



(12) 发明专利申请

(10) 申请公布号 CN 117911975 A

(43) 申请公布日 2024. 04. 19

(21) 申请号 202310367655.1

B60W 60/00 (2020.01)

(22) 申请日 2023.04.07

G06V 10/26 (2022.01)

G06V 20/40 (2022.01)

(30) 优先权数据

10-2022-0133839 2022.10.18 KR

(71) 申请人 现代自动车株式会社

地址 韩国首尔

申请人 起亚株式会社

梨花女子大学校产学协力团

(72) 发明人 朴炯郁 申章浩 赵悻瑛 姜制远

李订炅

(74) 专利代理机构 北京路浩知识产权代理有限公司

公司 11002

专利代理师 崔龙铉 赵赫

(51) Int. Cl.

G06V 20/58 (2022.01)

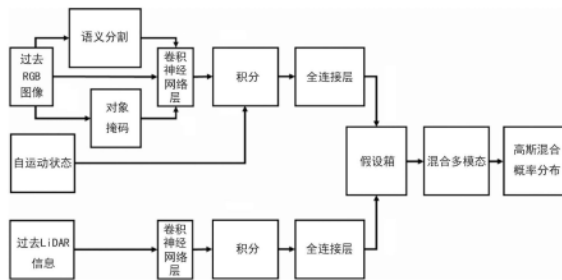
权利要求书1页 说明书8页 附图17页

(54) 发明名称

在车辆中预测对象在未来时间点的位置的方法

(57) 摘要

本公开涉及一种预测对象在未来时间点的位置以控制车辆的预测操作的方法。在在车辆中预测对象在未来时间点的位置的方法中,可以提取通过车辆的摄像机获取的当前时间点和当前时间点之前的多个时间点的视频图像信息作为语义分割图像。可以提取对每个视频图像中存在的对象的属性和位置信息进行成像的掩码图像。可以通过接收当前时间点和当前时间点之前的时间点的视频图像、多个语义分割图像、多个掩码图像和车辆的自运动信息,通过深度学习导出关于对象在未来时间点的位置的多个假设,并且将多个假设计算为高斯混合概率分布,来预测对象的位置分布。



1. 一种在车辆中预测对象在未来时间点的位置的方法,包括:
 - 由处理器提取通过所述车辆的摄像机获取的视频图像;
 - 由所述处理器提取所述视频图像作为语义分割图像;
 - 由所述处理器提取对所述视频图像中存在的对象的属性和位置信息进行成像的掩码图像;
 - 由所述处理器混合所述视频图像、所述语义分割图像、所述掩码图像以及所述车辆的自运动信息;
 - 由所述处理器预测所述对象的位置分布以导出所述对象在未来时间点的预测位置的多个假设;
 - 由所述处理器执行利用关于通过预测所述对象的位置分布导出的多个假设的学习数据的拟合;以及
 - 由所述处理器生成混合模型。
2. 根据权利要求1所述的方法,其中,所述视频图像信息包括通过提取并拼接通过所述车辆的所述摄像机获取的两个或更多个视频图像信息而获得的宽视点图像。
3. 根据权利要求2所述的方法,其中:
 - 所述宽视点图像是RGB二维图像即RGB 2D图像,并且
 - 所述方法包括基于自我中心视点利用合成所述RGB 2D图像和LiDAR信息的多视点来预测路线。
4. 根据权利要求3所述的方法,其中,通过混合基于所述视频图像、所述语义分割图像和所述掩码图像的RGB 2D模型和基于所述LiDAR信息的LiDAR模型的输出值来生成所述混合模型。
5. 根据权利要求4所述的方法,进一步包括利用所述混合模型来生成高斯混合概率分布。
6. 根据权利要求5所述的方法,其中,预测所述对象的位置分布包括利用深度学习注意力机制合成来自所述视频图像的最终向量和来自所述LiDAR信息的最终向量。
7. 根据权利要求3所述的方法,其中,生成所述混合模型包括通过混合所述多个假设和基于所述LiDAR信息的LiDAR模型的输出值来生成所述混合模型。
8. 根据权利要求1所述的方法,其中,所述车辆的所述自运动信息包括对应于当前时间点 t 和未来时间点($t + \Delta t$)的信息。
9. 根据权利要求1所述的方法,其中,针对当前时间点 t 和多个过去时间点提取所述视频图像、所述语义分割图像和所述掩码图像。
10. 根据权利要求1所述的方法,进一步包括,在提取通过所述车辆的所述摄像机获取的所述视频图像之前预测所述对象的位置,其中预测所述对象的位置包括从在通过所述车辆的所述摄像机获取的、在当前时间点 t 提取的视频图像中导出多个假设。
11. 根据权利要求10所述的方法,其中,混合所述视频图像、所述语义分割图像、所述掩码图像和所述车辆的所述自运动信息进一步包括混合所述多个假设。
12. 根据权利要求10所述的方法,其中,预测所述对象的位置包括从通过所述车辆的所述摄像机获取的、在所述当前时间点 t 提取的所述视频图像,在所述当前时间点 t 提取的所述语义分割图像以及所述车辆的所述自运动信息中的一个或多个导出所述多个假设。

在车辆中预测对象在未来时间点的位置的方法

技术领域

[0001] 本公开涉及一种预测对象在未来时间点的位置以控制车辆的预测操作的方法。

背景技术

[0002] 随着车辆的自主续航里程逐渐增加,车辆在紧急情况下的安全要求也越来越高。为了满足更高的安全要求,复杂和多种突发情况下基于智能图像信号分析的风险预防和预测技术得以应用。可以通过利用诸如图像和3D信息的各种传感器数据对周围环境中的其他车辆和行人进行抢占式预测(preemptive prediction),提高车辆的有效响应和稳定性以防止事故发生。因此,提出了各种用于车辆中未来对象预测的方法。

[0003] 在一种这样的方法中,基于先验地图信息和路线信息的多模态预测利用先验地图来预测周围车辆的未来路线。先前的研究已经展示了利用详细的先验地图进行预测的深度神经网络模型,该详细的先验地图包括行驶车道并明确表达诸如通过交叉路口的有效路线的道路规则。然而,由于实际上不可能明确地表达所有区域的所有道路规则,所以提出了基于生成器的生成对抗网络(GAN)框架。鉴别器判断预测轨迹是否遵循道路规则,并且生成器预测遵循预测轨迹的轨迹。

