Computer Science and Application 计算机科学与应用, 2023, 13(7), 1343-1351
Published Online July 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.137132

Hans汉斯

自动驾驶中点云与图像多模态融合研究综述

孟　玥，李士心*，陈范凯，刘　宸，丛笑含

天津职业技术师范大学电子工程学院，天津

收稿日期：2023年6月6日；录用日期：2023年7月5日；发布日期：2023年7月12日

摘　要

针对复杂多变的道路环境，综合国内外研究现状，本文从激光雷达和摄像头方面论述了汽车自动驾驶中的网络输入的格式，并以两种传感器融合为例，归纳了自动驾驶汽车环境感知任务中多模态传感器融合的分类方法，在此基础上，又从融合阶段的角度总结出另一种分类，简化了融合方法的分类和理解，强调了融合程度的区别以及融合方法的整体性，这种分类对于推动融合方法的研究和发展具有创新价值。最后分析传感器融合所遗留的问题，对未来的发展趋势进行预测。

关键词

激光雷达，摄像头，多模态，传感器融合

Research Review of Multimodal Fusion of Point Cloud and Image in Autonomous Driving

Yue Meng, Shixin Li*, Fankai Chen, Chen Liu, Xiaohan Cong

College of Electronic Engineering, Tianjin University of Technology and Education, Tianjin

Received: Jun. 6th, 2023; accepted: Jul. 5th, 2023; published: Jul. 12th, 2023

Abstract

In view of the complex and changeable road environment, this paper discusses the format of network input in auto driving from the aspects of laser radar and camera, and summarizes the classification method of multimodal sensor fusion in the environmental perception task of autonomous vehicle, based on which, another classification is summarized from the perspective of fusion stage,

*通讯作者。

文章引用: 孟玥, 李士心, 陈范凯, 刘宸, 丛笑含. 自动驾驶中点云与图像多模态融合研究综述[J]. 计算机科学与应用, 2023, 13(7): 1343-1351. DOI: 10.12677/csa.2023.137132

simplifying the classification and understanding of fusion methods, emphasizing the differences in fusion levels and the integrity of fusion methods, and this classification has innovative value for promoting the research and development of fusion methods. Finally, the issues left by sensor fusion and predict future development trends is analyzed.

Keywords

LiDAR, Camera, Multimodal, Sensor Fusion

1. 引言

不同的传感器数据经过预处理或同一种传感器数据经过不同的预处理，都可视作不同的数据模态，传感器的融合也就是数据的融合。当今，自动驾驶是车辆工程的研究热点，车辆安装摄像头、激光雷达、毫米波雷达等传感器用于感知道路环境信息，由于原始数据噪音大，信息利用率低以及多模态传感器未对齐等原因，很难提高感知的准确性和容错率，如图 1 所示，RGB 相机能够获取具有颜色，纹理，轮廓等稠密的特征信息，但在光照不足、曝光过度情况下效果较差，并且缺少深度信息；而雷达拥有获取距离信息的能力，但因为点云本身具有稀疏性和不规则性，容易出现小目标漏检的状况。基于上述情况，一些研究将摄像头和激光雷达这两个分支组合使用，融合图像与点云数据，两者互补在感知任务上的性能。

