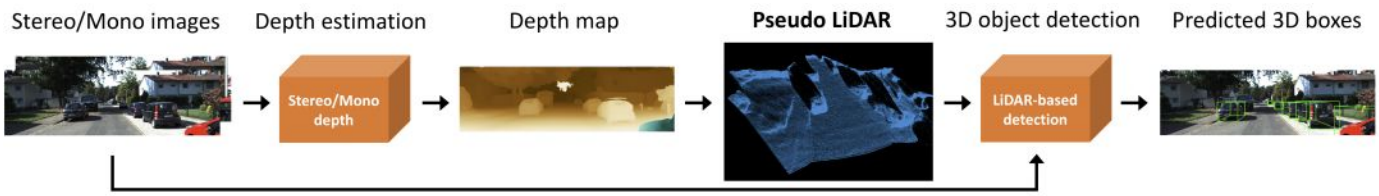
近日，CVPR 2019发布接收论文ID列表，共计1300篇论文被接收，接受率为25.2%。本文整理了无人驾驶方面的优秀论文，一起来看看该领域最前沿的研究课题。

CVPR 是首屈一指的年度计算机视觉盛会，在机器学习领域享有盛名。今年的 CVPR 将于 6 月 16 日-20 日于美国加州的长滩市举行。

CVPR 作为计算机视觉领域的顶级学术会议，今年共收到了 5165 篇有效提交论文，比去年 CVPR2018 增加了 56%。不久之前，CVPR 2019 官网放出了最终的论文接收结果。据统计，本届大会共接收了 1300 论文，接收率接近 25.2%。本文智车科技整理了本届会议上与无人驾驶相关的优秀论文及项目，并附有下载链接。

1.

题目：Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving（伪激光雷达）
作者：Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, Kilian Q. Weinberger
论文链接：arxiv.org/abs/1812.0717...
项目链接：mileyan.github.io/pseud...
代码链接：github.com/mileyan/pseu...



摘要：3D物体检测是自动驾驶中的基本任务。如果从精确但昂贵的LiDAR技术获得3D输入数据，则最近的技术具有高度准确的检测率。迄今为止，基于较便宜的单目或立体图像数据的方法导致精度显着降低 - 这种差距通常归因于基于图像的深度估计不良。然而，在本文中，我们认为数据表示（而不是其质量）占据了差异的大部分。考虑到卷积神经网络的内部工作原理，我们建议将基于图像的深度图转换为伪LiDAR表示 - 基本上模仿LiDAR信号。通过这种表示，我们可以应用不同的现有基于LiDAR的检测算法。在流行的KITTI基准测试中，我们的方法在现有的基于图像的性能方面取得了令人印象深刻的改进 - 提高了30米范围内物体的检测精度，从先前的22%到现在的前所未有的74% 。在提交时，我们的算法在KITTI 3D对象检测排行榜上保持最高条目，用于基于立体图像的方法。

2.

题目：ApolloCar3D: A Large 3D Car Instance Understanding Benchmark for Autonomous Driving（数据集）
作者：Xibin Song, Peng Wang, Dingfu Zhou, Rui Zhu, Chenye Guan, Yuchao Dai, Hao Su, Hongdong Li, Ruigang Yang
论文链接：arxiv.org/abs/1811.1222...

