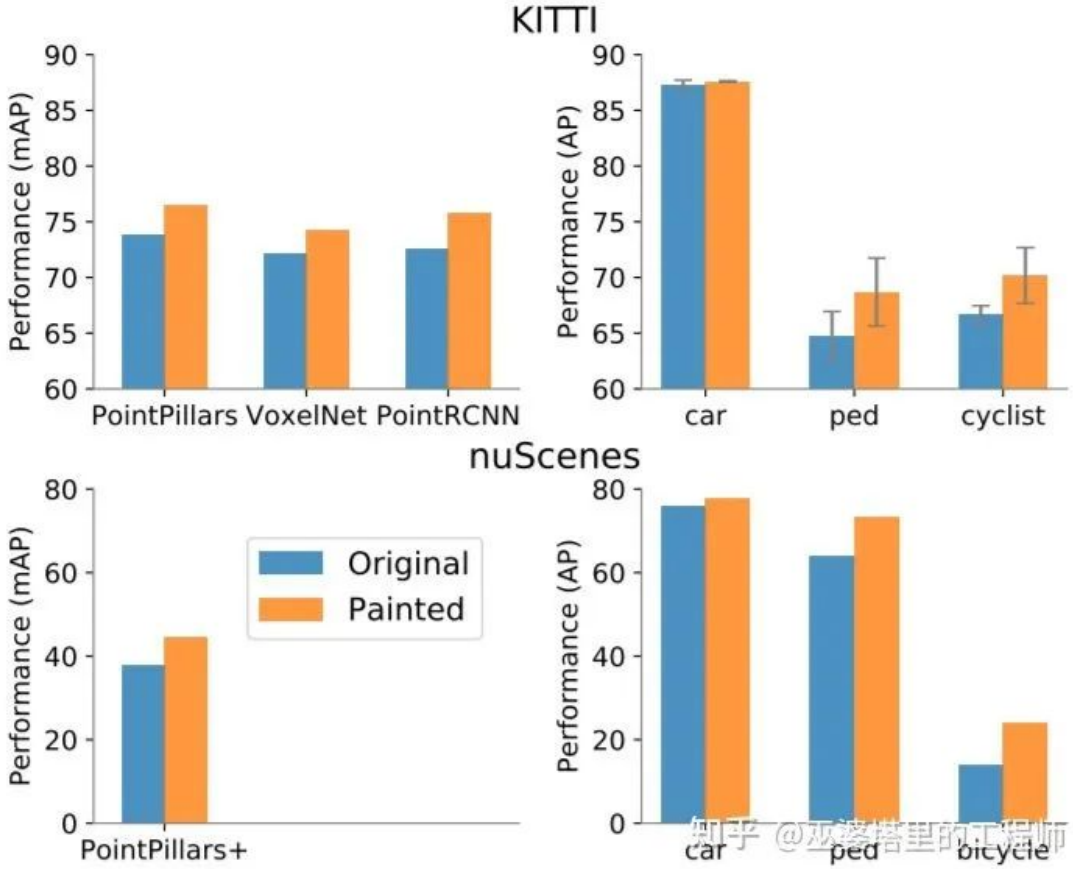
## 2

## 结果对比

这里我们来总结和定量的对比一下前面介绍的各种融合方法。准确度指标采用采用KITTI数据库上3D车辆检测中等难度的AP（70% IoU），速度指标采用FPS（运行的硬件不同，因此不具备完全的可比性）。下表中融合方法一栏中的D+F表示决策+特征层融合，之后的2D/3D表示是在2D图像还是3D点云上提取物体候选框。F表示特征层融合，之后的BEV和Point表示融合的位置。总体来说，特征层融合的效果较好，基于Point的融合也优于基于BEV的融合。

| 算法 | 融合方法 | 准确度（AP） | 速度（FPS） |
| --- | --- | --- | --- |
| F-PointNet | D+F, 2D | 69.79 | 5.9 |
| IPOD | D+F, 2D | 72.57 | 5.0 |
| SIFRNet | D+F, 2D | 72.05 | - |
| MV3D | D+F, 3D | 63.63 | 2.8 |
| AVOD | D+F, 2D+3D | 71.76 | 12.5 |
| ContFuse | F, BEV | 68.78 | 16.7 |
| PointPainting | F, Point | 71.70 | 2.5 |
| MVX-Net | F, Point | 77.43 | - |

作为对比，只基于点云数据的VoxelNet其AP为64.17，MVX-Net将图像特征附加到点云上之后再采用VoxelNet就可以将AP提升到77.43，提升的幅度还是非常可观的。PointPainting中的对比实验也展示了类似的提升。下图是分别在KITTI和NuScenes上进行的对比实验。PointPillar，VoxelNet，和PointRCNN这三个点云物体检测的常用方法在结合了图像特征后都有了很大幅度的提升。尤其是对于行人和骑车的人这两个类来说，提升的幅度更大，这也证明了分辨率较高的图像特征对小目标的检测有很大的帮助。






*参考文献*

*[1] Cui et.al., Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review, 2020.*

*[2] Qi et.al., Frustum Pointnets for 3d Object Detection from RGB-D Data, 2018.*

*[3] Yang et.al., IPOD: Intensive Point-based Object Detector for Point Cloud, 2018.*

*[4] Zhao et.al., 3D Object Detection Using Scale Invariant and Feature Re-weighting Networks, 2019.*

*[5] Chen et.al., Multi-View 3D Object Detection Network for Autonomous Driving, 2016.*

*[6] Ku et.al., Joint 3D Proposal Generation and Object Detection from View Aggregation, 2017.*

*[7] Liang et.al., Deep Continuous Fusion for Multi-Sensor 3D Object Detection, 2018.*

*[8] Vora et.al., PointPainting: Sequential Fusion for 3D Object Detection, 2019.*

*[9] Sindagi et.al., MVX-Net: Multimodal VoxelNet for 3D Object Detection, 2019.*

8、图像三维重建课程：视觉几何三维重建教程（第2期）：稠密重建，曲面重建，点云融合，纹理贴图

9、重磅来袭！基于LiDAR的多传感器融合SLAM 系列教程：LOAM、LeGO-LOAM、LIO-SAM

10、系统全面的相机标定课程：单目/鱼眼/双目/阵列 相机标定：原理与实战

扫描学习SLAM、三维重建

全国最棒的SLAM、三维视觉学习社区↓


## 技术交流微信群

欢迎加入公众号读者群一起和同行交流，目前有**SLAM、三维视觉、传感器、自动驾驶、计算摄影**、检测、分割、识别、**医学影像、GAN、算法竞赛**等微信群，请添加微信号 chichui502 或扫描下方加群，备注：”名字/昵称+学校/公司+研究方向“。**请按照格式备注，否则不予通过**。添加成功后会根据研究方向邀请进入相关微信群。请勿在群内发送**广告**，否则会请出群，谢谢理解~

投稿、合作也欢迎联系：simiter@126.com

扫描关注视频号，看最新技术落地及开源方案视频秀 ↓

扫一扫二维码，关注我的视频号



Read more

People who liked this content also liked

上线十年，影响一代ML工程师，吴恩达经典《机器学习》课程迎来重磅更新

机器之心

当深度学习邂逅物理学：机器之心X中科院自动化所技术论坛第一期开讲

机器之心

杜克大学陈怡然：高效人工智能系统的软硬件协同设计

机器之心