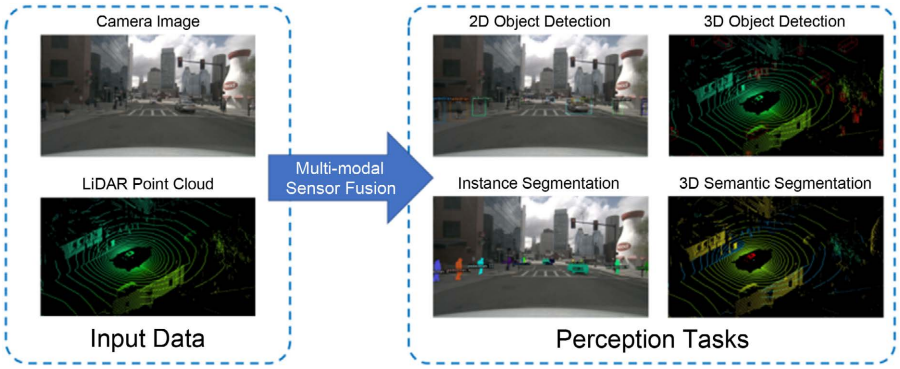

Figure 1. Image and radar perception tasks
图 1. 图像与雷达的感知任务

综合激光雷达点云数据对空间的位置信息和 RGB 图像丰富的语义信息，该方法有着巨大优势，遵循传统融合分类方法可将其分为前融合(Early Fusion)、深度融合(Deep Fusion)、后融合(Late Fusion)三种[1]。Early Fusion 以 PointPainting [2]为代表，是由 Vora 等人提出用图像语义分割的结果来给点云“染色”的方法。Deep Fusion 是多模态数据融合的主流方式，如 Qi 等人提出的 F-PointNet (Frustum PointNet) [3]，在 PointNet [4]与 PointNet++ [5]之上融合 RGB 图像所提出的一种两阶段的方法。与之相比，Liang 等人提出的 ContFuse [6]是 Deep Fusion 更好的范例，将前视视角的 RGB 特征转换到鸟瞰视角(BEV)是其主要创新点。Yoo 等人提出的 3D-CVF [7]利用跨域特征映射模块来提取多个摄像头的图像特征转换到鸟瞰视

角并连接。与 PointPainting 类似，Huang [8]等人提出了一种利用图像的语义特征来为点云增加信息量的方法。Late Fusion 方法以 Chen [9]等人提出的 MV3D 作为典型。Ku，Mozifian [10]等人则在 MV3D 基础上进一步提出了 AVOD。本文首先对两种传感器网络的输入格式进行介绍，其次描述了融合方法与创新点，最后分析融合方法存在的问题并对未来的发展趋势进行预测。

2. 图像和激光雷达的特征表示

输入数据的表示形式对深度学习模型影响巨大，为了提高模型的性能和泛化能力，需要对原始数据进行预处理。因此，本节介绍图像和激光雷达数据的表示格式，首先基于图像可细分为单目图像和双目图像，然后基于激光雷达可分为将点云数据投影和直接利用原始点云。

2.1. 图像

摄像头作为 2D 数据采集或 3D 目标检测和语义分割任务中最常用的传感器，其优点是检测信息全面、价格相对低廉，缺点是会受到雨雪天气和光照的影响。其中单目图像在 2D 检测中能够提供丰富的纹理信息，但是在 3D 检测中，因为缺乏深度信息而难以判断物体的尺寸、姿态、大小。2016 年，Chen [11]等人提出了 Mono3D 方法，该方法在计算损失函数时，误差会不断地累积，从而导致误差越来越大，所以精度不高，另外，大量的先验信息与密集的候选框让网络十分复杂，所以无法实现端到端的检测。Simonelli [12]等人提出了 MonoDIS 方法解决了误差累积问题。2019 年，Brazil [13]等人研究发表 M3D-RPN 网络。2020 年，Qian [14]等人提出了一种端到端的基于图片生成伪雷达的方法。相比单目图像，双目图像在精度方面有明显提升。例如，Qin [15]等人提出了 TLNet 模型，和基于单目的 3D 目标检测相比，放弃了深度图的输入，直接在三维空间中进行操作，在一定程度上降低了信息的损失，并使用了三角测量法提升了检测精度。单目图像通过深度估计预测出的深度信息偏差比双目图像要大，两者均存在信息丢失的情况，所以仅用图像进行 3D 目标检测精度较低。

2.2. 激光雷达

点云是由大量的离散点构成的三维数据集合，每个点都具有自己的坐标位置，点云数据和图像数据相比，具有更精确的深度信息，能更加稳健地克服现实场景中的气候、光照等环境因素，但是点云数据的稀疏和无序不规则的特性，会消耗大量的计算资源和时间。为了解决这一问题，目前常见的对点云数据进行间接处理方法包括两种，一是将三维点云场景投影到鸟瞰图视角(BEV)或者前视角(FV)，二是将点云体素化[16] [17]处理，然后体素级的数据投影到平面。这两种方法在使点云进行规则化的过程中都会带来一定程度的信息损失，另外，研究者们又提出了直接处理点云数据的网络。Charles Qi 等人提出 PointNet 网络，又在此基础上改进提出 PointNet++网络。PointRCNN [18]网络是第一个采用两阶段和 anchor-free 方法直接处理点云数据的三维目标检测网络，它能够从点云数据中检测出目标物体的位置、朝向和大小等信息，并且生成一个包围盒来描述目标，该方法也不需要使用预定义的锚点，不仅计算效率高，而且可以适应更广泛的目标形状和大小变化。