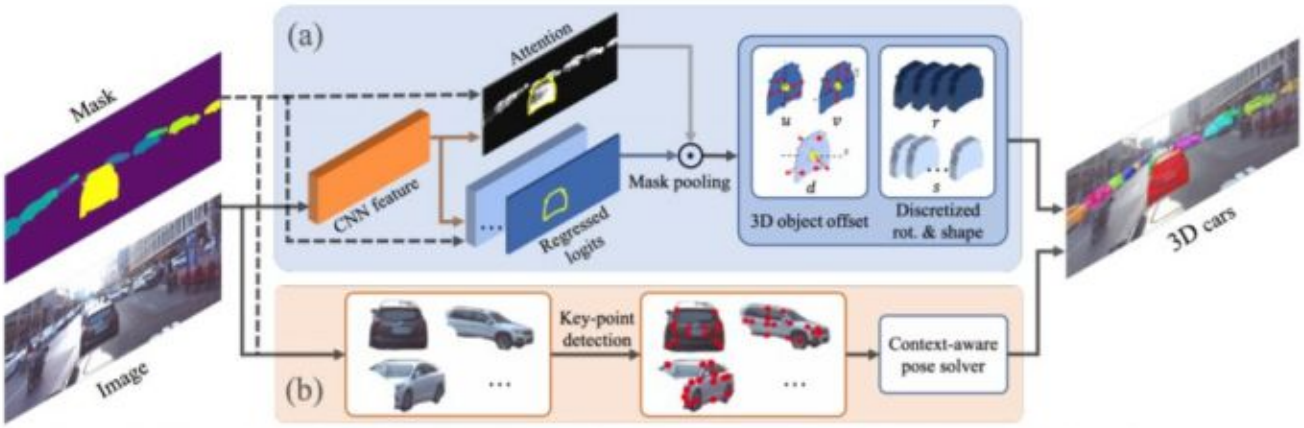


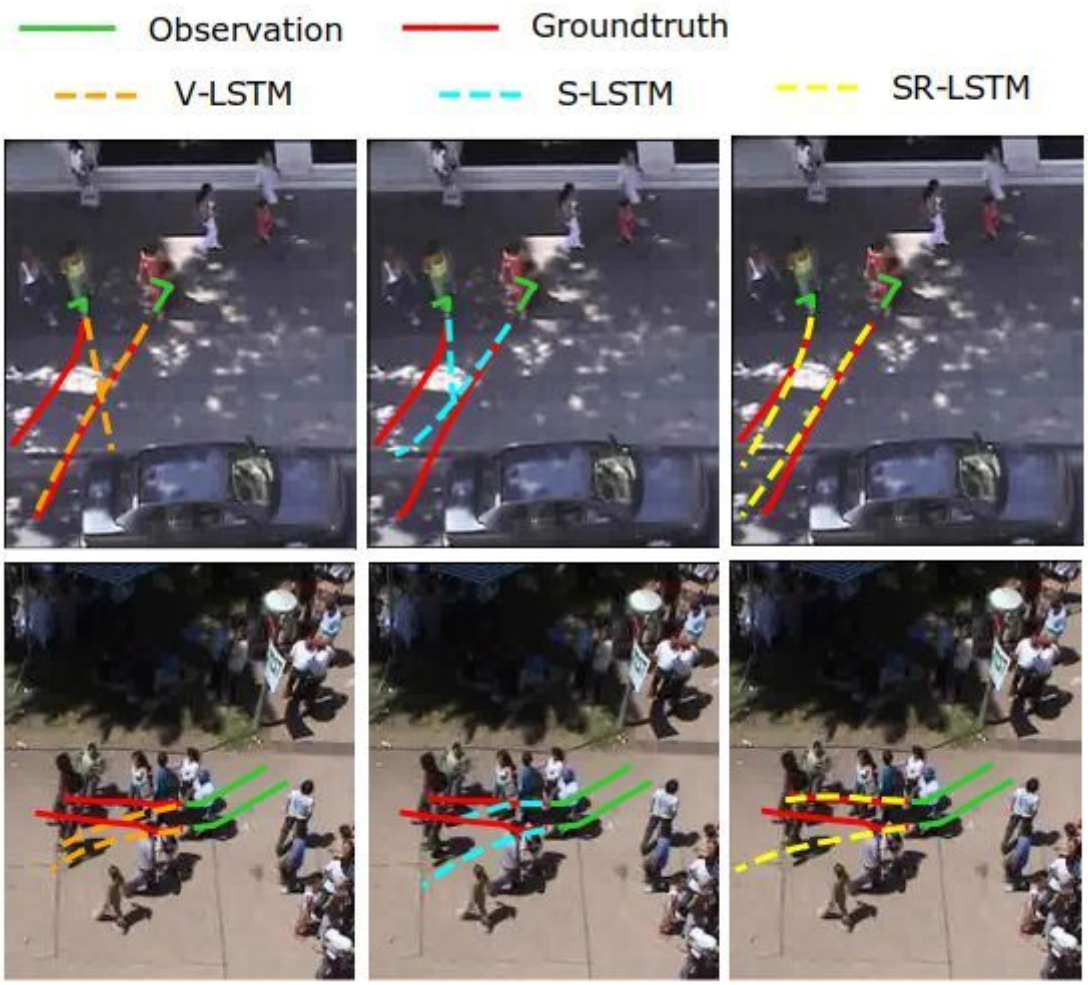
Figure 5: Training pipeline for 3D car understanding. Upper (a): direct approach. Bottom (b): key point based approach.

摘要：自动驾驶引起了业界和学术界的极大关注。一个重要的任务是估计道路上移动或停放的车辆的3D特性（例如，翻译，旋转和形状）。这项任务虽然至关重要，但在计算机视觉领域仍未得到充分研究 - 部分原因在于缺乏适合自动驾驶研究的大规模和完全注释的3D汽车数据库。在本文

中，我们贡献了第一个适合3D汽车实例理解的大型数据库 - ApolloCar3D。该数据集包含5,277个驾驶图像和超过60K的汽车实例，其中每辆汽车都配备了具有绝对模型尺寸和语义标记关键点的行业级3D CAD模型。该数据集比PASCAL3D +和KITTI（现有技术水平）大20倍以上。为了在3D中实现高效标记，我们通过考虑单个实例的2D-3D关键点对应关系和多个实例之间的3D关系来构建管道。配备这样的数据集，我们使用最先进的深度卷积神经网络构建各种基线算法。具体来说，我们首先使用预先训练的Mask R-CNN对每辆车进行分段，然后基于可变形的3D汽车模型，使用或不使用语义关键点，对其3D姿势和形状进行回归。研究表明，使用关键点可以显着提高拟合性能。最后，我们开发了一个新的3D度量，共同考虑3D姿势和3D形状，允许进行全面的评估和消融研究。

3.

