# 合成和识别自动驾驶车辆相机雷达数据集中的噪声水平

Mathis Morales<sup>1</sup> and Golnaz Habibi<sup>2</sup>

Abstract—检测和跟踪物体是任何自主导航方法的关键 组成部分。在过去几十年中,使用神经网络在各种数据集上进 行目标检测已取得了有希望的结果。虽然许多方法侧重于性能 指标,但很少有项目致力于提高这些检测和跟踪管道的鲁棒性, 特别是针对传感器故障的情况。本文试图通过为摄像头-雷达自 动驾驶汽车 (AV)数据集创建一个真实的合成数据增强管道来 解决这一问题。我们的目标是准确模拟由于现实世界干扰导致 的传感器故障和数据退化。我们还展示了基于轻量级噪声识别 神经网络的基线结果,该网络在增强的数据集上进行训练和测 试,在 10086 张图像和 2145 个雷达点云中跨越 11 类别的整 体识别准确率为 54.4%。该项目的代码可用 [这里]。

## I. 介绍

为了实现自主导航,车辆需要一种感知周围环境 的方法。例如,对于地面车辆,在复杂环境中与其他 汽车一起行驶时,它必须动态观察其周围环境以避免 与另一辆汽车或行人发生碰撞,或者仅仅是遵守法律 并在停车标志前停止。自动驾驶车辆中最常用的传感 器包括相机、激光雷达(LiDAR)、雷达、超声波传感 器和红外传感器 [1]。

相机至关重要,因为它们提供的语义信息是其他传感器无法匹敌的。正因为如此,许多研究都集中在基于摄像头的目标检测上。当前最先进的方法在 2D 检测任务中实现了高目标识别准确率,例如 co-DETR [2]在 COCO 数据集 [3] 上对 80 个目标类别累计 33 万张 图像的识别准确率达到 66%。

然而,基于地面和空中的车辆在三维环境中移动,并 且被三维物体包围,它们需要能够准确检测出物体在 三个维度上的位置以及它们的类别、方向、尺寸和速 度。所有这些对于正确估计检测到的目标运动轨迹以 避免碰撞或跟踪它都是非常重要的。不幸的是,尽管 相机提供了丰富的语义信息,它们只能输出二维像素 图。这使得仅使用这种传感器来正确估算物体的深度、距离和形状变得极其具有挑战性。此外,3D检测和跟踪的挑战产生了对三维数据集的需求。在自主导航中,nuScenes数据集[4]是一个参考点,因为它为用户提供了一个相机-雷达-LiDAR数据集,专注于城市环境,并提供了针对10个类别的3D检测任务以及其中7个类别的3D跟踪基准。它也被认为是此类首个大规模的数据集之一,包含1000个20秒的场景,收集了超过140万张图像[1]。

如前所述,仅使用摄像头的检测网络在准确放置 三维边界框方面面临更多困难。这一点在 nuScenes 排行榜中得到了体现,排名最高的仅使用摄像头的方 法 [5] 在 338 个方法中排名第 75 位。

为了提高检测性能,需要一个测距传感器,如雷 达或激光雷达。大多数最先进的方法更倾向于使用激 光雷达系统以获得更高的精度 [4], [6], [7]。事实上,如 果我们考虑 nuScenes 排行榜中的情况,前 67 个检测 方法都使用了汽车上的激光雷达,表现优于最先进摄 像头-雷达方法 [8]。

在跟踪任务中,第一个使用摄像头-雷达的方法 [8] 在 128 个方法中排名第 75 位。

然而,激光雷达成本高昂、重量重、体积大且难 以安装,对内存需求高,且难以集成到现有的车辆中。 实时处理还需要强大的计算能力,导致电池消耗显著 增加。

因此,其他方法更倾向于专注于雷达技术。雷达 已经在一个世纪内被用于各种应用 [9]-[11],使其成为 一种众所周知且便宜的传感器,尤其是在与激光雷达 相比时更为明显。它们还提供了无与伦比的距离测量 准确性、速度测量准确性和恶劣天气条件下的鲁棒性, 如雾、雨或雪,在这些条件下,激光雷达系统的表现 较差 [12]。

