**北京航空航天大学计算机学院**

**硕士学位论文开题报告**

**题 目**：基于短轨迹关联优化的多假设跟踪方法研究

**专 业**：计算机科学与技术

**研究方向**：计算机视觉

**研 究 生**：叶珍

**学 号**：SY1806608

**指导教师**：盛浩

**北京航空航天大学计算机学院**

2019年12月19日

目录

[1. 课题背景与意义 1](#_Toc26978373)

[2. 国内外研究现状 2](#_Toc26978374)

[2.1. 在线多目标跟踪算法概述 2](#_Toc26978375)

[2.2. 离线多目标跟踪算法概述 4](#_Toc26978376)

[3. 研究目标和研究内容 7](#_Toc26978377)

[3.1. 多目标关联方法研究 7](#_Toc26978378)

[3.2. 多假设剪枝方法研究 8](#_Toc26978379)

[3.3. 多目标全局优化研究 8](#_Toc26978380)

[4. 拟解决的关键问题及对应的技术路线 9](#_Toc26978381)

[4.1. 基于深度学习和短轨迹的特征关联技术 9](#_Toc26978382)

[4.2. 基于多假设概率神经网络的剪枝技术 10](#_Toc26978383)

[4.3. 基于深度预测模型的全局关联优化技术 10](#_Toc26978384)

[5. 论文工作计划 11](#_Toc26978385)

[6. 部分参考文献 11](#_Toc26978386)

# 课题背景与意义

语义分割一直是计算机视觉中非常重要的领域。从传统语义分割的发展到深度学习的普及，语义分割的任务和方法也得到了广泛的扩展和改进。在本章中，我们首先解释语义分割的背景和定义，然后从本文开始概述近年来的各种语义分割解决方案，包括这些算法提出的背景，设计思路和需要解决的问题，同时我们仍将介绍语义分割领域中当前使用的主流数据集。本章简要介绍了深度学习上下文中从浅层模型到深层模型的各种语义分割技术，包括这些技术的背景，方法和问题。如今，语义分割已成为计算机视觉的关键问题之一。通过语义分割生成的分割结果被应用到静态2D图像，视频甚至3D数据和体数据中，被广泛用于计算机视觉的各种高级应用中。从宏观意义上讲，语义分割是一个高级任务，为场景理解铺平了道路。作为计算机视觉的核心问题，场景理解的重要性越来越突出，因为现实中越来越多的应用场景需要从图像中推断出相关的知识或语义（即从具体到抽象的过程）。这些应用包括自动驾驶，人机交互，计算摄影，图像搜索引擎，增强现实等。使用各种传统的计算机视觉和机器学习技术，已经解决了这些问题。尽管这些方法非常流行，但是深度学习革命带来了相关领域的翻天覆地的变化，不仅大大提高了语义分割的准确性，而且加快了场景分割结果的生成速度。因此，包括语义分割，视觉问题在内的许多计算机开始使用深度网络架构来解决，通常是卷积神经网络CNN，而CNN的准确性甚至效率远远超过传统方法。但是，与固有的计算机视觉和机器学习分支相比，深度学习的发展还远远没有成熟。因此，语义分割的发展空间仍然很大。这个领域的快速发展使得对初学者的教育更加困难。此外，由于已经陆续提出了大量工作，因此跟上发展的步伐也非常耗时。因此，很难跟上与语义分割有关的工作，合理地解释它们的论据，过滤掉底层工作，并验证相关的实验结果是非常困难的。

# 国内外研究现状

语义分割一直是计算机视觉中非常重要的领域。从传统语义分割的发展到深度学习的普及，语义分割的任务和方法也得到了广泛的扩展和改进。通过语义分割生成的分割结果被应用到静态2D图像，视频甚至3D数据和体数据中，被广泛用于计算机视觉的各种高级应用中。从宏观意义上讲，语义分割是一个高级任务，为场景理解铺平了道路。作为计算机视觉的核心问题，场景理解的重要性越来越突出，因为现实中越来越多的应用场景需要从图像中推断出相关的知识或语义（即从具体到抽象的过程）。这些应用包括自动驾驶，人机交互，计算摄影，图像搜索引擎，增强现实等。使用各种传统的计算机视觉和机器学习技术，已经解决了这些问题。尽管这些方法非常流行，但是深度学习革命带来了相关领域的翻天覆地的变化，不仅大大提高了语义分割的准确性，而且加快了场景分割结果的生成速度。因此，包括语义分割，视觉问题在内的许多计算机开始使用深度网络架构来解决，通常是卷积神经网络CNN，而CNN的准确性甚至效率远远超过传统方法。但是，与固有的计算机视觉和机器学习分支相比，深度学习的发展还远远没有成熟。因此，语义分割的发展空间仍然很大。