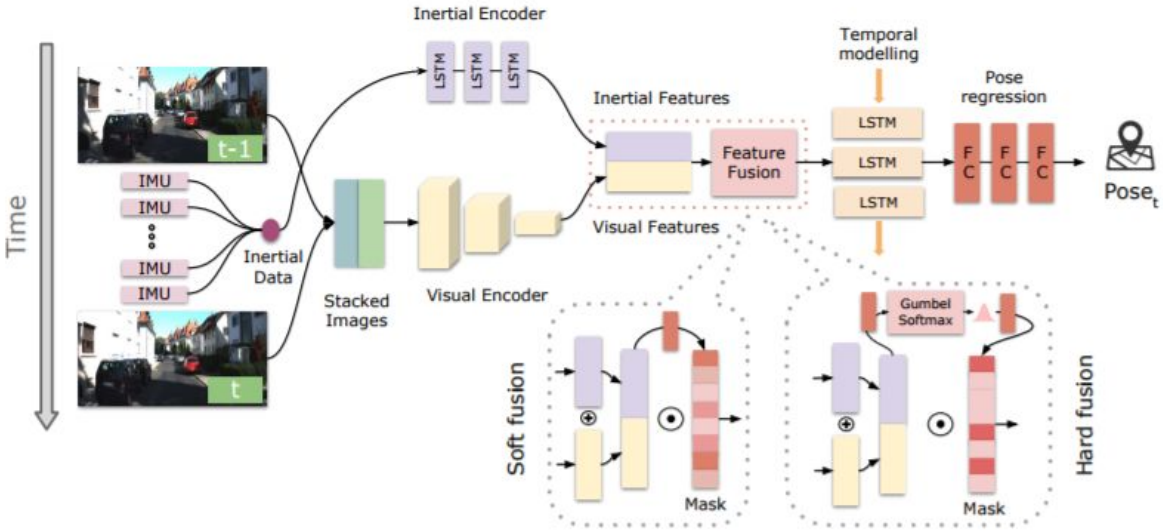
题目：SR-LSTM: State Refinement for LSTM towards Pedestrian Trajectory Prediction（行人预测）
作者：Pu Zhang, Wanli Ouyang, Pengfei Zhang, Jianru Xue, Nanning Zheng
论文链接：https：//arxiv.org/abs/1903.0279...



摘要：在人群场景中，行人的可靠轨迹预测需要深刻理解他们的社交行为。大量研究已经很好地研究了这些行为，而规则很难充分表达。最近基于LSTM网络的研究表明，学习社交行为的能力很强。然而，这些方法中的许多方法依赖于先前的相邻隐藏状态，但忽略了邻居的重要当前意图。为了解决这个问题，我们提出了一个用于LSTM网络（SR-LSTM）的数据驱动状态细化模块，它激活了对邻居当前意图的利用，并共同和迭代地改进了人群中所有参与者的当前状态。通过消息传递机制。为了有效地提取邻居的社会影响，我们进一步介绍了一种社会意识信息选择机制，包括逐元素运动门和行人注意力，以便从邻近的行人中选择有用的信息。两个公共数据集（即ETH和UCY）的实验结果证明了我们提出的SR-LSTM的有效性，并且我们实现了最先进的结果。

4.

题目：Selective Sensor Fusion for Neural Visual-Inertial Odometry（视觉惯性测距）
作者：Changhao Chen, Stefano Rosa, Yishu Miao, Chris Xiaoxuan Lu, Wei Wu, Andrew Markham, Niki Trigoni
论文链接：arxiv.org/abs/1903.0153...