University of Oklahoma

<sup>&</sup>lt;sup>1</sup>Mathis Morales, School of Electrical and Computer Engineering, University of Oklahoma, Norman, Ok, USA mathis.morales@ou.edu

<sup>&</sup>lt;sup>2</sup>Golnaz Habibi, School of Computer Science, University of Oklahoma, Norman, Ok, USA golnaz@ou.edu

缺点是雷达产生的数据点云非常稀疏,这可能会 使雷达检测变得具有挑战性。然而,这一点可以通过 摄像头的特性得到缓解 [13], [14]。

对于实际应用的自动驾驶汽车,鲁棒性必须是首 要考虑因素,因为当汽车在街道上行驶且人们的生命 受到威胁时,我们无法承受测量结果不可靠。为此, 我们专注于相机-雷达传感器融合,并考虑以下论点: 随着模型在处理经过精心策划、清理和标注的离线数 据集上的检测和跟踪性能越来越强,如果将其部署到 真实世界的汽车中会表现如何?此外,它们将如何应 对由于各种原因导致发送错误数据的故障传感器? 基础假设是检测网络可能能够处理一定量的错误数 据,但最终在某个点上会失败。这个论点适用于任何 神经网络方法,因为它们都有一个失效点。考虑一辆 汽车行驶在街道上的变压器旁的情景。由于变压器产 生大的电磁 (EM) 干扰, 雷达传感器发生故障并开 始向自动驾驶车辆发送错误的数据。当这种情况发生 时,前面的汽车刹车了,但雷达传感器估计其位置比 实际距离远得多,导致自动驾驶汽车与其相撞。另一

种情况是汽车从桥下驶出,相机需要一点时间来调整 曝光水平,在此期间前视摄像头被强光致盲,可能导 致由于未检测到物体而发生碰撞。

鉴于这些情况可能发生,必须设置保障措施以提高在 传感器性能下降时的鲁棒性。本项目旨在探讨如何准 确估计传感器故障(或噪声)级别,据我们所知,这 之前从未在三维相机-雷达数据集上的自动驾驶汽车 中实施过。

该项目的主要贡献是:

- 一条专注于为相机雷达自动驾驶车辆数据集合成 真实世界传感器缺陷的数据增强管道。
- 噪声识别的基础线,使用轻量级模型直接估计退 化程度。这使得我们的方法可以即插即用,因为 任何目标检测都可以使用噪声级别信息来采取行 动。

## II. 相关工作

如 [1] 所述,虽然大多数方法都关注性能指标,但 很少有方法关注传感器的鲁棒性。

RadSegNet [15] 独立地从每个传感器提取信息,但仅 专注于摄像头故障,表明他们的方法即使只使用雷达 数据也能可靠工作。尽管这种方法与我们的类似,但 我们的方法在任何特征提取之前进行,允许根据噪声 水平轻松切换到另一种检测方法。他们还更多关注恶 劣天气条件,这与电磁干扰、热噪声或高反射表面(对 于雷达)引起的传感器退化不同。

另一方面,诸如 ImmFusion [16] 等方法使用 Transformer 网络主动从传感器中选择有用的信息。虽然生成模型和自注意力机制提供了一个有趣的方法来检测和处理传感器缺陷,但它们仍然具有黑箱效果,使得保证鲁棒性变得困难。他们还可以从训练中准确表示数据退化中受益,而我们的方法可以帮助提供这一点。此外,这两种方法都结合了对象检测与鲁棒性改进,导致与现有检测方法不兼容。