传统的语义分割方法多是仅使用场景中的RGB色彩信息作为算法的输入生成用于语义分割的场景特征，而由于存在类别繁杂、遮挡严重、目标外观差异较大等原因，复杂场景中的分割任务往往更难实现。为此研究者们选择引入了场景深度信息，因为深度信息可为RGB图像提供对应的几何关系，这类额外结构信息能够有效辅助复杂和困难场景下的分割，提升分割精度。而现有的大多数工作只是简单地假设深度测量是准确的，且与RGB像素能够良好地对齐，所以获取场景准确的深度信息则显得尤为重要。而基于光场的深度提取方法在场景深度提取领域中表现出了其突出的优势。

## 基于RGB图像语义分割方法综述

在完全卷积网络出现之前，早期的研究经常使用图像块方法进行语义分割，即将图像切成块并输入到深度模型中，然后对图像块中的像素进行分类。这些分类网络通常包含一个全连接层，并且全连接层的输入需要固定大小的图像块，因此图像被切成块。构造神经网络模型的一种相对基本的方法是简单地堆叠几个卷积层（使用相同的填充以确保维数不变）并输出最终的分割图。【1】2014年，提出了全卷积网络[1]（Fully Convolutional Networks，FCN），这促进了先前分类网络的CNN结构。FCN该结构可以接受任何大小的输入图像，编码器-解码器结构也首先引入了FCN体系结构。【2】SegNet [3]的建议概括了以前的Encoder-Decoder结构，其在编码网络中的连续下采样导致输出特征图上的每个像素与输入图像中的大量空间信息重叠。最大池化层在解码网络中被重用。此应用程序改进了边界划分，并减少了端到端训练参数的数量。后，Ronneberger等。通过扩展解码器模块的容量，改进了上述全卷积架构。具体来说，此U-Net [4]结构“包括用于捕获上下文的收缩路径和用于精确定位的对称扩展路径”。 在UNet的基础上，Drozdzal [5]和其他人用堆栈中的剩余模块替换了基本卷积模块。该剩余模块在模块内部包含跳过连接，同时在编码器和解码器对应的特征图之间保留与U-Net相同的跳过连接。他们声称这种方法可以使网络收敛更快。但是，编码和解码网络的结构不能支持从输入图像中提取足够的特征信息。在后来的发展中，为了弥补这一不足，语义分割领域出现了越来越多的网络，例如PSPNet [8]，PSPNet。网络结构中有一个复杂的金字塔池层，用于提取不同规模的信息。在功能图中。与UNet相比，PSPNet具有更复杂的网络结构，并且在多类分类任务中通常更强大。对于实时视频语义分割，TDNet [11]本文提出了一种基于时间分布式网络的视频语义分割算法TDNet（时间分布式网络）。 ，其设计思想来自这样的观察：可以通过组合一系列浅网络输出的特征来获得较深网络输出的特征。在视频分割任务中，视频具有定时属性，视频分割也具有定时属性。因此，可以在一定的时间范围内将浅层网络用于特征提取，然后组合以实现深层网络输出特性。[12]本文提出了一种轻型版本的RefineNet，即轻型RefineNet，它将实时分割任务的速度从20FPS提高到55FPS，同时保持着高精度的语义分割效果。