[0004] 提出的另一种预测方法利用自我中心视点,其中自我中心视点是指在车辆的当前位置从驾驶员的视点的图像信息。当使用单视点RGB摄像机获取自我中心视点时,存在驾驶员的视点具有局部且狭窄的视角的问题。另外,由于指示车辆的行进方向的自运动,因此存在视点从自我中心视点随时间改变的问题。为了解决该问题,提出了从自我中心视点预测其他车辆和行人的未来位置的方法。预测问题通过学习对象在未来时间可能出现的位置的先验(prior)来解决。不是获得诸如地图信息的外部信息,而是利用从当前图像分析的信息来学习在图像中检测到的对象的类别可能存在的位置的先验。另外,建立学习模型以计算在下一场景中对象可能突然出现的位置的概率分布。

[0005] 另一个研究开发了一种利用来自自我中心视点的RGB图像来预测行人的多路径的方法,如图1所示。该研究通过接收行人的过去路线和大小以及RGB图像来预测路线。当考虑行人的来自自我中心视点的过去路线时,由此产生的未来路线分布包括多条路线。因此,由于多行人路线预测,多模态预测会受到多对一映射的影响。利用条件变分自动编码器(CVAE)解决了该问题。由于CVAE模型对高维输出空间的分布进行建模,因此提出了用于将CVAE模型与循环神经网络(RNN)编码器-解码器结构耦合的网络。

[0006] 现在参照图2,在利用路线信息和道路信息的多模态方法中,当确定未来路线时,可以存在各种路线。除了车辆的确定之外,还可以考虑到周围情况和对象来预测路线。为了减少可能的路线搜索情况的数量,预测了大概的未来最后位置。基于大概的未来最后位置,生成路线,然后通过对可能的未来路线的假设选择路线。

[0007] 现在参照图3,提出了利用光探测和测距(LiDAR)信息的基于点柱结构的数据处理方法。点柱结构是指在将三维点云信息转换成二维伪图像的过程中获得的结构。在传统研究中,像图像中的像素一样,点云信息被划分成称为体素的单元,并且体素的特征被用作三

维卷积神经网络 (CNN) 的输入。所提出的方法通过利用被转换成填充符 (filler) 或体素的信息来处理连续的点云数据并且通过仅计算具有足够信息的体素,解决了在每个体素中稀疏分布的点云数据的问题。

[0008] 现在参照图4,另一个研究开发了用于基于LiDAR传感器检测对象并且基于结果预测对象的未来位置的深度神经网络结构。不是通过检测在当前帧中的位置并且预测未来位置的方法,而是通过未来对象检测方法回顾问题,并且从每个轨迹的起始点预测未来对象的位置。通过以多对一的方式连接未来和现在的位置来估计多种未来。

[0009] 现在参照图5,提出的另一种方法利用神经网络模型来预测先前给定的未来LiDAR帧。通过用于预测点云之间运动的FlowNet3D模型生成关于LiDAR数据的时间之间运动的信息。

[0010] 背景技术部分的内容旨在促进对本公开的背景的理解,并且可能包括本公开所属领域的普通技术人员先前未知的事项。

发明内容

[0011] 本公开旨在提供一种在自动驾驶车辆中利用多模态合成来预测对象在未来时间点的位置的方法。在实施例中,可以通过混合各种模态的未来预测来执行未来预测。在示例中,可以通过合成在每个模态的未来预测中可以从车辆获取以判断当前事故并且基于判断的当前事故预测未来运动的各种信息(多模态)来执行未来预测。

[0012] 根据本公开的实施例,一种在自动驾驶车辆中预测对象的未来位置的方法包括:由处理器提取通过自动驾驶车辆的摄像机获取的视频图像;由处理器提取视频图像作为语义分割图像;由处理器提取对视频图像中存在的对象的属性和位置信息进行成像的掩码图像;由处理器混合视频图像、语义分割图像、掩码图像以及车辆的自运动信息;由处理器预测对象的位置分布,以导出对象在未来时间点的预测位置的多个假设;由处理器执行利用关于通过预测对象的位置分布导出的多个假设的学习数据的拟合;由处理器生成混合模型。

[0013] 在一方面,视频图像信息可以是通过提取并拼接通过车辆的摄像机获取的两个或多个视频图像信息而获得的宽视点图像。

[0014] 在一方面,宽视点图像可以是RGB二维 (2D) 图像,该方法可以包括基于自我中心视点利用合成RGB 2D图像和LiDAR信息的多视点来预测路线。

[0015] 在一方面,该方法可以进一步包括通过混合基于视频图像、语义分割图像以及掩码图像的RGB 2D模型和基于LiDAR信息的LiDAR模型的输出值来生成混合模型。

[0016] 在一方面,该方法可进一步包括利用混合模型来生成高斯混合概率分布。

[0017] 在一方面,预测对象的位置分布可以包括利用深度学习-注意力机制混合(合成)来自视频图像的最终向量和来自LiDAR信息的最终向量。

[0018] 在一方面,该方法可以包括通过混合多个假设和基于LiDAR信息的LiDAR模型的输出值来生成混合模型。

[0019] 在一方面,车辆的自运动信息可以包括对应于当前时间点 t 和未来时间点($t + \Delta t$)的信息。

[0020] 在一方面,可以针对当前时间点 t 和多个过去时间点提取混合的视频图像、语义分

割图像和掩码图像。

[0021] 在一方面,该方法可以进一步包括:在提取通过车辆的摄像机获取的视频图像之前预测对象的位置,其中预测对象的位置包括从通过车辆的摄像机获取的、在当前时间点 t 提取的视频图像中导出多个假设。

[0022] 在一方面,混合视频图像、语义分割图像、掩码图像以及车辆的自运动信息进一步包括混合多个假设。

[0023] 在一方面,预测对象的位置可以包括导出多个假设,该多个假设包括通过行驶中的车辆的摄像机获取的、在当前时间点 t 提取的一个或多个视频图像,在当前时间点 t 提取的语义分割图像以及车辆的自运动信息。

[0024] 本公开的各种实施例基于对象未来预测网络。本公开的实施例i) 利用由LiDAR提供的多视点执行多视点合成,ii) 利用合成的多视点执行对象未来预测,以及iii) 利用图像信息和LiDAR信息执行多模态预测。在一些实施例中,当混合多个模态时,可以利用诸如添加多模态的结构和注意力结构的方法。此外,可以执行利用概率分布的最终合成方法。因此,可以克服受限视角的图像信息的限制,并且通过提出的模态之间的注意力结构,可以有效地混合多模态。