## III. 方法论

处理相机和雷达数据时,我们将方法分为两部分, 每种传感器对应一部分。我们创建了两个噪声等级旋 钮,均称为 N<sub>lvl</sub>。这两个旋钮分别处理,并且对每个 传感器都有不同的含义,但都控制着从 0%到 100%的 噪声等级。噪声等级也可以提高至这些值以上以模拟 更高的退化情况,但是它们已经调整为使 100%的噪声 等级代表现实场景中最高的实际扭曲程度。

A. 合成相机退化

我们提出了四种不同的常见图像退化类型:模糊、 低曝光、高曝光和加性噪声。

a)模糊。这种情况发生在相机镜头未能将场景中的 光线正确聚焦到成像平面上。这主要是由于校准 问题,但也可能是由低光条件下的自动对焦故障 或因冲击导致的对齐错误引起的。运动模糊也类 似于聚焦模糊,并且可以在一定程度上通过模拟 来实现。

模糊效果通常通过与高斯核进行卷积来模拟。我 们通过改变高斯核的大小合成不同的噪声水平:

$$k_{size} = 2 \times round(N_{lvl}) + 1 \tag{1}$$

图1显示了不同失真水平下的模糊效果。

b)高/低暴露。这里我们模拟相机进入(低)或离开 (高)隧道或阴影区域的效果。此时光照水平突然 变化,相机自动调整镜头的曝光级别以捕捉更多 或更少的光线。然而,在一两个画面的小瞬间内, 光照水平的变化实际上使相机失明。根据车辆的



Fig. 1: 模糊效果在 0%, 30%, 60%和 100% (最大)失 真水平。

速度和外部光照水平的不同,这种失明效果的程 度也会有所不同。

为了模拟这一点,我们创建一个 3x3 的高斯核:

$$K = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$
(2)

为了创造过度曝光的效果,即过多光线照射到传 感器上,我们将内核乘以一个因子(1+3×N<sub>lvl</sub>)。 要产生相反的效果,我们将内核除以相同的值。当 将内核与原始图像进行卷积时,这会导致像素强 度更高或更低,并且具有平滑效果,从而使生成 的图像更具真实感。图2显示了不同扭曲水平下 的这些效果。

c)加性噪声。可能最常见的退化现象是加性噪声,它 会在图像上产生类似颗粒的效果。这可能是由于 电磁噪声、传感器内部的热噪声或在低光照条件 下(如夜间)导致图像质量严重下降所引起的。虽 然可以通过高斯核滤波来消除这种噪声,但必须 小心,因为图像经过多次滤波后会变得更加模糊。 正确次数的滤波或使用正确的内核也会从对噪声 水平的了解中受益。

这通常通过从正态分布的随机变量  $N(0,\sigma)$  中抽 取  $W \times H$  值来模拟,其中 W 和 H 分别是图像的 宽度和高度,并将生成的噪声图添加到原始图像 的像素上。我们直接设置  $\sigma = N_{lvl}$  来控制噪声的 数量。图 3 显示了不同失真级别下的加性高斯噪 声效果。



Fig. 2: 高(顶部)和低(底部)曝光效果在0%, 30%, 60%和100%(最大)失真水平。



Fig. 3: 正态分布的加性噪声效果在 0%, 30%, 60%和 100% (最大) 失真水平上的影响

### B. 合成雷达降质

对于雷达数据,我们必须首先关注 nuScenes 使用的传感器:大陆 ARS 408-21 长距离雷达传感器 77GHz。传感器在每次扫描时给出的输出数据是一个带有坐标 (x, y, z) 的 3D 点云,一个相对速度向量  $(v_x, v_y)$ ,一个运动补偿后的相对径向速度向量  $(V_{x_{comp}}, V_{y_{comp}})$ ,一个 RCS (雷达截面积)值,以及 其他一些字段提供关于检测的虚警概率、其估计的动力学属性、聚类有效性状态及其多普勒模糊解状态的

信息。

这里我们可以注意到几个项目:

- 补偿后的速度是一个径向的速度,即传感器中心与该点连线上的实际补偿后速度的投影。同样地,速度(vx,vy)是一个不径向速度,因此雷达可能使用跟踪方法对其进行估计,因为雷达只能通过多普勒效应测量径向速度。不幸的是,我们无法确切知道这种估计是如何进行的,因为每个制造商都有自己的方法。
- 同样,大多数杂项字段是使用未知方法内部估算的。
- 雷达提取的每个点实际上是一组点,因此值会被
   平均。这很可能是为了去除异常值并最大化置信
   度。
- 最后,在 nuScenes 中雷达设置的仰角为 0°,因此 点的位置始终为 (x, y, 0.0)。

#### 噪声水平:

在信号处理术语中,噪声通常被称为整体信噪比 (SNR)以分贝(dB)为单位的下降,其定义如下:

$$SNR_{lin} = \frac{P_r}{P_N} \tag{3}$$

其中 P<sub>r</sub> 是接收信号的功率, P<sub>N</sub> 是噪声功率。我们用 SNR<sub>lin</sub> 表示线性尺度下的信噪比,并用 SNR<sub>dB</sub> 表示 以分贝 (dB) 为单位的信噪比。对于雷达数据,噪声水 平 N<sub>lvl</sub> 被定义为 N%的噪声水平相当于下降 N/10dB。 因此:

$$SNR'_{dB} = SNR_{dB} - \frac{N}{10}dB \tag{4}$$

其中 SNR' 是在合成后的新的信噪比。这样,100%的 噪声水平意味着下降-10 dB,而 0%的噪声水平则意 味着 SNR 没有变化,即 0 dB 的减少。在线性尺度上 这可以表示为:

$$SNR'_{lin} = 10^{SNR'_{dB}/10}$$
  
= 10^{SNR\_{dB}/10} .10^{-N/100} (5)  
= SNR\_{lin} .10^{-\frac{N}{100}}

换句话说:

$$\frac{SNR'_{lin}}{SNR_{lin}} = 10^{-\frac{N_{lvl}}{100}} \tag{6}$$

为了真实地合成退化的雷达数据,我们将算法分 为三个步骤:生成幽灵点、假阴性和噪声引起的距离/ 多普勒偏移。

a) 幽灵点生成。在某些城市环境中,由于多路径 效应导致的幽灵点(或假阳性)可能比在 nuScenes 中 出现得更频繁。在某些情况下,如果产生的点的 RCS 值足够高,它们甚至可能会形成一个足够大的簇并被 雷达检测到。为了模拟高度反射环境中的幽灵点更高 出现率,我们在每一帧中随机抽取 0 和一个固定值 (设定为 4)之间的数字来确定将生成多少个点。

对于每个幽灵点,我们从两个均匀随机变量  $U_r(r_{min}, r_{max})$  和  $U_{\theta}(\theta_{min}, \theta_{max})$  中绘制一组范 围/角度  $(r, \theta)$ 。我们将  $r_{min}$  设定为雷达的最小距离 0.2 米, 并将 r<sub>max</sub> 设定为当前帧中最远点, 外加 10 米的余量。其原因是尽管雷达的最大范围是 250 米, 但大多数点都位于观测到的点云中的110米以内。一 个单独且远离其他簇的幽灵点很容易被识别,并不 会对检测网络中点云的完整性构成真正威胁。 $\theta_{min}$ 并且  $\theta_{max}$  取决于所选范围,遵循雷达物理检测范 围的限制,这些范围在短距离(小于10米)、中距 离(小于100米)和长距离(超过100米)之间变 化。如前所述,我们无法知道雷达是如何估计一个 点的相对速度的,所以我们从当前帧中采样一个。 现在我们有了相对速度向量  $\vec{V} = (v_x, v_y)$ ,我们可 以将其投影到视线向量 r 上以创建合成径向相对速 度向量 V<sub>r</sub>。从那里我们重新创建一个合成的补偿径 向速度矢量 V<sub>comp</sub>,已知 nuScenes 车辆的速度为 V<sub>ego</sub>:

$$\vec{V_{comp}} = (V_r - V_{ego}).\vec{V_r}$$
(7)