3. 融合分类方法

在自动驾驶领域中，多模态融合是指将来自多个不同传感器的数据进行集成和处理的技术，不同模态的信息进行融合，能够提高机器学习模型的性能和准确度，多种传感器呈现互补关系，在融合过程中，每个模态都可以提供不同的信息和特征，从而具有更好的表达能力。从传感器融合的角度分析，可分为三类，分别是前融合、深度融合和后融合，然而最近所研究的工作无法直接归为以上三类，例如文献[19]中针对激光雷达的 3D 目标检测任务，提出了新的神经网络架构，该架构将点云数据映射到一组刻度不

变的空间中，并使用特征重新加权方法对不同距离区域的目标进行加权融合，增强对不同尺度目标的检测能力。又如文献[20]中，为了优化点云数据本身的稀疏性，提出了一种称为 Faraway-Frustum 的方法，其中“Faraway”指的是激光雷达数据能够获取更远距离的信息，“Frustum”指的是摄像头数据在 3D 空间中的投影区域，通过将两种传感器的数据进行融合，可以更加准确地确定物体的边界框。根据特征融合的阶段，本小节归纳出另一种分类方法，将所有的融合方式划分为强融合和弱融合，如图 2 所示。下面我们分别介绍强融合中的前融合、深度融合、后融合和不对称融合。

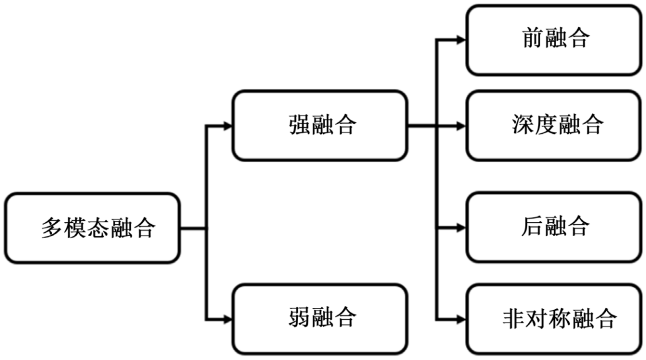

Figure 2. Multimodal fusion classification
图 2. 多模态融合分类

3.1. 强融合

3.1.1. 前融合

前融合是基于数据的融合方法，以雷达和图像融合为例，指的是对雷达分支的数据和相机分支的数据或特征进行融合。如图 3 所示，图像经过语义分割后与雷达点早期信息交互，这样操作拓展了前融合阶段中图像数据的数据级定义，更加利于 3D 目标检测，语义分割的目标是将图像中每个像素分到预定义的类别中，其中 Enet [21]深度神经网络是最有效的模型之一，它使用了特殊的编码器－解码器结构来减少计算量。此后，T. Samann [22]等人用信道修剪法[23]应用于 ENet 网络来提高效率。

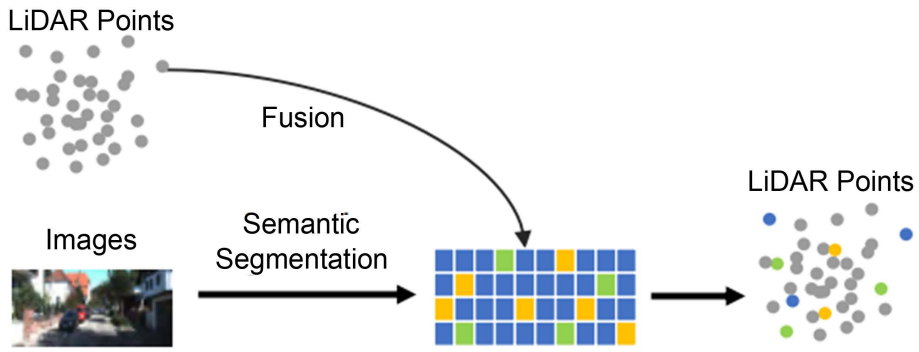

Figure 3. Early fusion
图 3. 前融合

雷达分支没有经过特征提取阶段，不会失去一部分可解释性，所以雷达分支的数据表示具有直观的可视化效果。雷达数据可以是具有反射强度的 3D 点、体素和由点云投影得到的 2D 图等。文献[24]将 3D 激光雷达点云转换为 2D 图像，并利用成熟的 CNN 技术融合图像中的特征级表示，从而实现更好的性能。文献[25]将图像分支中的语义特征和雷达点云预处理得到的体素融合在一起，来提高目标检测和追踪模型