[0025] 因此,可以有效地预测对象的未来位置。

[0026] 另外,可以在自动行驶期间的意外情况下增加自动驾驶车辆的安全性,并且预测和预防风险。

[0027] 如上所述,利用本公开的实施例,可以通过执行抢占式预测提高自动驾驶车辆的有效响应和稳定性。

[0028] 更具体地,可以利用传感器信号来实现更有效的响应,该传感器信号除了用于全面地检测具有狭窄视角的传统的自我中心视点摄像机的图像之外,还通过多视图合成全面地检测具有宽视角的图像和全向环境。在各种实施例中,自动驾驶车辆可以利用可以连续360度监控车辆周围的情况的各种类型的高性能摄像机传感器、LiDAR技术、雷达等来准确地检测状况并支持快速响应。实施例可以利用传感器硬件和电子装置通过混合和耦合智能软件来产生协同作用。

附图说明

[0029] 从以下结合附图进行的详细描述,本公开的上述和其他目的、特征以及优点可以更加明显,其中:

[0030] 图1至图5示出在传统的自动驾驶车辆中预测未来的方法;

[0031] 图6示意性示出根据本公开的实施例的预测方法;

[0032] 图7是示出根据本公开的实施例的预测对象的位置和位置分布的方法的视图;

[0033] 图8是更详细地示出根据本公开的实施例的预测对象的位置的方法的视图;

[0034] 图9是更详细地示出根据本公开的实施例的预测对象的位置分布的方法的示图;

[0035] 图10示意性地示出根据本公开的实施例的利用自动驾驶车辆的多模态合成来预测对象在未来时间点的位置的方法;

[0036] 图11至图13示出混合多模态的过程;

[0037] 图14和图15示出利用多视图图像预测对象在未来时间点的位置的方法;以及

[0038] 图16和图17示出当混合多模态时使用的注意力结构。

具体实施方式

[0039] 为了充分地理解本公开、本公开的操作优点以及实施本公开所实现的目的,应当参照示出本公开的实施例的附图以及在附图中描述的内容。

[0040] 在本公开的实施例的描述中,将减少或者省略可能不必要地模糊本公开的主旨的公知技术的描述或者重复描述。

[0041] 图6示意性示出根据本公开的实施例的方法。

[0042] 图7是示出根据本公开的实施例的预测对象的位置和位置分布的方法的视图。图8是更详细地示出根据本公开的实施例的预测对象的位置的方法的视图。图9是更详细地示出根据本公开的实施例的预测对象的位置分布的方法的视图。

[0043] 图10示意性地示出根据本公开的实施例的利用自动驾驶车辆的多模态合成预测对象在未来时间点的位置的方法。

[0044] 在下文中,参照图6至图10描述根据本公开的实施例的利用自动驾驶车辆的多模态合成预测对象在未来时间点的位置的方法。

[0045] 本公开提供通过结合包括从车辆前方获取的二维RGB图像信息的从RGB图像中提取的图像的各种分析信息、随着时间的变化信息、LiDAR信息等,来预测在车辆的自我中心视点中在未来时间出现的对象的运动和出现概率的方法。

[0046] 预测对象的未来位置的传统方法包括利用鸟瞰图或自我中心视点的图像信息的方法。但是,在利用鸟瞰图时,存在处理所需的数据变得过于庞大的问题。另外,自我中心视点的图像信息的视角受限。

[0047] 为了克服该问题,本公开提出通过拼接RGB二维(2D)信息的图像来利用宽视角的方法,并且提出将具有宽视角的RGB 2D信息和LiDAR信息相混合以用于预测对象的未来运动的多视点合成。

[0048] 传统方法利用基于作为先验信息准备的地图执行的多视点合成来预测路线。另一方面,在本公开中,如图6中所示,利用基于宽视角的自我中心视点执行的多视点合成来预测对象的未来运动。另外,本公开描述当混合LiDAR信息和RGB 2D信息时,混合两种不同模态的方法。在传统系统中,使用了利用单模态预测未来位置或者当每个模态融合时简单地相加最终预测结果概率的方法。另一方面,本公开的实施例可以通过在一个深度学习结构中的注意力计算合成LiDAR和RGB信息来预测未来。

[0049] 如图6所示,在根据本公开的实施例的预测未来的方法中,可以首先利用多模态之一来预测未来。作为一种类型的模态,可以利用通过LiDAR传感器检测的数据(例如通过摄像机收集的2D RGB图像信息)、合成的多视点图像以及LiDAR和雷达信息。在每个模态的未来预测中,可以判断当前事件并且可以基于所判断的当前事件来预测未来运动。最后,可通过混合多模态的未来预测来执行未来预测。

[0050] 如上所述,本公开的各种实施例基于对象未来预测网络。本公开的实施例可以i)利用由LiDAR传感器提供的多视点执行多视点合成,ii)利用合成的多视点执行对象未来预测,以及iii)利用图像信息和LiDAR信息执行多模态预测。在一些实施例中,当多模态混合时,可以利用诸如添加多模态的结构和注意力结构的方法。此外,可以执行利用概率分布的

最终合成方法。因此,可以克服受限视角的图像信息的限制,并且通过提出的模态之间的注意力结构,可以有效地混合多模态。因此,能够有效地预测对象的未来位置。

[0051] 参照图7至9更详细地描述根据实施例的最终合成方法,并且图10是根据实施例的最终合成方法的示意图。

[0052] 本文中公开一种通过自动行驶预测系统装置利用从行驶车辆的视点获取的视频(自我中心视点)来预测周围车辆和行人的未来位置的方法。

[0053] 在实施例中,对象位置预测网络首先估计考虑对象的背景和属性每个对象在考虑车辆的自运动的未来时间点可以存在的若干候选位置。

[0054] 换言之,在收集来自过去时间点的图像信息、车辆的自运动信息以及来自过去时间点的LiDAR信息并对其进行一定的处理之后,可以基于学习并利用多视图来预测特定对象在未来时间点的位置。

[0055] 在实施例中,对象位置分布预测网络可以通过预测每个对象的未来位置的概率分布来最终估计未来位置。在示例中,两个网络基于ResNet-50。在本公开中,可以通过在对象位置分布预测网络中附加利用多模态信息来提高对象未来预测的准确性。