## 基于RGBD图像语义分割方法综述

RGB-D信息就是标准的RGB图像信息中引入了深度信息，而深度信息可为RGB图像提供对应的几何关系。引入深度信息后，其提供的额外结构信息能够有效辅助复杂和困难场景下的分割。比如，与室外场景相比，由于语义类别繁杂、遮挡严重、目标外观差异较大等原因，室内场景的分割任务要更难实现。此时，在结合深度信息的情况下，能够有效降低分割的难度。基于ReNet [14]，LSTM-CF使用长时记忆和短时记忆以及DeepLab来实现分段。 LSTM-DF主要包括四个部分：用于垂直深度上下文提取的层，用于垂直光度上下文提取的层，用于将光度和深度上下文集成到2D全局上下文中的内存融合层以及像素级场景分割Floor。输入深度信息后，LSTM-CF使用HHA描述将深度信息转换为视差，表面法线和高度的三个通道。随后，使用ReNet提取不同方向的上下文信息，并在两个方向上双向传播。同时，对于RGB通道信息，网络使用上图所示的卷积结构来提取特征，并使用插值将所有级别的特征恢复为相同的分辨率，并将它们级联。之后，ReNet还用于获取上下文信息。结果表明，分割的主要精度仍来自于RGB信息，但是随着深度信息和上下文信息的引入，网络的精度可以得到一定程度的提高。在RGBD的语义分割领域上论文【16】提出了一种基于三维点云的三维图神经网络，它在点云的基础上构造了一个k近邻图并且在此基础上提取RGBD信息获取语义分割结果。论文【17】中，我们提出了一个注意补充网络（ACNet），它选择性地从RGB和深度分支中收集特征。主要贡献在于注意补充模块（ACM）和具有三个并行分支的架构。更确切地说，ACM是一个基于通道注意的模块，可从RGB和深度分支中提取加权特征。该架构提供原始RGB和深度分支的推断，并同时启用融合分支。基于上述结构，ACNet能够利用来自不同渠道的更高质量的功能. 15】提出了现有的大多数工作只是简单地假设深度测量是准确的，且与RGB像素能够良好地对齐，由此将该问题建模为交叉模式特征融合以获得更好的特征表示从而实现更准确的分割，但是，通常传感器无法获得令人满意的精准的深度结果，实际的深度数据通常比较嘈杂，这可能会随着网络的深入而降低分割的准确性。提出了一个统一而有效的跨模态引导的编码器，该编码器不仅可以有效地重新校准RGB特征响应，而且还可以通过多个阶段来提取准确的深度信息，并将两者交替合并在一起重新校准的表示形式，同时提出的编码器可以轻松地注入到先前的编码器-解码器结构中，以提高其在RGB-D语义分割任务上的性能。

## 基于光场的场景深度提取方法综述

前人的研究已经充分证明了在RGB图像的语义分割上引入场景的深度信息对于语义分割的精度是有一定提升的，而提升的效果除了在于如何引入深度信息之外，更重要的事在于如何获取精确的场景深度信息，而传统的深度提取方法都存在着精确性和鲁棒性的缺陷，因此在本文中，我们引入了在深度估计领域上前沿的算法，也就是光场深度估计，引入光场信息的深度估计在精确度跟鲁棒性上相较于传统方法都有着十分明显的提升。本章节将具体介绍光场深度提取的相关概念与内容。与传统图像不同，光场相机的特殊成像方法使光场图像包含各种不同的表示形式，Jeon 等人[24] 设计了基于学习的匹配项算法，针对不同场景下的光场图像采用随机森林算法从 31 个不同的匹配项中选择最为可靠的匹配代价结合方式以及计算方法，进一步得到准确的深度估计结果。Johannsen 等人[26] 提出了基于稀疏编码的深度估计框架，通过移动中心视角图像对应点的位置仿真出不同深度标签下的光场图像构成字典元素，并采用该字典对真实的光场图像进行稀疏表示。对应稀疏编码的系数可以用于计算深度标签，系数分布的统计分析可以用来判断场景中是否存在由于透明或者反射造成的多层深度，并进一步通过高斯混合模型求得不同表面的深度信息。针对光场极面图像的斜率估计，Wanner 等人[28] 提出了基于结构张量的方法衡量直线方向，并通过建立能量函数得到全局优化的深度图。不同于传统的点与点之间匹配计算，基于结构张量的方法计算速度较快且精度不受标签的限制。Shin等人[1]在2018年提出一种基于极面图像与机器学习的光场深度提取方法，他们提出了基于全卷积神经网络的EPINet，一种可以在光场深度提取中保持着高精确性与速度的深度网络结构，该网络对光场图像中四个方向的极面图像都进行了特征提取与融合，获取的深度提取结果更具精确性与鲁棒性。