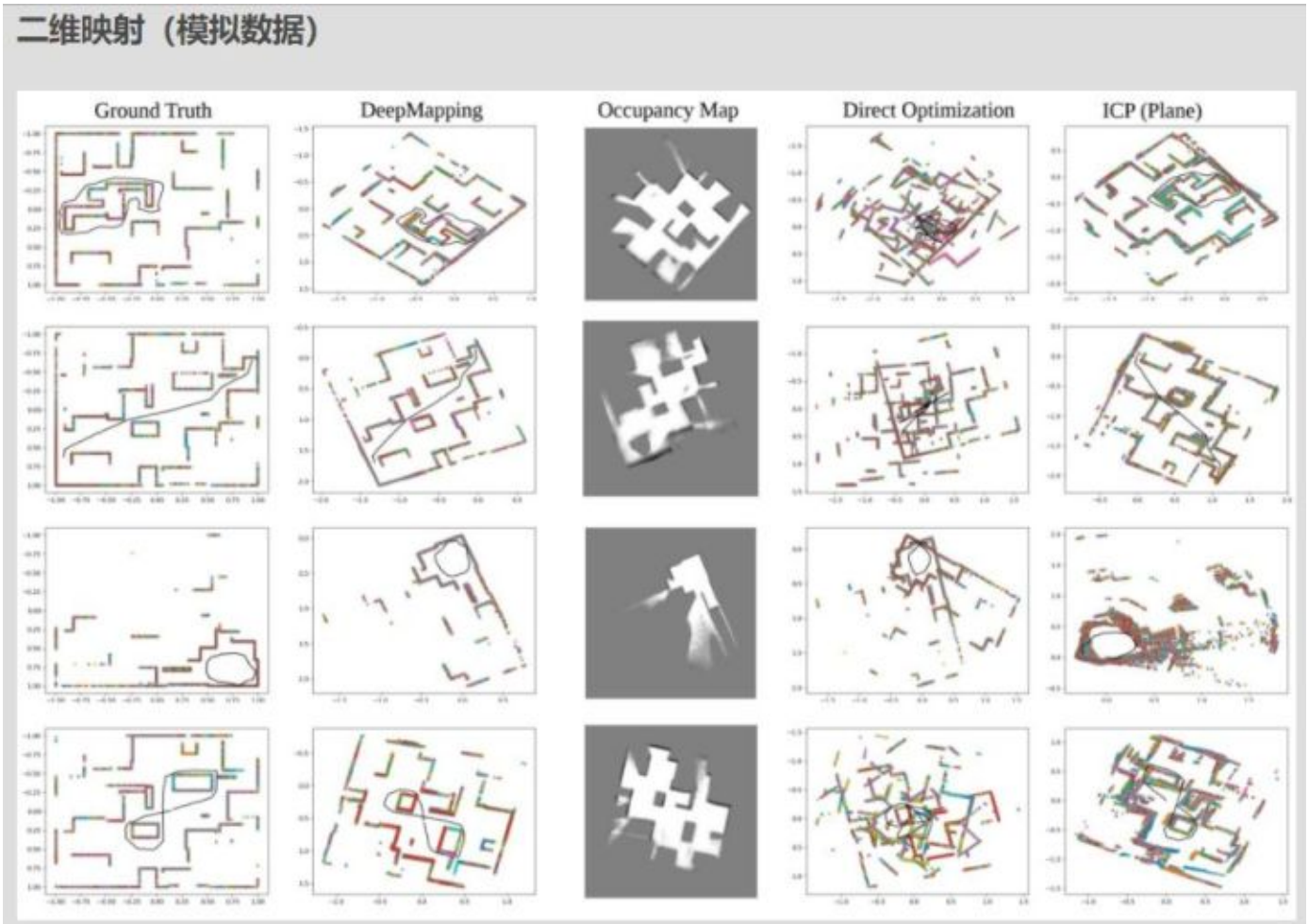


摘要：视觉惯性测距（VIO）的深度学习方法已被证明是成功的，但他们很少专注于结合稳健的融合策略来处理不完美的输入感觉数据。我们提出了一种新颖的端对端选择性传感器融合框架，用于单眼VIO，融合单眼图像和惯性测量，以估计轨迹，同时提高对实际问题的鲁棒性，如丢失和损坏的数据或不良的传感器同步。特别地，我们提出了两种基于不同掩蔽策略的融合模态：确定软性融合和随机硬融合，并与先前提出的直接融合基线进行比较。在测试期间，网络能够选择性地处理可用传感器模态的特征并且产生大规模的轨迹。我们对三种公共自动驾驶，微型飞行器（MAV）和手持VIO数据集的性能进行了全面调查。结果证明了融合策略的有效性，与直接融合相比，其提供了更好的性能，特别是在存在损坏的数据的情况下。此外，我们通过可视化不同场景中的掩蔽层和不同的数据损坏来研究融合网络的可解释性，揭示融合网络与不完美的传感输入数据之间的有

关性。

5.

题目：DeepMapping: Unsupervised Map Estimation From Multiple Point Clouds
作者：Li Ding, Chen Feng
论文链接：arxiv.org/abs/1811.1139...
项目链接：ai4ce.github.io/DeepMap...



摘要：我们提出DeepMapping，一种新颖的注册框架，使用深度神经网络（DNN）作为辅助功能，将多点云从头开始对齐到全局一致的帧。我们使用DNN来模拟高度非凸映射过程，该过程传统上涉及手工制作的数据关联，传感器姿态初始化和全局细化。我们的关键新颖之处在于，正确定义无监督损失以通过反向传播来“训练”这些DNN等同于解决基础注册问题，但是对ICP的要求实现良好初始化的依赖性更小。我们的框架包含两个DNN：一个估计输入点云姿态的本地化网络，以及一个通过估计全局坐标的占用状态来模拟场景结构的地图网络。这允许我们将配准问题转换为二进制占用分类，这可以使用基于梯度的优化来有效地解决。我们进一步表明，通过在连续点云之间施加几何约束，可以很容易地扩展DeepMapping以解决激光雷达SLAM的问题。在模拟和真实数据集上进行实验。定性和定量比较表明，与现有技术相比，DeepMapping通常能够实现更

健和准确的多点云全局注册。在模拟和真实数据集上进行实验。定性和定量比较表明，与现有技术相比，DeepMapping通常能够实现更加稳健和准确的多点云全局注册。在模拟和真实数据集上进行实验。定性和定量比较表明，与现有技术相比，DeepMapping通常能够实现更加稳健和准确的多点云全局注册。

6.

题目：Stereo R-CNN based 3D Object Detection for Autonomous Driving
作者：Peiliang Li, Xiaozhi Chen, Shaojie Shen
研究机构：香港科技大学、大疆
论文下载链接：
arxiv.org/abs/1902.0973...