[0056] 图7是根据实施例的用于更加详细地描述预测对象的位置的操作的示图。参照图7,可以提取通过行驶车辆的摄像机获取的RGB视频图像(S11)。然后,可以从提取的RGB视频图像信息中提取语义分割图像(S12)。可以估计并提取通过删除诸如车辆和行人的动态对象而获得的静态语义分割图像(S13)。这是为了理解背景的属性,然后通过识别背景的属性 and 对象的属性之间的关系来预测对象的位置(S10)。例如,具有车辆属性的对象位于道路上,而具有行人属性的对象位于人行道上。在该过程中,可以通过自动行驶预测系统装置来收集车辆的自运动(S14),并且可以基于所收集的自运动来预测对象的未来位置,并且可以导出多个预测位置的假设(S20)。例如,可以导出总共20个预测位置的假设(S20)。

[0057] 现在参考图8中示出的预测对象的位置分布的操作,对象位置分布预测网络可以接收,例如0.3秒之前(或者诸如0.15秒的另一时间之前)以及当前时间点的视频图像信息帧(S31),来预测对象的例如0.3秒之后的未来位置。在实施例中,可以根据技术目标来调整过去输入和未来输出时间的间隔。在示例中,对象位置分布预测网络可以接收三个过去RGB图像、三个语义分割图像(S32)、通过对对象的属性信息和对象的位置进行成像而获得的三个掩码图像(S33)、车辆的自运动信息(S34)以及从预测对象的位置的操作(S10)输出的二十个假设(S20),并且可以导出二十个新假设(S31)。然后可以通过使新假设通过拟合网络(S40)来计算高斯混合概率分布(S50)。

[0058] 图11至图13示出根据实施例的在利用多模态预测来预测对象位置的方法中的混合多模态的过程。图14和图15示出根据实施例的利用多视点图像来预测对象在未来时间点的位置的方法。

[0059] 传统的研究专注于基于鸟瞰图的合成多视点的方向。参照图11,本公开提出一种基于通过拼接两个或更多的单一自我中心视点图像生成的具有宽视角的自我中心视点来合成LiDAR信息和RGB 2D图像信息来生成自我中心视点图像的方法。例如,可以通过拼接左前方图像、前方图像和右前方图像来生成一个宽RGB 2D图像(宽视点图像)。在通过使生成的图像通过ImageNet而提取特征之后,可以将该特征扭曲(warp)为距离视点图像。另外,LiDAR信息被配置为距离视点图像。在示例中,可以使LiDAR信息通过LidarRVNet,然后可以

与之前生成的距离视点图像执行距离视点融合。这在下面用于预测路线。

[0060] 图12示出根据实施例的上述输入因子和网络配置的示例。在该结构中,利用两个或更多个单视点图像来生成多视点图像,并利用多视点图像来预测对象的未来。如图12所示,可以通过利用左前方图像、前方图像和右前方图像合成具有宽视点的图像和LiDAR信息来生成多视点图像。在一些示例中,各种视点还可以用于合成不同视角。在合成具有宽视角的图像之后,可以利用预测未来对象的传统方法来执行未来预测。

[0061] 如上所述,在本公开中,除了2D图像之外的LiDAR信息用于解决2D图像视角受限的问题。因此,实现了由图像和LiDAR信息组成的多模态预测,并且由此预测了未来对象。在各种实施例中,可以选择如图3至图5中所示的处理LiDAR信息的方法。通过混合通过处理LiDAR信息获得的输出值和通过利用RGB图像获得的输出值可以获得最终结果值。在各种示例中,可以使用如图12所示的多视点RGB网络,或者可以使用利用单RGB图像的网络。

[0062] 图13示出根据实施例的用于通过利用点柱的柱特征网处理LiDAR信息的网络的示例。可以通过点柱从由(x,y,z)组成的点云生成具有通道C、宽度W和高度H的伪图像,并且可以通过卷积层生成最终向量。图13示出利用网络处理单RGB图像的示例。在示例中,可以通过在利用多个RGB图像最终输出的向量中累积LiDAR模型的信息并且通过全连接层来预测最终的未来位置。可选地,如在传统方法中一样,可以使用当每个模态混合时简单地相加最终预测结果概率的结构。

[0063] 接下来,参考将多模态的结果混合到图14中的概率分布中的方法,当混合RGB模态和LiDAR模态的输出值时,可以利用混合模型(S50)生成高斯混合概率分布(S60)。对于从每个模态输出的N个未来位置的假设,可以利用两个全连接层(fc)来生成和估计软分配。通过估计的软分配,可以计算k个高斯概率分布模型的参数(π_k, μ_k, σ_k)。最后,可以利用最终计算的均值、方差和分布权重来生成高斯混合概率分布。

[0064] RGB 2D图像已经在先前的研究中被广泛地用于估计对象的未来位置。本公开描述一种通过融合若干单视点图像来预测多视点情形下的未来对象的方法。在示例中,利用三个过去图像来估计未来位置。此外,在各种实施例中,除了RGB图像之外,以下因素的一个或多个组合可以用作输入。

[0065] (1) 可以使用过去帧的语义分割图像信息。由于在场景中出现的可能性和运动速度可能根据每个对象的对象信息和属性而不同,所以语义分割图像信息可以反映对象信息。

[0066] (2) 类似地,通过在过去对象的边界框的位置处生成指定对象的类别的掩码图像,可以将表达对象的信息和属性的信息用作输入。

[0067] (3) 可以使用通过合成RGB 2D图像和LiDAR信息获得的距离视点图像。另外,可以将上述语义分割和掩码图像中的每一个的若干过去图像串接(concatenate)并用作输入。

[0068] (4) 可以将车辆的计划自运动用作输入。

[0069] (5) 可以将诸如LiDAR信息或点柱的三维LiDAR特征信息用作输入。

[0070] (6) 可以使用诸如车道的道路信息。

[0071] 在各种实施例中,使用(1)至(6)作为输入的网络通过全连接层生成M个(例如,20个)边界框假设。对于学习网络,可以利用EWTA来学习网络。EWTA损耗可以表示为以下表达式:

$$[0072] \quad L_{FLN} = \sum_{k=1}^K w_k I(h_k, \hat{y}), l=L2 \text{ norm}$$

[0073] 其中 $\hat{y}$ 是指地面真值 (ground-truth), h 是指估计的假设。可以计算地面真值和每个假设的L2范数, 其中 w 指权重且为0或1。可以更新K个优胜者 (最佳假设), 其中K个权重为1, 并且 $(M-K)$ 个权重为0。在示例中, K从M到1持续减半, 并且学习了该网络使得可以输出各种假设。