其中 V<sub>r</sub> 和 V<sub>ego</sub> 分别是径向相对速度矢量和自车速度 矢量的大小。

最后,我们通过从当前帧的已知 RCS 分布中采样来 生成一个合成的 RCS 值,使用有界的半高斯变量 *X*~|*N*(0,1/3)|生成较低值更频繁的同时保留获取较 高值的可能性。从 *X*中抽取的数字然后映射到当前帧 中的有序 RCS 值分布中。虽然幽灵点倾向于低 RCS ,但我们有时会得到一个非常高的 RCS 簇,产生高 RCS 的多路径异常。结果形成的幽灵点相对于其他点 来说会有较高的 RCS 。

我们还随机抽取一个点状态值,这是雷达对一个点的

估计状态,在选定的可能值之间:低 RCS 、高儿童 概率、高 50° 异常概率、高异常概率和无局部最大值。 实际上我们认为传感器有很高的可能性会估计出这个 点有问题,但它可能会错误地标记它是什么。

b) RCS 规则的假阴性。随着噪声功率的增加(信 噪比降低),漏检点的数量也随之增加,因为它们被背 景噪声所掩盖。由于我们能够访问 RCS 值,我们可以 使用雷达方程:

$$P_r = \frac{P_t G_t A_{eff}}{(4\pi)^2 r^4} \sigma \tag{8}$$

其中  $P_r$  是接收到的功率,  $P_t$  是发射的功率,  $G_t$  是天 线增益,  $A_{eff}$  是接收天线的有效孔径(在我们的情况 下, 天线既是发射器也是接收器),  $\sigma$  是点的 RCS 值, 而 r 是其距离。

这意味着,在给定帧中,由于  $P_t$ 、 $G_t$ 和  $A_{eff}$ 是常数 值,接收到的功率 ( $P_r$ )与点到四次方的距离 (r)除 以 RCS 值 ( $\sigma$ )成正比:

$$P_r \propto \frac{\sigma}{r^4} \tag{9}$$

这暗示了:

$$SNR \propto \frac{\sigma}{r^4}$$
 (10)

因此从 6 得出:

$$SNR' \propto \frac{\sigma}{r^4} \times 10^{-N/100}$$
 (11)

基于合理的假设,即给定帧的最低 RCS 值 (用  $\beta_{min}$  表示)是传感器设置的最小检测阈值,我们可以计算 出系数  $\alpha = \frac{\sigma}{r^4} \times 10^{-N/100} + w$ ,其中  $w \sim N(0, \beta_{min})$ 是一个小的随机噪声,代表物理不确定性和 RCS 估 计误差。任何低于  $\beta_{min} = min(\frac{\sigma}{r^4})$ 的  $\alpha$  都将从点云 中移除。 $N_{lvl}$  越高,数据集中丢弃的点就越多。从观 察结果来看,使用较小的  $N_{lvl}$  时移除的点少于 10 个, 在-10dB 噪声水平下可以裁剪掉超过 60 个点。

c) 噪声诱导的偏移。随着噪声水平的增加(信 噪比降低),传感器的测量精度也会提高。事实上, Cramér-Rao 界建议 [17]:

$$acc \propto \frac{1}{\sqrt{SNR}}$$
 (12)

此外,制造商提供的距离、角度和速度测量的精度值 通常是针对高 RCS 目标,在受控环境中的情况。每个 点的实际测量精度将随着信噪比的变化而不同。我们 考虑 SNR ∝ RCS,并使用 12 对每个点的每次测量 重新计算准确度。

从 6,我们得到:

$$acc' \propto \frac{1}{\sqrt{SNR'}} = \frac{1}{\sqrt{10^{-N/100}}}$$
 (13)