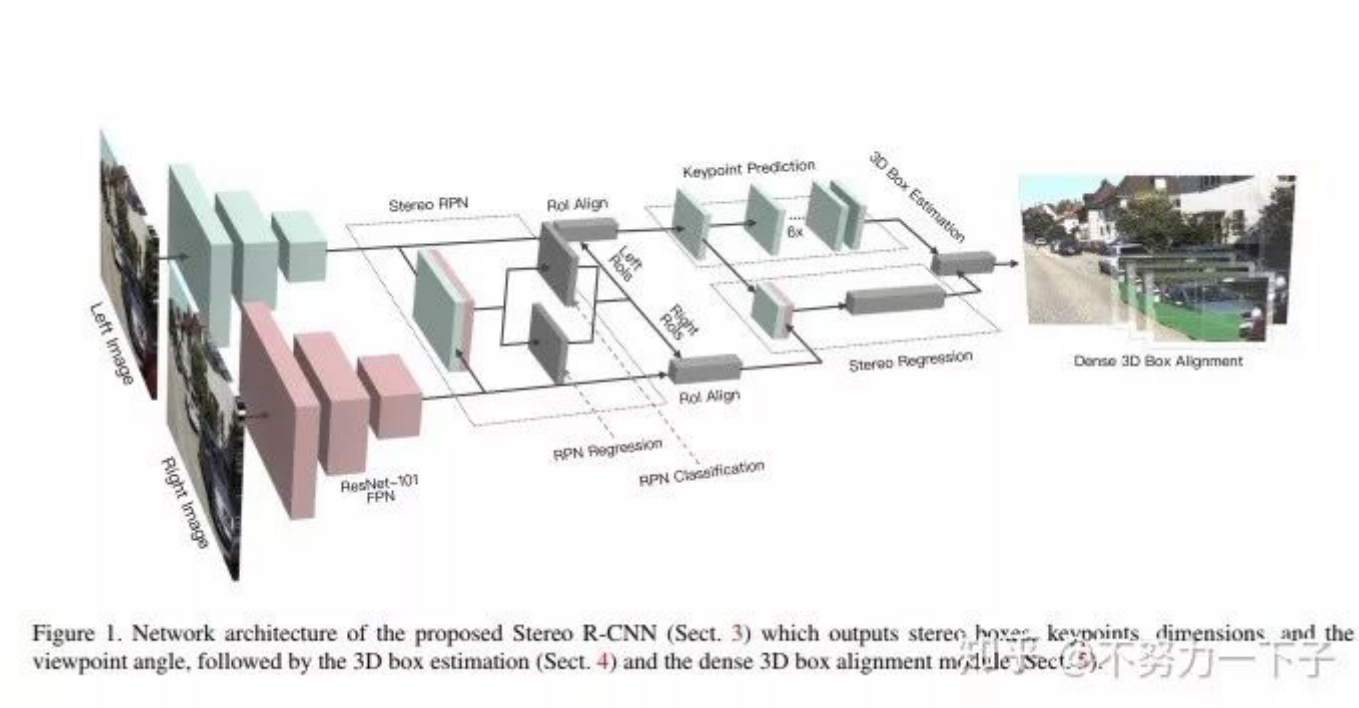


Figure 1. Network architecture of the proposed Stereo R-CNN (Sect. 3) which outputs stereo boxes, keypoints, dimensions, and the viewpoint angle, followed by the 3D box estimation (Sect. 4) and the dense 3D box alignment module (Sect. 5).

摘要 ：我们通过充分利用立体图像中的稀疏，密集，语义和几何信息，提出了一种用于自动驾驶的三维物体检测方法。 我们的方法，称为Stereo R-CNN，扩展了更快的R-CNN用于立体声输入，以同时检测和关联左右图像中的对象。 我们在立体声区域提议网络（RPN）之后添加额外分支来预测稀疏关键点，视点和对象维度，这些关键点与2D左右框组合以计算粗略的3D对象边界框。 然后，我们通过使用左右RoI的基于区域的光度对准来恢复精确的3D边界框。 我们的方法不需要深度输入和3D位置监控，但是，优于所有现有的完全监督的基于图像的方法。 在具有挑战性的KITTI数据集上的实验表明，我们的方法在3D检测和3D定位任务上的性能优于最先进的基于立体的方法约30％AP。

7.

题目：Group-wise Correlation Stereo Network
作者：Xiaoyang Guo,Kai Yang,Wukui Yang,Xiaogang Wang,Hongsheng Li
团队：香港中文大学电子工程系、商汤科技
论文链接：arxiv.org/abs/1903.0402...

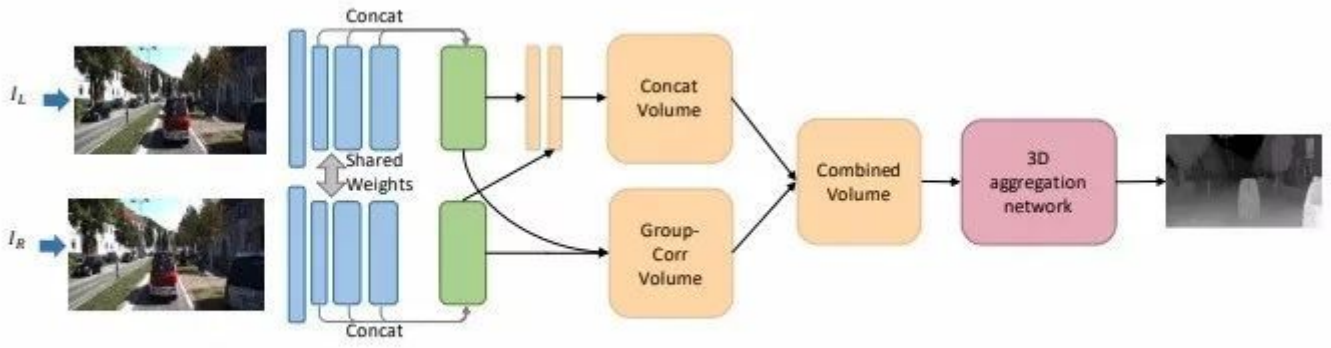


Figure 1: The pipeline of the proposed group-wise correlation network. The whole network consists of four parts, unary feature extraction, cost volume construction, 3D convolution aggregation, and disparity prediction. The cost volume is divided into two parts, concatenation volume (*Cat*) and group-wise correlation volume (*Gwc*). Concatenation volume is built by concatenating the compressed left and right features. Group-wise correlation volume is described in Section 3.2.