[0074] 在示例中, 当混合多模态时可以使用注意力结构。图16示意性示出根据实施例的自注意力概念。图17示出根据实施例的计算实际的基于矩阵的自注意力的过程。

[0075] 在深度学习技术中, 注意力机制是通过计算当前模型学习的特征与其自身或其他特征相关的程度来执行学习以增强计算的特征中的有用特征的技术。从自然语言处理到图像处理的深度学习模型都使用注意力机制。

[0076] 下面描述从自然语言处理开发的自注意力技术。下面的主要示例利用自然语言处理进行描述, 但是可以被应用于各种传感器数据, 例如图像和视频。

[0077] 将初始输入到编码器的词的每个词向量乘以权重矩阵以获得查询向量、键向量和值向量。

[0078] 在各种实施例中, 可以获得给定查询的所有键的注意力分数中的每一个, 然后将相应的相似度用作权重并且反映到与键映射的每个值。最后, 可以通过对相似度所反映的所有值进行加权求和来获得注意力值。

[0079] 参照图16, 示例附图示出从输入 “I am a student.” 中计算词 “student” 的自我注意力的方法。通过将 “student” 的词向量乘以权重矩阵 W^Q, W^K, W^V , 可以分别获得词 “student” 的查询向量、键向量和值向量 $Q_{\text{student}}, K_{\text{student}}, V_{\text{student}}$ 。然后, 可以通过利用两个向量的点积执行缩放的点积注意力通过表达式 $\text{scorefunction}(q, k) = q \cdot k / \sqrt{d_k}$ 获得注意力分数。可以通过获得注意力分数 Q_{student} 和所有键向量, 然后获得注意力分布, 然后利用相应的相似度作为权重来反映与映射诸如键的每个值的相似度, 并且对相似度所反映的值向量加权, 获得最终的注意力值。

[0080] 在本公开的实施例中, 在利用注意力有效地混合多模态的方法中, 可以针对每个通道堆叠多模态或通过添加维度向量来处理多模态。在实施例中, 利用注意力结构混合模态以混合2D图像和LiDAR信息。

[0081] 将在时间 t 利用2D RGB信息输出的一组向量定义为 X_t^R , 并且将从LiDAR信息生成的一组向量定义为 X_t^L 。 $x^R \in X_t^R$ 和 $x^L \in X_t^L$ 两者具有相同大小的维度。可以通过分别选择以下方法来设置查询、键和值。

[0082] 查询被设置为 $x^R \in X_t^R$, 键和值被设置为 $x^L \in X_t^L$, 以及

[0083] 将查询设置为 $x^L \in X_t^L$, 并且将键和值设置为 $x^R \in X_t^R$ 。

[0084] 通过针对 Ω_q 中的第K个键表示为 $A(\cdot, \cdot)$ 的注意力权重, 输出向量可以被表示为以下表达式。

$$[0085] \quad y_q = \sum_{k \in \Omega_q} A(\Phi_Q(x_q), \Phi_K(x_k)) \cdot \Phi_V(x_k)$$

[0086] Φ_Q, Φ_K, Φ_V 是用于处理查询、键和值的值的线性层。通过利用注意力权重计算查询键值, 获得具有具有最高相似度的值的权重总和。利用通过上述表达式获得的输出值来执行最终预测。

[0087] 图15示出根据实施例的用于多模态预测的注意力结构。更具体地, 图15示出根据实施例的利用深度学习注意力结构A通过从LiDAR模态输出的向量L和从RGB模态输出的向量R之间的相似度混合的网络的示例。

[0088] 如本领域所公知的, 从功能块、单元、部件和/或模块的角度来看, 可以在附图中示出一些示例形式。本领域的普通技术人员应当理解的是, 这样的块、单元、部件和/或模块在物理上由诸如逻辑电路、分立组件、处理器、硬连线电路、存储器装置和布线连接的电子(或光学)电路来实现。当块、单元、部件和/或模块由处理器或其他类似硬件实施时, 块、单元和模块可以通过软件(例如, 代码)编程和控制, 以执行本公开中论述的不同功能。此外, 块、单元、部件和/或模块中的每一个可以由专用硬件或者用于执行一些功能的专用硬件和用于执行另一功能的处理器(例如, 一个或多个编程处理器和相关电路)的组合来实施。

[0089] 上面参照附图描述了本公开, 但本公开不限于所描述的实施例, 并且对于本领域普通技术人员应当显而易见的是, 在不背离本公开思想和范围的情况下, 可以对本公开进行各种修改和改变。因此, 修改的示例或改变的示例应当被包括在本公开的权利要求中, 并且本公开的范围应当基于所附权利要求书来解释。