的性能。前融合方法可以保留各个信息源的原始信息，能够简单高效的实现和部署，但融合结果的准确性和完整性不足，由于缺乏信息源之间的交互和协调，无法充分利用信息源之间的相关性和互补性，因此研究人员逐渐提出了其他更复杂的融合方法。

3.1.2. 深度融合

深度融合是基于模型的融合方法，以雷达和图像融合为例，指的是对雷达分支的特征和相机分支的数据或特征进行融合。深度融合的主要思想是利用神经网络结合不同传感器的数据进行特征融合和决策融合，从而提高目标检测和跟踪的准确性和鲁棒性。如图 4 所示，先使用体素化的方法将雷达点云数据转化为 3D 体素网络数据，再对每个体素进行特征提取，如点云的密度、高度、垂直角度等，得到一个体素特征向量，在处理图像时，可以使用卷积神经网络(CNN)来提取 2D 特征，这些特征图可以表示图像的不同层次和抽象程度的特征信息，之后确定不同特征之间的权重，将两个模态的特征进行融合。文献[26]使用特征提取器分别获取雷达点云和相机图像的特征表示，并通过一系列下游模块将特征融合以进行更准确的目标检测。文献[27]提供了一个极端天气条件下的多模态数据集，并采用深度融合的方式将不同传感器的数据进行融合，从而有效提高了自动驾驶模型在极端天气下的鲁棒性。深度融合方法能够通过深度学习模型学习复杂的特征表示，挖掘信息源关联性，实现端到端的训练，但由于需要大量的数据，所以需要较高的计算资源和时间成本，这就会使深度融合难以应用。最早的深度融合方法主要基于传统的深度学习模型，如卷积神经网络(CNN)和循环神经网络(RCNN)，为了更好地关注不同信息源的重要性，引入了注意力机制，从而提高融合结果的质量。

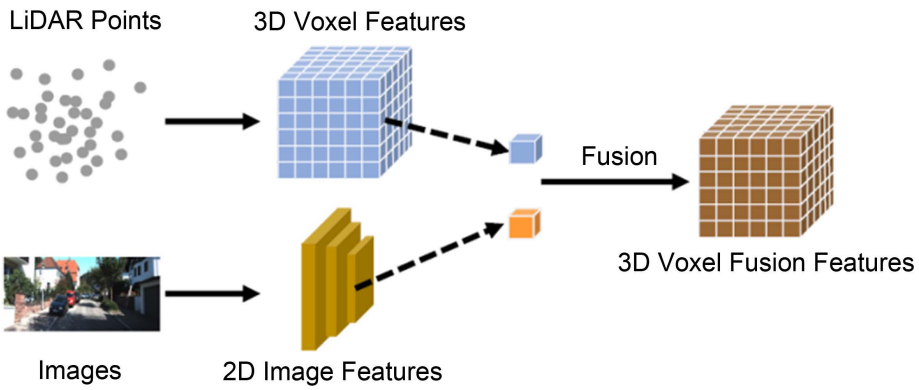

Figure 4. Deep fusion
图 4. 深度融合

3.1.3. 后融合

后融合也是基于数据的融合方法，图 5 为例，将雷达数据与图像数据单独处理，得到两个不同的目标检测和分类结果，再进行对齐，即对于同一目标再两个结果中的标记进行匹配，确定它们的空间位置和特征，接着使用融合方法将结果进行融合，过程中不需要进行特征融合，也无法利用各个模态之间的相互依赖关系，所以可能需要大量的分类器或者检测器，这也就意味着会造成计算资源的浪费。文献[28]采用后期融合方法，对图像分支的 2D 数据与雷达分支的 3D 数据进行处理，通过对每个 3D 区域方案进行二次细化得到最终结果，对于重叠区域，采用多个统计特征，例如置信度、距离和 IoU 等，进一步筛选和优化，提高目标检测和分类的准确性和可靠性。文献[29]采用后融合方法及卡尔曼滤波器对移动的目标进行跟踪，利用传感器输出的历史数据和当前数据来预测目标的运动轨迹和状态。后融合方法具有较高的可扩展性，可以方便地添加、替换或调整信息源，保留信息的多样性，提供清晰的决策过程和解释