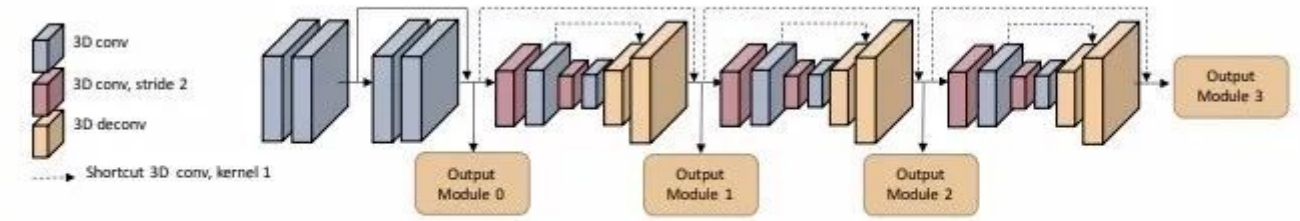


Figure 2: The structure of our proposed 3D aggregation network. The network consists of a pre-hourglass module (four convolutions at the beginning) and three stacked 3D hourglass networks. Compared with PSMNet [2], we remove the shortcut connections between different hourglass modules and output modules, thus output modules 0, 1, 2 can be removed during inference to save time. $1\times1\times1$ 3D convolutions are added to the shortcut connections within hourglass modules.

摘要：立体匹配估计整流图像对之间的差异，这对深度感测、自动驾驶和其他相关任务非常重要。先前的工作建立了在所有视差水平上具有交叉相关或串联左右特征的成本量，然后利用2D或3D卷积神经网络来回归视差图。在本文中，我们建议通过分组相关来构建成本量。左边特征和右边特征沿着通道维度被分成组，并且在每个组之间计算相关图以获得多个匹配成本提议，然后将其打包到成本量中。分组相关为测量特征相似性提供了有效的表示，并且不会丢失过多的信息，如完全相关。与以前的方法相比，它在减少参数时也能保持更好的性能。在先前的工作中提出的3D堆叠沙漏网络被改进以提高性能并降低推理计算成本。实验结果表明，我们的方法在Scene Flow，KITTI 2012和KITTI 2015数据集上优于以前的方法。此代码可通过xy-guo/GwcNet（代码待更新）获得。

8.

题目：Hierarchical Discrete Distribution Decomposition for Match Density Estimation
研究结构：伯克利DeepDrive

作者：Zhichao Yin
论文链接：arxiv.org/abs/1812.0626...

Figure 1: Illustration of HD^3. We aim to estimate discrete match distribution in this work. For reducing the infeasible computational cost, the overall distribution is decomposed into multiple scales hierarchically at learning time. The full match information can be recovered by aggregating predictions from all levels. Please refer to Sec. 3.2 for more details.

摘要：用于像素对应的现有深度学习方法输出运动场的点估计，但不表示完全匹配分布。匹配分布的显式表示对于许多应用是期望的，因为它允许直接表示对应概率。使用深度网络估计全概率分布的主要困难是推断整个分布的高计算成本。在本文中，我们提出了分层离散分布分解，称为HD3，以学习概率点和区域匹配。它不仅可以模拟匹配不确定性，还可以模拟区域传播。为了实现这一点，我们估计了不同图像尺度下像素对应的层次分布，而没有多假设集合。尽管它很简单，但我们的方法可以在既定基准上实现光流和立体匹配的竞争结果，而估计的不确定性是错误的良好指标。此外，即使区域在图像上变化，也可以将区域内的点匹配分布组合在一起以传播整个区域。

9.

题目：Deep Rigid Instance Scene Flow
研究机构：Uber ATG部门、MIT、多伦多大学
作者：Wei-Chiu Ma 、Shenlong Wang 、Rui Hu、Yuwen Xiong、 Raquel Urtasun
论文链接：
people.csail.mit.edu/we...