# 研究目标和研究内容

语义分割一直是计算机视觉中十分重要的领域，从传统语义分割的发展到随着深度学习的流行，语义分割任务与方法也得到了大量的拓展和进步。RGB-D信息就是标准的RGB图像信息中引入了深度信息，而深度信息可为RGB图像提供对应的几何关系可以提升语义分割的精确度。为获取精确的场景信息，我们决定从光场中提取场景的深度信息。

## 基于RGBD图像的语义分割研究

语义分割是对图像的像素级分类，语义图像分割的目标是用对应的所表示的类来标记图像的每个像素。而现阶段大多基于深度学习的语义分割算法都是采用Encoder使用池化层逐渐减少空间维度从而获取场景特征图，而Decoder使用上采样的方法恢复空间维度和细节信息，最终获取与输入大小一致的特征图，通过高低层网络之间的跃接，在保持着特征提取效率同时保留了足够的像素级细节。而近年来，在面对复杂环境下的场景，仅仅基于图像的RGB信息已经很难达到一个十分高的分割结果，所以引入深度信息后，其提供的额外结构信息能够有效辅助复杂和困难场景下的分割。比如，与室外场景相比，由于语义类别繁杂、遮挡严重、目标外观差异较大等原因，室内场景的分割任务要更难实现。此时，在结合深度信息的情况下，能够有效降低分割的难度

## 基于光场的场景深度提取研究

光场图像中包含来自场景的多视角信息，这使得深度估计成为可能。相对于传统的多视角深度估计问题，光场视角更加丰富且均匀整齐地排列在规则网格上，视角间场景几何信息变换规律更加清晰，使深度提取的难度大大降低。现有的光场深度提取算法在大多数情况下都能保持较高的准确性。常用的光场四维简化通过两个平行平面参数化表示空间中所有光线，可以从中提取场景的完整三维信息。机器学习方法可以整合更大范围的场景信息与视角信息，也可以从大规模数据中总结丰富的几何映射规律，大大提高了光场深度提取的精确性。而为了提供精确的场景深度信息给语义分割模型，我们选择合适的深度网络提取光场信息来进行场景深度提取。

## 基于光场的场景语义分割研究

光场能为场景语义分割提供更丰富和准确的场景信息，比如场景的深度信息。但对于语义分割的精度的提升，除了在于准确的场景信息外，还在于合适的融合方法。为了让语义分割网络能够充分学习到光场深度提取的信息，我们提出了显示与隐示的融合方式。显示融合即将光场深度信息与场景RGB信息融合作为场景的RGBD信息进入特征提取模块；而隐示融合的思路来源于，在光场深度提取的过程中我们已经充分利用了场景的RGB信息，基于此我们没必要继续重复提取场景的RGB信息，而是在深度提取的过程中对网络做一些调整，直接完成从光场图像到语义分割结果的生成。

# 拟解决的关键问题及对应的技术路线

## 基于深度学习的RGBD图像语义分割技术

论文提出了RefineNet 的轻量化版本Light-Weight RefineNet ，针对实时分割任务，将速度从20FPS提升到了55FPS。总体网络结构，分为下采样的encoder部分和上采样的decoder部分。网络主要包含4个模块，RCU，CRP，FUSION，CLF。为了轻量化该网络，分别使用RCU-LW，CRP-LW。该网络在不影响RefineNet语义分割的精度的前提下大幅降低了网络深度，提升了分割的效率。但是该结果在物体边缘的分割结果却存在着许多明显的错误，为解决该问题，我们又将金字塔型池化层引入，金字塔pooling模型通过应用大尺寸核的pooling层用来提取特征图中不同尺度的信息，使得分割网络可以结合上下文场景的信息进行语义分割。如图与图。