结果，便于分析和理解系统的行为，缺点是信息交互仍有不足，在处理多模态数据融合时需要额外的机制来处理不同类型信息源之间的融合和交互。最早的后融合方法采用简单的融合方式，如加权平均或投票机制，之后研究者们提出学习融合权重的方法，模型能够自适应地决定不同信息源的重要性，随着深度学习发展，又将神经网络引入后融合方法中，通过设计多个分支网络或使用多层感知机(MLP)，使后融合模型能学习到更丰富的特征表示和融合模式。

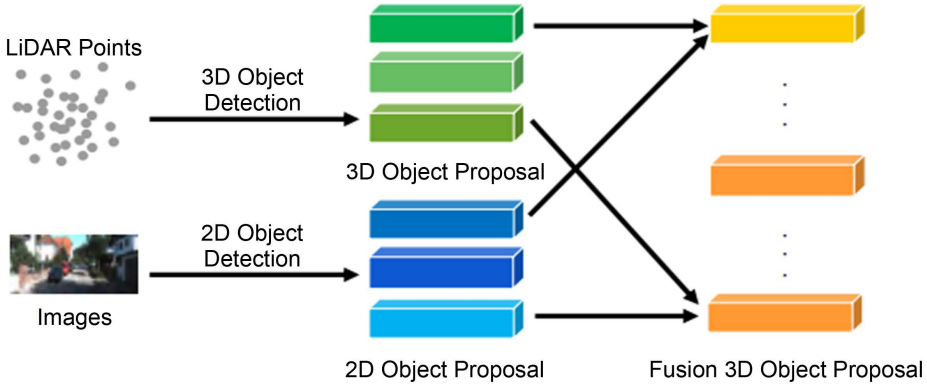

Figure 5. Late fusion
图 5. 后融合

3.1.4. 不对称融合

按不同的权限处理不同的传感器数据，将目标级别的信息从一个分支与数据级别或特征级别的信息从另一个分支进行融合的方法定义为不对称融合(Asymmetry-Fusion)。不对称融合强调至少有一个分支占据主导位置，其他分支则用于提供辅助信息预测最终结果。如图 6 所示，当雷达与图像融合时，对点云数据使用 3D 目标检测算法提取 3D 目标区域，对图像数据使用语义分割算法将图像中的每个像素分配到不同的语义类别中，然后进行匹配，对于匹配成功的目标将其融合，未匹配成功的，可根据上下文信息、形状等进行推断，或使用其他方法生成新的目标检测或语义分割结果。例如文献[30]中重点关注 2D 检测，利用激光雷达分支的 3D 区域，来指导 2D 数据进行融合，以便进一步细化。文献[31]中先用雷达点云预测 3D 候选区域，再用候选区域和 RGB 图像获取目标的多视图图像，并进一步利用多视图图像的特征对之前的检测结果进行修正。由此可见不对称融合可以提高系统的鲁棒性，能够通过合理分配权重和主导分支，根据不同的环境和任务需求进行调整和优化。不对称融合方法具有灵活性、多样性和适应性等优点，但在实际应用中，如何准确地确定不对称融合的权重可能具有一定的挑战性，如果对某个信息源的权重设置过低，可能无法充分利用该信息源提供的有效信息，从而影响融合结果的完整性。随着深度学习的发展，研究者们通过引入注意力机制或设计多个分支网络，能将不同信息源更好地进行建模和融合。

3.2. 弱融合

弱融合与上述强融合不同，弱融合不直接从多模态分支融合，通常使用基于规则的方法，利用一种模式中的数据作为监督信号，以指导另一种模式的交互，融合过程更加灵活，不受严格的模态对齐限制。弱融合与不对称融合不同，在某种情况下，它能直接将选中的原始雷达信息输入到雷达主干中，过程中不直接与图像分支主干进行特征交互，会通过一些弱连接的方式，比如将信息送入 loss 函数中训练，进行最后的信息融合。两者的区别在于信息的交互方式，弱融合更注重数据之间的独立处理和监督信号的使用，而不对称融合则更注重信息交互和特征融合。