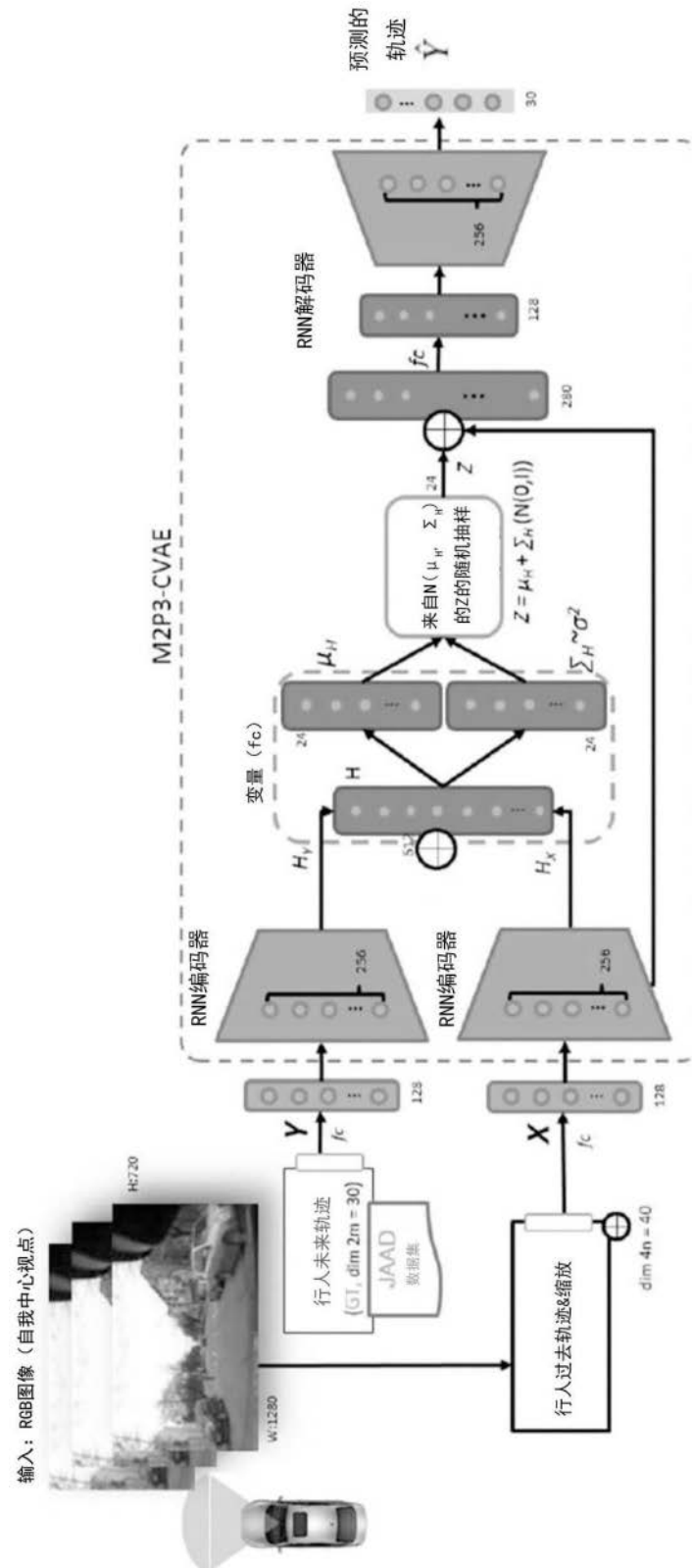


图1

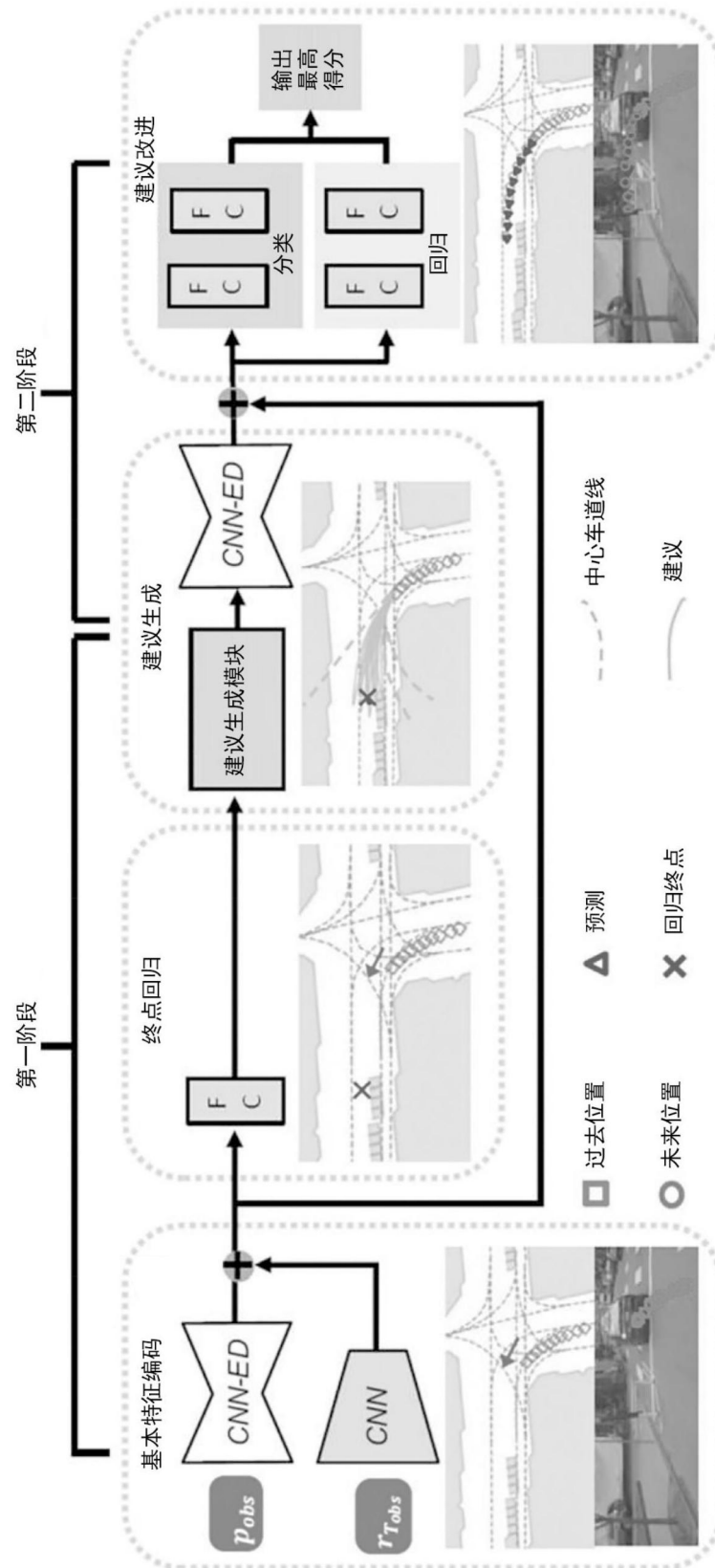


图2