## 基于光场的场景深度特征提取技术

EPINet[1]是可以获取光场图像中不同视角间的联系和信息的一种卷积神经网络，能够以较快的速度完成光场深度提取。如图 3‑1所示，是EPINet的网络结构图。

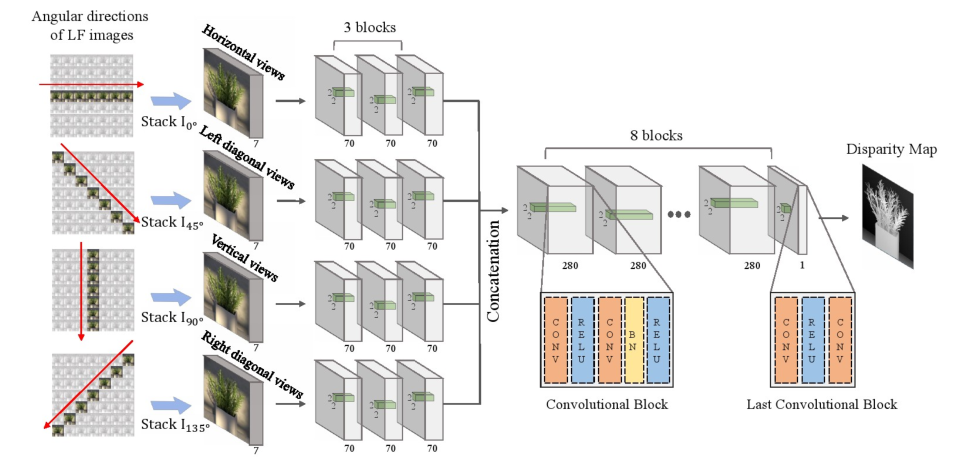


图 3‑1 EPINet网络结构图

从EPINet网络图我们可以解析其结构，从左往右的网络层依次是：

1．多流网络(Multi-stream network)，多流网络是为具有一致基线的四个视点所构建的，与单流网络训练后的效果进行对比后，多流网络的表现要比单流网络更精确10%左右。

2. 三个全卷积神经网络，由于完全卷积网络已被认为是用于像素密集预测的有效架构。为了处理光场相机中视角间基线过短，使用大小为2\*2，步长为1的卷积核来更多地保留图像特征信息。

3. 合并网络由八个卷积块组成，它们通过多流网络查找通过的特征之间的关系

4. 为了用子像素精度推断视差值，EPINet使用Conv(卷积层)-ReLU(整流线性单元)-Conv结构构造最后一个块。

EPINet具有着在精确性和速度上的优越性，其在BadPix0.07(预测像素值与真实像素值差值大于0.07的像素比例)、MSE(Mean-square error，均方根误差)等指标上表现十分突出，所以我们选择EPINet作为本研究的光场深度提取的基础网络。

## 基于深度学习与光场的语义分割技术

将光场信息融入到语义分割网络当中我们拟采取两种融合方法，显示融合以及隐示融合。显示融合即将从光场提出出的深度信息与场景的RGB信息融合作为特征提取网络的输入参与到语义分割当中，这是一种比较直观与直接的融合方式，但是这种融合方式的弊端也是十分明显的，因为是先从光场中提取出深度再融合进语义分割的网络中，这方式会明显增加整个深度网络的深度，而在提取深度过程中也已经充分利用了图像的RGB信息，所以整个网络重复提取了图像RGB信息。为提升融合效率，我们也提出了隐示融合的方案，即修改深度提取的部分网络结构，使得提取的RGB信息既可以用于场景深度提取，也可以用于场景语义分割。两种融合的方式如图与图，而隐示融合的网络结构正在尝试修改与优化。

# 论文工作计划

* 2019年12月～2020年2月

整理资料，学习研究近几年来的RGBD语义分割方法

* 2019年3月～2020年5月

研究基于深度学习的光场深度提取方法

* 2020年6月～2020年8月

研究语义分割与光场的融合，从显示融合到隐示融合，最终完善出一个高效的光场语义分割网络结构。

* 2020年9月～2020年12月

资料整理和论文撰写

# 部分参考文献

[1]. W. Luo, J. Xing, X. Zhang, X. Zhao, and T.-K. Kim. Multiple object tracking: A literature review. arXiv preprint arXiv:1409.7618, 2014.

[2].S. Oron, A. Bar-Hillel, D. Levi, S. Avidan, Locally orderless tracking, in: 610 Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 1940–1947.

[3] D. Wang, H. Lu, M.-H. Yang, Online object tracking with sparse prototypes, Image Processing, IEEE Transactions on 22 (1) (2013) 314–325.