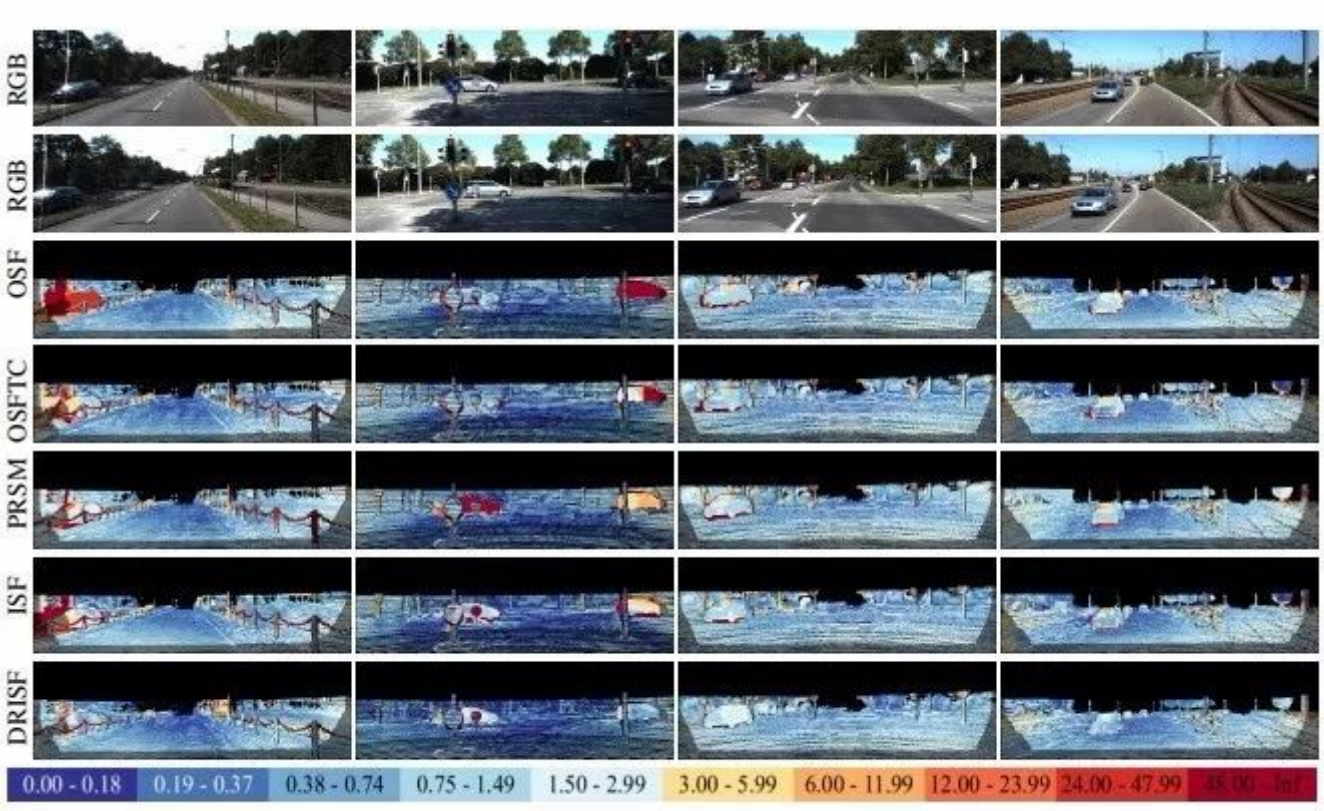


Figure 4: **Qualitative comparison on test sest:** Our method can effecitvely handle occlusion and texture-less regions. It is more robust to the illumination change as well as large displacement.

摘要： 在本文中，我们解决了自动驾驶环境下的场景流量估计问题。 我们利用深度学习技术以及强大的先验，因为在我们的应用领域中，场景的运动可以由机器人的运动和场景中的演员的3D运动来组成。 我们将问题表达为深度结构化模型中的能量最小化，这可以通过展开高斯 - 牛顿求解器在GPU中有效地求解。 我们在具有挑战性的KITTI场景流数据集中的实验表明，我们以超大的优势超越了最先进的技术，同时快了800倍。

10.

题目：MagicVO: End-to-End Monocular Visual Odometry through Deep Bi-directional Recurrent Convolutional Neural Network（单目视觉测距）
作者：Jian Jiao,Jichao Jiao,Yaokai Mo,Weilun Liu,Zhongliang Deng
研究结构: 北邮
论文链接：arxiv.org/abs/1811.1096...

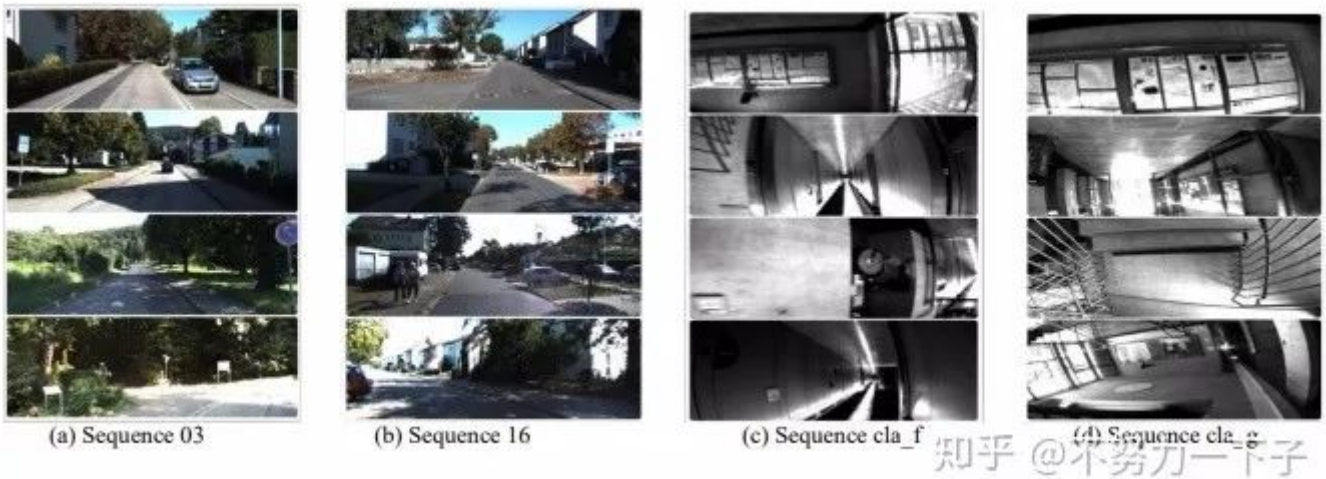


Fig.8. Sequence 03, 16 of KITTI dataset and Sequence cla_f , cla_g of ETH-asl cla dataset with 4 sample images for each sequence.