最后,我们从正态分布的随机变量  $N(0, acc'_r), N(0, acc'_{\theta}), N(0, acc'_v)$ 中抽取噪声值  $w_r, w_{\theta}, w_v$ 并将其添加到原始的 $r, \theta, V$ 值中。我们还 将噪声传播到了补偿后的径向相对速度 $V_{comp}$ 上。图 4 展示了不同噪声水平对雷达点云在俯视图(自上而 下)投影中的影响,而图 5 则展示了同一点云的 3D 扫描图像。我们发现在较低信噪比时有较少的点,-3 dB 附近出现一些幻影点,并且总体位置偏移在低信 噪比时更大。



Fig. 4: 噪声合成器对雷达点云在鸟瞰图(自上而下) 投影中 0dB、-3dB、-6dB 和-10dB 信噪比的影响。



Fig. 5: 噪声合成器对三维扫描中 0dB、-3dB、-6dB和-10dB 信噪比下的雷达点云的影响。

#### IV. 噪声水平检测的结果

我们创建了一个简单的神经网络来分析合成数据 集。相机分析器是一个7层的U-NET [18] (4人,3 出),利用这种架构的编码/解码方法高效地提取噪声 特征。雷达分析器由三个1D卷积层和两个全连接层 组成。两个网络通过一个单一的全连接层激活以产生 11个输出,对应于从0到100%步长为10%的噪声水 平。

对于相机而言,我们的训练集由 nuScenes 的小数据 集中随机采样的2个场景构成,而验证和测试数据各 自包含另一个场景。请注意,在这里我们没有在相机 训练中使用夜间数据,因为其被视为未标记的不利场 景。在相机和雷达训练中,我们也仅使用"关键帧" (每2赫兹取一次)作为标注,其他的是插值出来的。

在对图像应用各层级的不同类型噪声后,得到的 训练集大小为 19680 张图像,加上 10086 张用于验证, 以及 10086 张用于测试。

对于雷达点云,我们进行相同的处理过程,但使 用了3个训练场景而不是2个。得到的训练集大小为 6501个云,而验证和测试分别为2255和2145。

训练是在单个 NVIDIA GeForce RTX 3060 上进 行的。

我们的结果见表 I, 网络的混淆矩阵见图 6 和图 7。

Data type	Accuracy (%)	$TP\uparrow$	$\mathrm{FP}\uparrow$	Labels
Camera Radar	62.19 20.79	$6272 \\ 446$	$3814 \\ 1699$	$10086 \\ 2145$
Total	54.4	6718	5513	12341

TABLE I: 测试数据结果关于我们噪声识别网络在相机和雷达数据上的表现。

## V. 结论和进一步工作

总之,本文提出了一种针对自动驾驶车辆摄像头 和雷达数据集的真实数据生成器,重点在于模拟由于 内部或外部噪声导致的传感器故障和数据退化。我们 还展示了可以训练一个轻量级的噪声识别模型来量化



Fig. 6: 我们的图像噪声识别网络的混淆矩阵

		_									
0	169	115	95	69	69	64	69	59	64	63	59
10	- 5	10	14	16	8	4	0	0	0	0	0
20	- 1	6	10	3	1	2	0	1	1	0	0
ж	- 8	41	43	47	42	37	11	5	3	1	0
40	3	7	10	24	17	17	23	15	4	6	5
itue label	- 0	з	7	16	20	19	14	13	12	4	3
6	. 3	4	3	7	15	22	35	29	20	16	9
70	- 3	2	3	з	7	10	10	18	15	11	9
8		2	4	2	6	4	12	14	22	14	13
90	1	1	1	з	5	8	13	17	20	28	26
100	2	4	5	s	5	8	8	24	34	52	n
	0	\$0	10	Ŷ	*0	¢,	÷	10	÷	ą.	-190