[4] D. Hall, P. Perona, From categories to individuals in real time – a unified boosting approach, in: Computer Vision and Pattern Recognition

(CVPR), 2014 IEEE Conference on, 2014, pp. 176–183.

[5] H. Possegger, T. Mauthner, P. Roth, H. Bischof, Occlusion geodesics for online multi-object tracking, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, 2014, pp. 1306–1313.

[6] Kalman R E. A New Approach to Linear Filtering and Prediction Problems[J]. Journal of Basic Engineering Transactions, 1960, 82:35-45.

[7] Vermaak J, Doucet A, Perez P. Maintaining multimodality through mixture tracking[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2003:1110.

[8] Okuma K, Taleghani A, Freitas N D, et al. A Boosted Particle Filter: Multitarget Detection and Tracking[C]// European Conference on Computer Vision. DBLP, 2004:28--39.

[9] Breitenstein M D, Reichlin F, Leibe B, et al. Robust tracking-by-detection using a detector confidence particle filter[C]// IEEE, International Conference on Computer Vision. IEEE, 2010:1515-1522.

[10]. Wu, Y., Lim, J., & Yang, M. H. (2013). Online object tracking: A benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2411-2418).

[11]. S.-H. Bae, K.-J. Yoon, Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning, 2014, pp. 1218–1225. 625

[12].J. Supancic, D. Ramanan, Self-paced learning for long-term tracking, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 2379–2386.

[13] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, “Online multiperson tracking-by-detection from a single, uncalibrated camera,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 33, no. 9, p. 1820, 2011.

[14] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, “Who are you with and where are you going?” in Computer Vision and Pattern Recognition, 2011, pp. 1345–1352.

[15] Bae S H, Yoon K J. Confidence-Based Data Association and Discriminative Deep Appearance Learning for Robust Online Multi-Object Tracking[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):1-1.

[16] Kieritz H, Becker S, Hubner W, et al. Online multi-person tracking using Integral Channel Features[C]// IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2016:122-130.

[17] Y. Wu, J. Lim, and M. H. Yang, “Online object tracking: A benchmark,” in IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2411–2418.

[18]. Sadeghian A, Alahi A, Savarese S. Tracking The Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies[J]. 2017.

[19] Y. Xiang, A. Alahi, and S. Savarese, “Learning to track: Online multiobject tracking by decision making,” in IEEE International Conference on Computer Vision, 2016, pp. 4705–4713.

[20] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,”IEEE transactions on pattern analysis and machine intelligence, vol. 32,no. 9, pp. 1627–1645, 2010.

[21] R. Girshick, “Fast r-cnn,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.

[22].L. Zhang, Y. Li, and R. Nevatia. Global data association for

multi-object tracking using network flows. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008.

[23].Yang, B., & Nevatia, R. (2012, June). An online learned CRF model for multi-target tracking. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on (pp. 2034-2041). IEEE.

[24]. Kim C, Li F, Ciptadi A, et al. Multiple Hypothesis Tracking Revisited[C] IEEE International Conference on Computer Vision. IEEE, 2015:4696-4704.

[25]J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, “Multiple object tracking using k-shortest paths optimization.” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 33, no. 9, pp. 1806–1819, 2011.

[26]A. A. Butt and R. T. Collins, “Multi-target tracking by lagrangian relaxation to min-cost network flow,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1846–1853.

[27]H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, “Globally-optimal greedy algorithms for tracking a variable number of objects,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1201–1208.

[28]L. Zhang, Y. Li, and R. Nevatia, “Global data association for multi-object tracking using network flows,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[29]A. Milan, K. Schindler, and S. Roth, “Detection-and trajectory-level exclusion in multiple object tracking,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3682–3689.

[30]A. Milan, L. Lealtaixe, K. Schindler, and I. Reid, “Joint tracking and segmentation of multiple targets,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5397–5406.

[31]Reid D B. An Algorithm for Tracking Multiple Targets[C]// Decision and Control Including the, Symposium on Adaptive Processes, 1978 IEEE Conference on. IEEE Xplore, 1979:1202-1211.

[32]C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, “Multiple hypothesis tracking revisited,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4696–4704.