摘要：本文提出了一种解决单眼视觉测距问题的新框架，称为MagicVO。 基于卷积神经网络（CNN）和双向LSTM（Bi-LSTM），MagicVO在摄像机的每个位置输出6-DoF绝对标度姿势，并以一系列连续单目图像作为输入。 它不仅利用CNN在图像特征处理中的出色表现，充分提取图像帧的丰富特征，而且通过Bi-LSTM从图像序列前后学习几何关系，得到更准确的预测。MagicVO的管道如图1所示.MagicVO系统是端到端的，KITTI数据集和ETH-asl cla数据集的实验结果表明MagicVO比传统的视觉测距具有更好的性能（ VO）系统在姿态的准确性和泛化能力方面。

11.

题目：SSA-CNN: Semantic Self-Attention CNN for Pedestrian Detection
作者：Chengju Zhou,Meiqing Wu,Siew-Kei Lam
研究机构：南洋理工大学
论文链接：arxiv.org/abs/1902.0908...

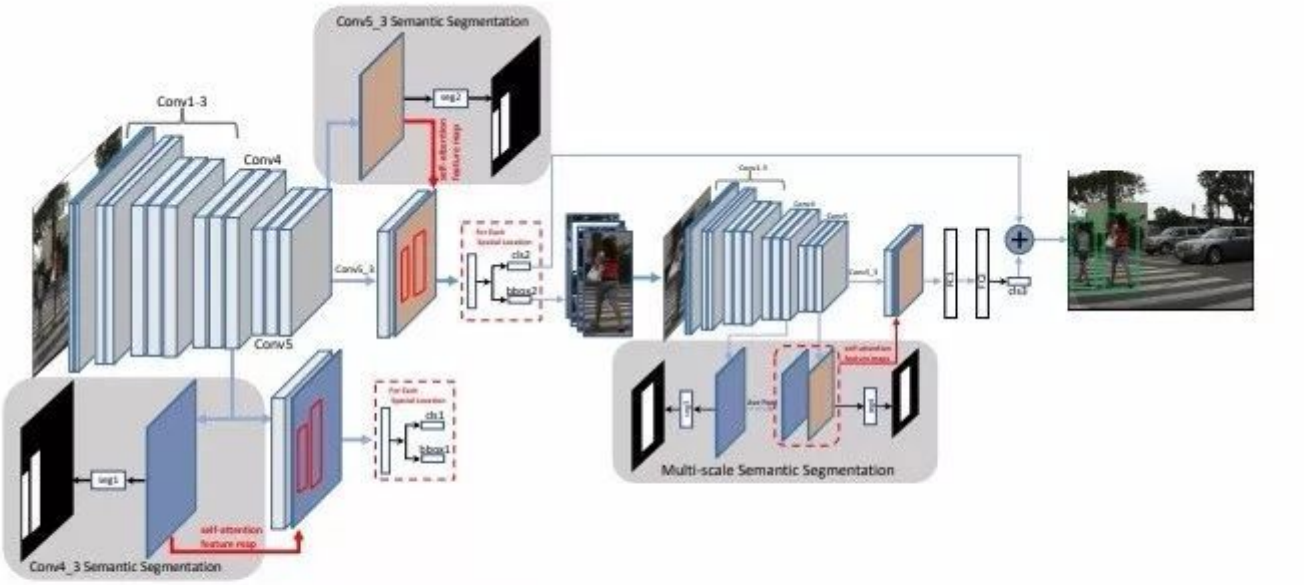


Figure 1. Overview of proposed framework. The framework consists of two stages: Semantic Self-Attention RPN (referred as SSA-RPN) and Semantic Self-Attention R-CNN (referred as SSA-RCNN). In each stage, multi-scale semantic segmentation branches are added to conv4_3 and conv5_3 layers of VGG-16 network to generate semantic segmentation results as self-attention feature maps. The self-attention feature maps are then concatenated with corresponding convolution feature maps to work as features for pedestrian detection in SSA-RPN and pedestrian classification in SSA-RCNN. The detection results are obtained by applying Non-Maximum Suppress (NMS) on candidates with combined confidence scores from *cls2* of SSA-RPN and *cls3* of SSA-RCNN. The detection and segmentation branches connected to conv4_3 layer in SSA-RPN are not used during inference.

摘要：行人检测在诸如自动驾驶的许多应用中起着重要作用。我们提出了一种方法，将语义分割结果作为自我关注线索进行探索，以显着提高行人检测性能。具体而言，多任务网络被设计为从具有弱框注释的图像数据集联合学习语义分割和行人检测。语义分割特征图与相应的卷积特征图连接，为行人检测和行人分类提供更多的辨别特征。通过联合学习分割和检测，我们提出的行人自我关注机制可以有效识别行人区域和抑制背景。此外，我们建议将来自多尺度层的语义注意信息结合到深度卷积神经网络中以增强行人检测。实验结果表明，该方法在Caltech数据集上获得了6.27%的最佳检测性能，并在CityPersons数据集上获得了竞争性能，同时保持了较高的计算效率。