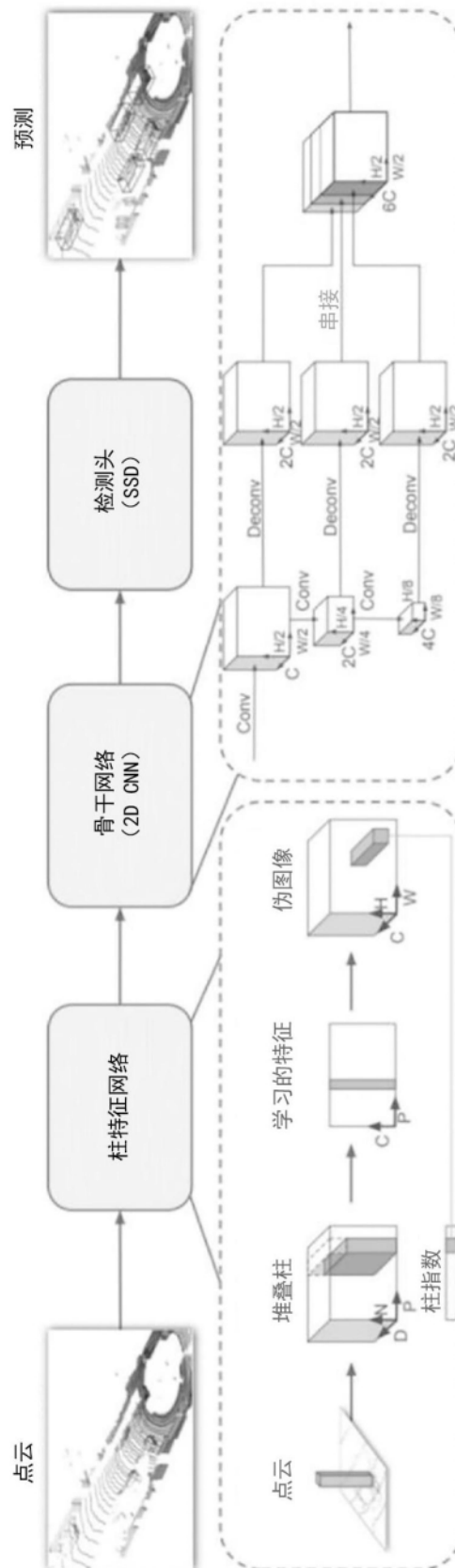


图3

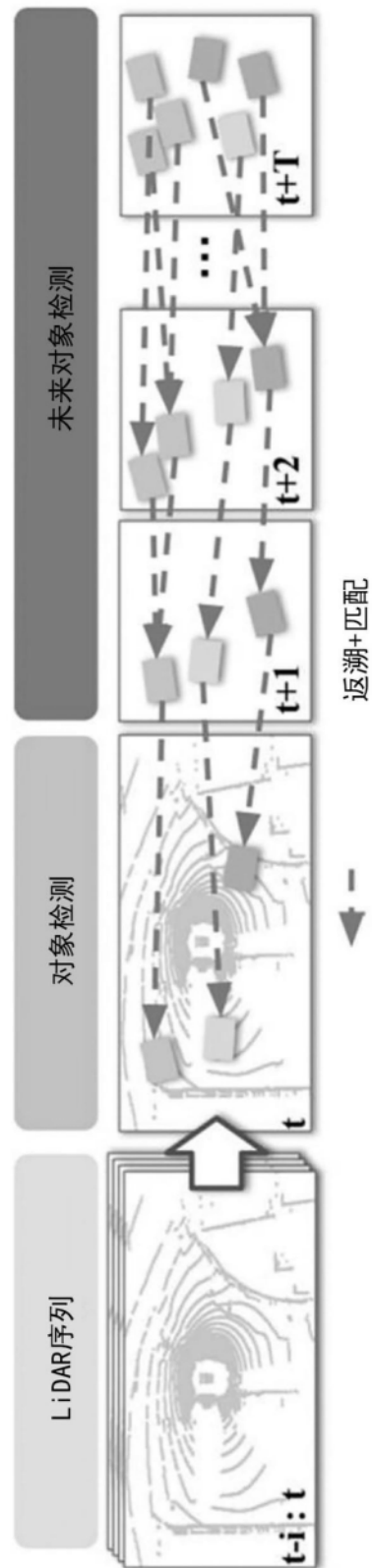


图4

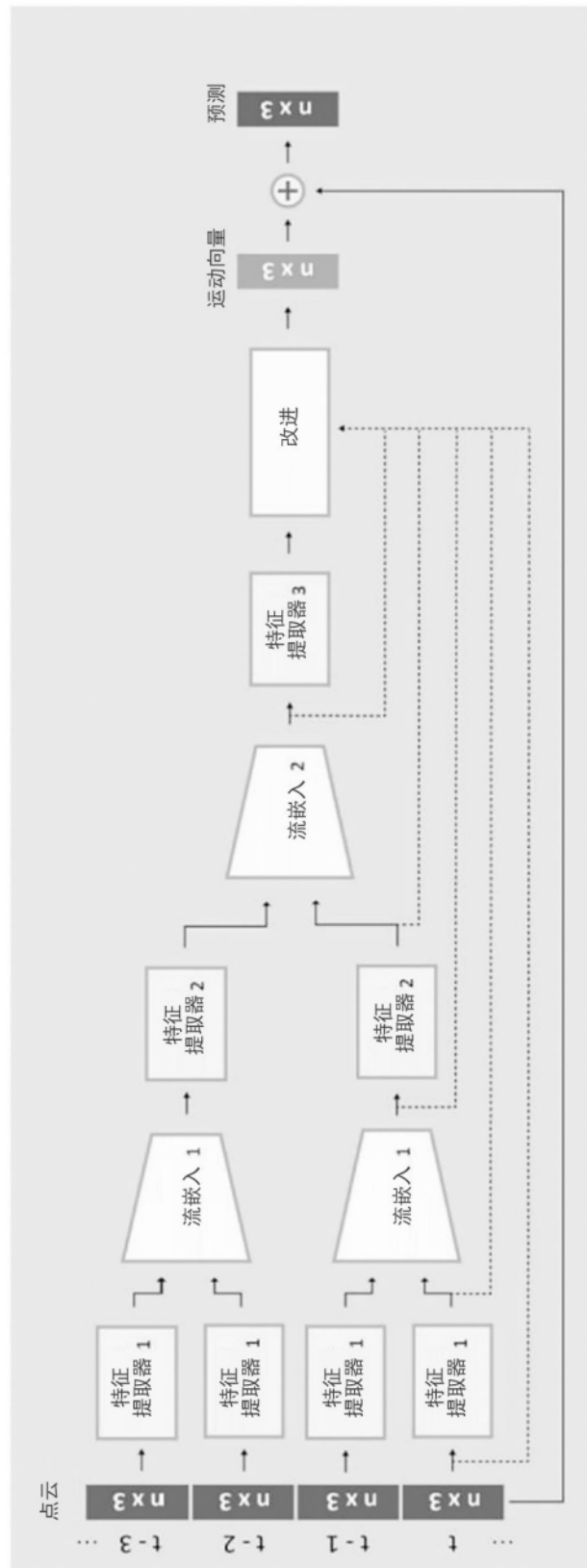