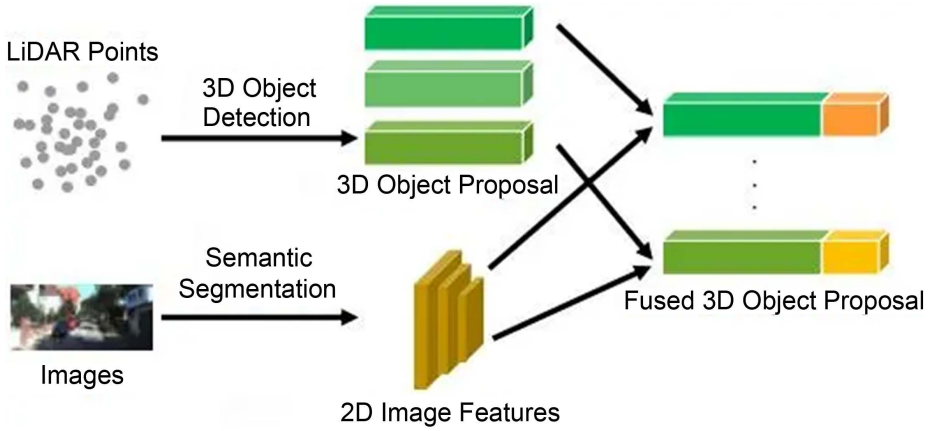

Figure 6. Asymmetry-fusion
图 6. 不对称融合

4. 融合模型存在的问题和解决办法

在自动驾驶领域中，多模态信息融合仍面临两大挑战，分别是融合模型的不对齐及信息丢失。例如，雷达与图像的相位不完全一致，相位差异会导致图像中的一些区域被错误地归为目标，而其他区域被错误地归为环境。不同传感器的测量误差和噪声也会影响融合模型在估计距离时产生偏差，导致目标的检测和跟踪失败。为解决此类问题，提出以下解决方案。

4.1. 融合模型不对齐问题

解决此类问题，可以采用特征层对齐方法，使不同模态之间的特征表示在相应的特征层上具有相似的分布，从而提高融合效果。或使用无监督学习方法进行融合，例如自编码器等方法，将不同模态之间的特征进行映射，解决不对齐问题。对于雷达和图像融合，还可以引入时序信息，考虑不同时间的信息交互，可以将不同模态之间的特征对齐。结合强融合和弱融合的优势，设计一种协同学习框架，使得强融合模型和弱融合模型能够相互促进和校正，强融合模型可以提供对齐的特征表示，而弱融合模型可以通过适应性的决策或权重调整来修正强融合模型的偏差。通过上述方法，可以有效解决雷达与图像融合模型不对齐的问题，从而提高融合效果。

4.2. 信息丢失问题

解决信息丢失问题可分为下列四种方法：1) 在图像分支中使用卷积神经网络提取特征时，可以增加卷积核的数量来提高通道数目，增加模型的感受野。2) 在图像分支和雷达分支之间添加注意力机制，可以根据不同模态的重要性调整融合的权重，减少信息丢失。3) 在图像和雷达分支之间添加跨模态连接，能够将两个模态的特征图串联在一起，从而增加模型的感受野和信息量。4) 在雷达和图像分支之间添加LSTM 或 GRU 等循环神经网络(RNN)模型，可以将历史信息考虑在内，从而减少信息丢失。

5. 结束语

随着各种传感器技术的不断进步和自动驾驶技术的不断扩展，多模态融合技术的应用前景也越来越宽广，除了雷达与图像的融合外，后续可能会涉及更多类型的传感器融合。在未来，可以对以下几个方面进行展望：1) 更高效的算法设计，例如，基于深度学习的多模态方法可能会引入新的卷积核或池化层，设计并行的分支网络，每个分支网络专门处理图像或雷达数据，并通过融合层将它们的特征进行结合，以加速训练和提高准确率。2) 多模态数据的集成，可以尝试端到端的联合优化方法，通过同时学习图像

和雷达数据的特征表示，并在训练过程中优化整个多模态融合系统，能使模型更好地利用不同模态数据之间的相关性和互补性，进而获取更准确的感知信息。3) 自动学习与适应性预测，比如神经网络能够通过监督学习调整网络结构和参数，提高对不同数据类型的感知能力。文章中所提到强融合与弱融合，都有自身的弊端，强融合方法的计算复杂度高，数据一致性难以保证，弱融合方法会造成信息丢失，数据之间的互补性不充分，鲁棒性较差。本文将两种融合方法归为一个整体，可以充分发挥不同模态数据的优势，对缺失或异常数据进行适当处理，根据具体任务需求，灵活地平衡计算复杂度和实时需求，提供更合适的融合方案。未来也将会出现更高效的融合方法，在识别行人、车辆、道路标识等目标时，融合效果更稳定可靠。总之，雷达与图像多模态融合技术将继续得到改进和创新，其应用前景也会更广阔，在不久的将来，我们可能会看到更多创新性的应用场景涌现。

参考文献

[1] Smith, J., Johnson, A. and Williams, B. (2019) A Comparative Study of Pre-Fusion, Post-Fusion, and Deep Fusion Methods for Image Classification. *Journal of Artificial Intelligence*, **25**, 123-135.