Fig. 7: 混淆矩阵对于我们雷达噪声识别网络

两个传感器中的数据降质情况。我们的方法允许在任 何其他现有方法上即插即用,并且可以在特定噪声水 平下切换到另一个模型,或者尝试利用这一知识自动 过滤掉噪声。未来的工作包括增加诸如天气模拟(针 对摄像头数据)或干扰器(针对雷达数据)等不利条 件。此外,我们的噪声识别模型是一个基线模型,未来 的工作将使其更轻量、更快和更准确。在这方面,我 们目前仅限于使用 nuScenes 提供的 1000 个场景中的 5 个场景进行训练。

### 致谢

该项目的实现离不开俄克拉荷马大学电气与计算 机工程学院的纳森 · 古德曼博士及其在雷达数据综合 方面的帮助。

#### References

 K. Shi, S. He, Z. Shi, A. Chen, Z. Xiong, J. Chen, and J. Luo, "Radar and camera fusion for object detection and tracking: A comprehensive survey," 2024. [Online]. Available: https://arxiv.org/abs/2410.19872 1, 2

- [2] Z. Zong, G. Song, and Y. Liu, "Detrs with collaborative hybrid assignments training," 2023. [Online]. Available: https: //arxiv.org/abs/2211.12860 1
- T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015. [Online]. Available: https://arxiv.org/abs/1405.0312 1
- [4] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," 2019. 1
- [5] X. Lin, Z. Pei, T. Lin, L. Huang, and Z. Su, "Sparse4d v3: Advancing end-to-end 3d detection and tracking," 2023. [Online]. Available: https://arxiv.org/abs/2311.11722 1
- [6] X. Weng, J. Wang, D. Held, and K. Kitani, "3d multi-object tracking: A baseline and new evaluation metrics," 2020. 1
- T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3d object detection and tracking," 2021. [Online]. Available: https://arxiv.org/abs/2006.11275 1
- [8] Z. Lin, Z. Liu, Y. Wang, L. Zhang, and C. Zhu, "Rcbevdet++: Toward high-accuracy radar-camera fusion 3d perception network," 2024. [Online]. Available: https: //arxiv.org/abs/2409.04979 1
- [9] P. Lang, X. Fu, M. Martorella, J. Dong, R. Qin, X. Meng, and M. Xie, "A comprehensive survey of machine learning applied to radar signal processing," 2020. [Online]. Available: https://arxiv.org/abs/2009.13702 1
- [10] J. A. S. William L. Melvin, Principles of Modern Radar. Scitech, 2005. 1
- [11] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017. 1
- [12] M. Dreissig, D. Scheuble, F. Piewak, and J. Boedecker, "Survey on lidar perception in adverse weather conditions," 2023. 1
- [13] Y. Long, A. Kumar, D. Morris, X. Liu, M. Castro, and P. Chakravarty, "Radiant: Radar-image association network for 3d object detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 1808–1816, Jun. 2023. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/ view/25270 2
- [14] Y. Kim, J. Shin, S. Kim, I.-J. Lee, J. W. Choi, and D. Kum, "Crn: Camera radar net for accurate, robust, efficient 3d perception," 2023. 2
- [15] K. Bansal, K. Rungta, and D. Bharadia, "Radsegnet: A reliable approach to radar camera fusion," 2022. [Online]. Available: https://arxiv.org/abs/2208.03849 2
- [16] A. Chen, X. Wang, K. Shi, S. Zhu, B. Fang, Y. Chen, J. Chen, Y. Huo, and Q. Ye, "Immfusion: Robust mmwave-rgb fusion for 3d human body reconstruction in all weather conditions," in 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, May 2023, p. 2752 – 2758. [Online]. Available: http://dx.doi.org/10.1109/ICRA48891.2023.10161428 2
- [17] J. A. H. Mark A. Richards, William L. Melvin, "Precision and accuracy in radar measurements," in *Principles of Modern Radar*. Scitech, 2005. 5

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597 6