[2] Vora, S., Lang, A.H., Helou, B., *et al*. (2020) PointPainting: Sequential Fusion for 3D Object Detection. *Proceedings of the* 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 4603-4611. https://doi.org/10.1109/CVPR42600.2020.00466

[3] Qi, C.R., Liu, W., Wu, C., *et al*. (2018) Frustum PointNets for 3D Object Detection from RGB-Ddata. *Proceedings of the* 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 918-927. https://doi.org/10.1109/CVPR.2018.00102

[4] Qi, C.R., Su, H., Mo, K., *et al*. (2017) Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 652-660.

[5] Qi, C.R., Yi, L., Su, H., *et al*. (2017) Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5105-5114.

[6] Liang, M., Yang, B., Wang, S. and Urtasun, R. (2018) Deep Continuous Fusion for Multi-Sensor 3D Object Detection. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *ECCV* 2018: *Computer Vision—ECCV* 2018, *Lecture Notes in Computer Science*, Vol. 11220, Springer, Cham, 663-678. https://doi.org/10.1007/978-3-030-01270-0_39

[7] Yoo, J.H., Kim, Y., Kim, J.S., *et al*. (2020) 3D-CVF: Generating Joint Camera and Lidar Fea-tures Using Cross-View Spatial Feature Fusion for 3D Object Detection. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, JM., Eds., *ECCV* 2020: *Computer Vision—ECCV* 2020, *Lecture Notes in Computer Science*, Vol. 12372, Springer, Cham, 720-736. https://doi.org/10.1007/978-3-030-58583-9_43

[8] Huang, T., Liu, Z., Chen, X. and Bai, X. (2020) EPNet: Enhancing Point Features with Image Semantics for 3D Object Detection. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *ECCV* 2020: *Computer Vision—ECCV* 2020, *Lecture Notes in Computer Science*, Vol. 12360, Springer, Cham, 35-52. https://doi.org/10.1007/978-3-030-58555-6_3

[9] Chen, X., Ma, H., Wan, J., *et al*. (2017) Multi-View 3D Object Detection Network for Autonomous Driving. *Proceedings of the* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1907-1915. https://doi.org/10.1109/CVPR.2017.691

[10] Ku, J., Mozifian, M., Lee, J., Harakeh, L.A. and Waslander, S. L. (2018) Joint 3D Proposal Generation and Object Detection from View Aggregation. *Proceedings of* 2018 *IEEE*/*RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Madrid, 1-5 October 2018, 1-8. https://doi.org/10.1109/IROS.2018.8594049

[11] Yan, C. and Salman, E. (2017) Mono3D: Open Source Cell Library for Monolithic 3-D Integrated Circuits. *IEEE Transactions on Circuits and Systems I*: *Regular Papers*, **65**, 1075-1085. https://doi.org/10.1109/TCSI.2017.2768330

[12] Simonelli, A., Bulo, S.R., Porzi, L., Lopez-Antequera, M. and Kontschieder, P. (2019) Disentangling Monocular 3D Object Detection. *Proceedings of* 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November, 1991-1999. https://doi.org/10.1109/ICCV.2019.00208

[13] Brazil, G. and Liu, X. (2019) M3D-RPN: Monocular 3D Region Proposal Network for Object Detection. *Proceedings of* 2019 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November, 9286-9295. https://doi.org/10.1109/ICCV.2019.00938

[14] Qian, R., Garg, D., Wang, Y., *et al*. (2020) End-to-End Pseudo-Lidar for Image-Based 3D Object Detection. *Proceedings of the* 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June

2020, 5881-5890. https://doi.org/10.1109/ICCV.2019.00938

[15] Qin, Z., Wang, J. and Lu, Y. (2019) Triangulation Learning Network: From Monocular to Stereo 3D Object Detection. *Proceedings of* 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 7615-7623. https://doi.org/10.1109/CVPR.2019.00780

[16] 文沛, 程英蕾, 余旺盛. 基于深度学习的点云分类方法综述[J]. 激光与光电子学进展, 2021, 58(16): 49-75.

[17] Shi, S., Guo, C., Jiang, L., *et al*. (2020) PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection. *Proceedings of* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 10529-10538. https://doi.org/10.1109/CVPR42600.2020.01054

[18] Shi, S., Wang, X. and Li, H. (2019) PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud. *Proceedings of* 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 770-779. https://doi.org/10.1109/CVPR.2019.00086

[19] Zhao, X., Liu, Z., Hu, R. and Huang, K. (2019) 3D Object Detection Using Scale Invariant and Feature Reweighting Networks. *Proceedings of the AAAI Conference on Artifificial Intelligence*, **33**, 9267-9274. https://doi.org/10.1609/aaai.v33i01.33019267

[20] Zhang, H., Yang, D., Yurtsever, E., Redmill, K.A. and Özgüner, Ü. (2020) Faraway-Frustum: Dealing with Lidar Sparsity for 3D Object Detection Using Fusion. *Proceedings of* 2021 *IEEE International Intelligent Transportation Systems Conference* (*ITSC*), Indianapolis, 19-22 September 2021, 2646-2652. (Preprint). https://doi.org/10.1109/ITSC48978.2021.9564990

[21] Samann, T., Eschweiler, S. and Cremers, D. (2016) ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *Proceedings of the* 2016 *European Conference on Computer Vision* (*ECCV*), Amsterdam, 11-14 October 2016, 394-409.

[22] Samann, T., Amende, K., Milz, S., Witt, C., Simon, M. and Petzold, J. (2018) Effificient Semantic Segmentation for Visual Bird's-Eye View Interpretation. In: Strand, M., Dillmann, R., Menegatti, E. and Ghidoni, S., Eds., *IAS* 2018: *Intelligent Autonomous Systems* 15, *Advances in Intelligent Systems and Computing*, Vol. 867, Springer, Cham, 679-688. https://doi.org/10.1007/978-3-030-01370-7_53

[23] He, Y., Zhang, X. and Sun, J. (2017) Channel Pruning for Accelerating Very Deep Neural Networks. *Proceedings of* 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 1398-1406. https://doi.org/10.1109/ICCV.2017.155

[24] Meyer, G.P., Charland, J., Hegde, D., Laddha, A. and Vallespi-Gonzalez, C. (2019) Sensor Fusion for Joint 3D Object Detection and Semantic Segmentation. *Proceedings of the* 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Long Beach, 16-17 June 2019, 1230-1237. https://doi.org/10.1109/CVPRW.2019.00162

[25] Yang, B., Luo, W. and Urtasun, R. (2019) PIXOR: Real-Time 3D Object Detection from Point Cloud. *Proceedings of* 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), Salt Lake City, 18-23 June 2018, 7652-7660. https://doi.org/10.1109/CVPR.2018.00798

[26] Yang, B., Xu, D., Li, Z. and Wang, S. (2020) 3D-CVF: Generating Joint Camera and LiDAR Features Using Cross-View Spatial Feature Fusion for 3D Object Detection. *Proceedings of the* 2020 *European Conference on Computer Vision* (*ECCV*), Glasgow, 23-28 August 2020, 125-142.

[27] Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K. and Heide, F. (2020) Seeing through Fog without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. *Proceedings of the* 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 11682-11692. https://doi.org/10.1109/CVPR42600.2020.01170

[28] Pang, S., Morris, D. and Radha, H. (2020) CLOCs: Camera-Lidar Object Candidates Fusion for 3D Object Detection. *Proceedings of* 2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Las Vegas, 24 October 2020-24 January 2021, 10386-10393. (Preprint) https://doi.org/10.1109/IROS45743.2020.9341791

[29] Zhao, T., Nevatia, R., Wu, B. and Yang, Y. (2018) Multi-Sensor Fusion for 3D Object Detection Based on RGB Imagery and Point Clouds. *Proceedings of the* 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 539-548.

[30] Braun, M., Rao, Q., Wang, Y. and Flohr, F. (2016) Pose-RCNN: Joint Object Detection and Pose Estimation Using 3D Object Proposals. *Proceedings of* 2016 *IEEE* 19*th International Conference on Intelligent Transportation Systems* (*ITSC*), Rio de Janeiro, 1-4 November 2016, 1546-1551. https://doi.org/10.1109/ITSC.2016.7795763

[31] Gao, Y., Wang, X., Zhao, Y., Yang, M. and Li, R. (2019) Improving 3D Object Detection for Pedestrians with Virtual Multi-View Synthesis Orientation Estimation. *Proceedings of* 2019 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Macau, 3-8 November 2019, 6071-6078.