图5

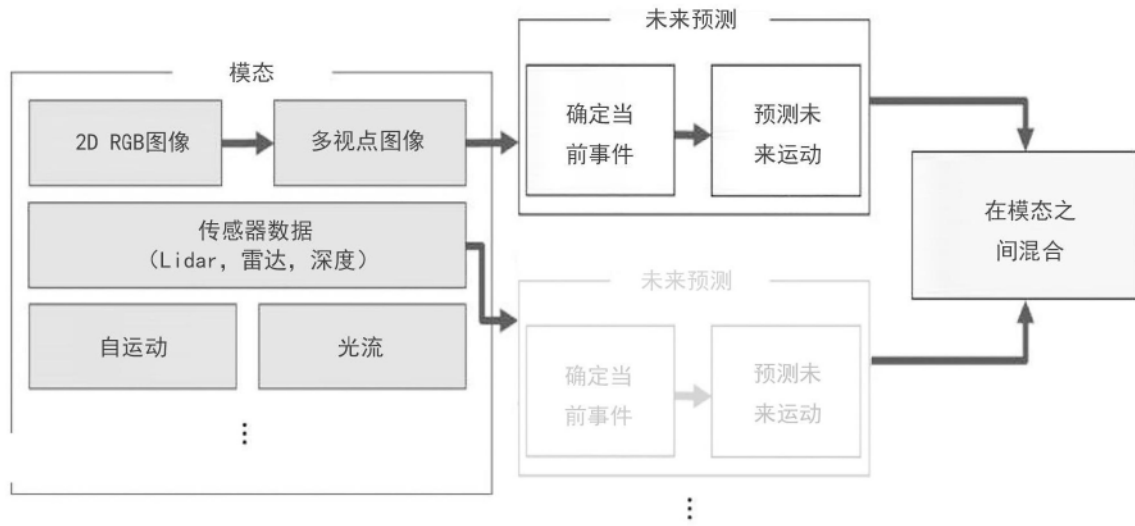


图6

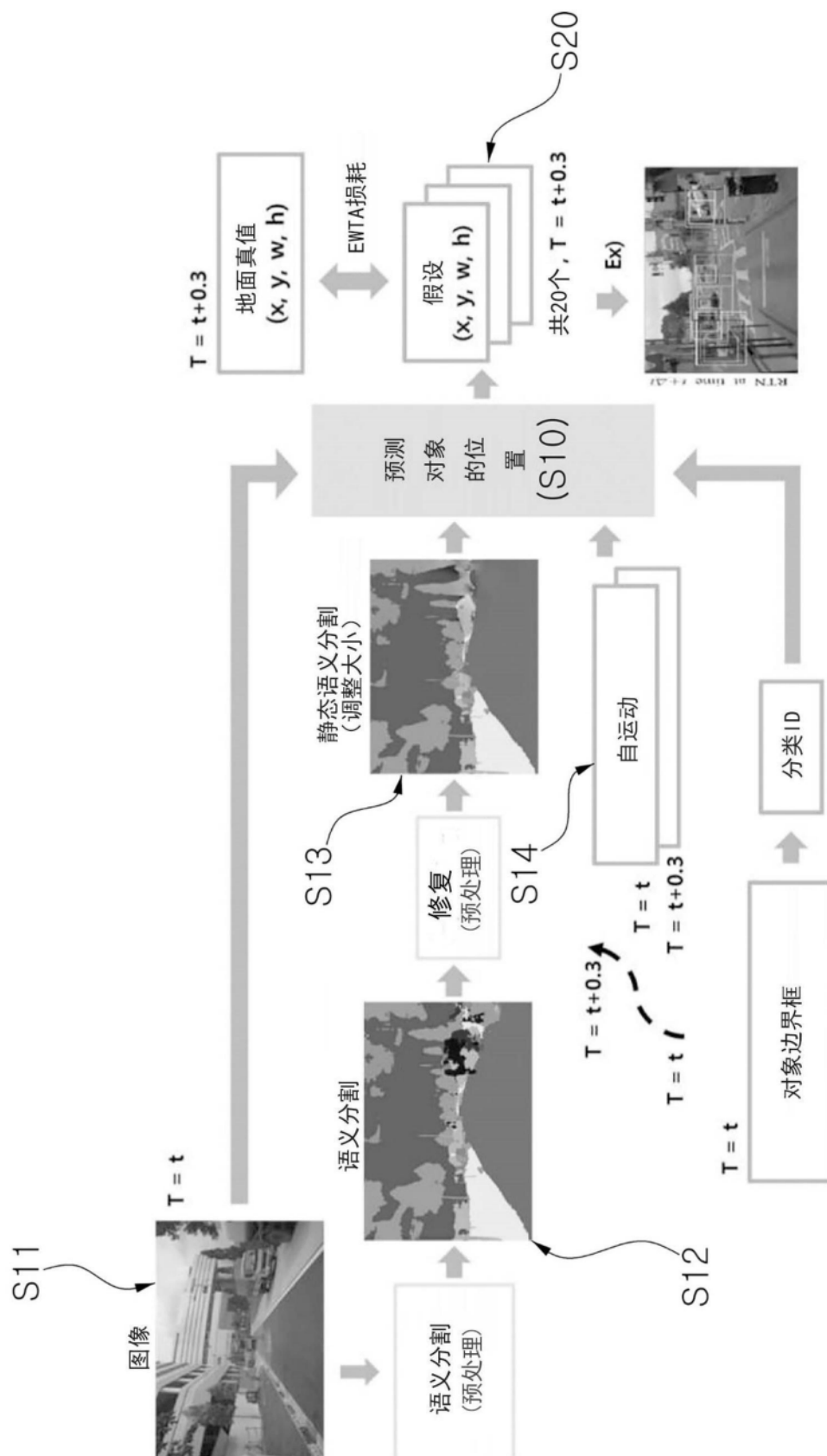


图8

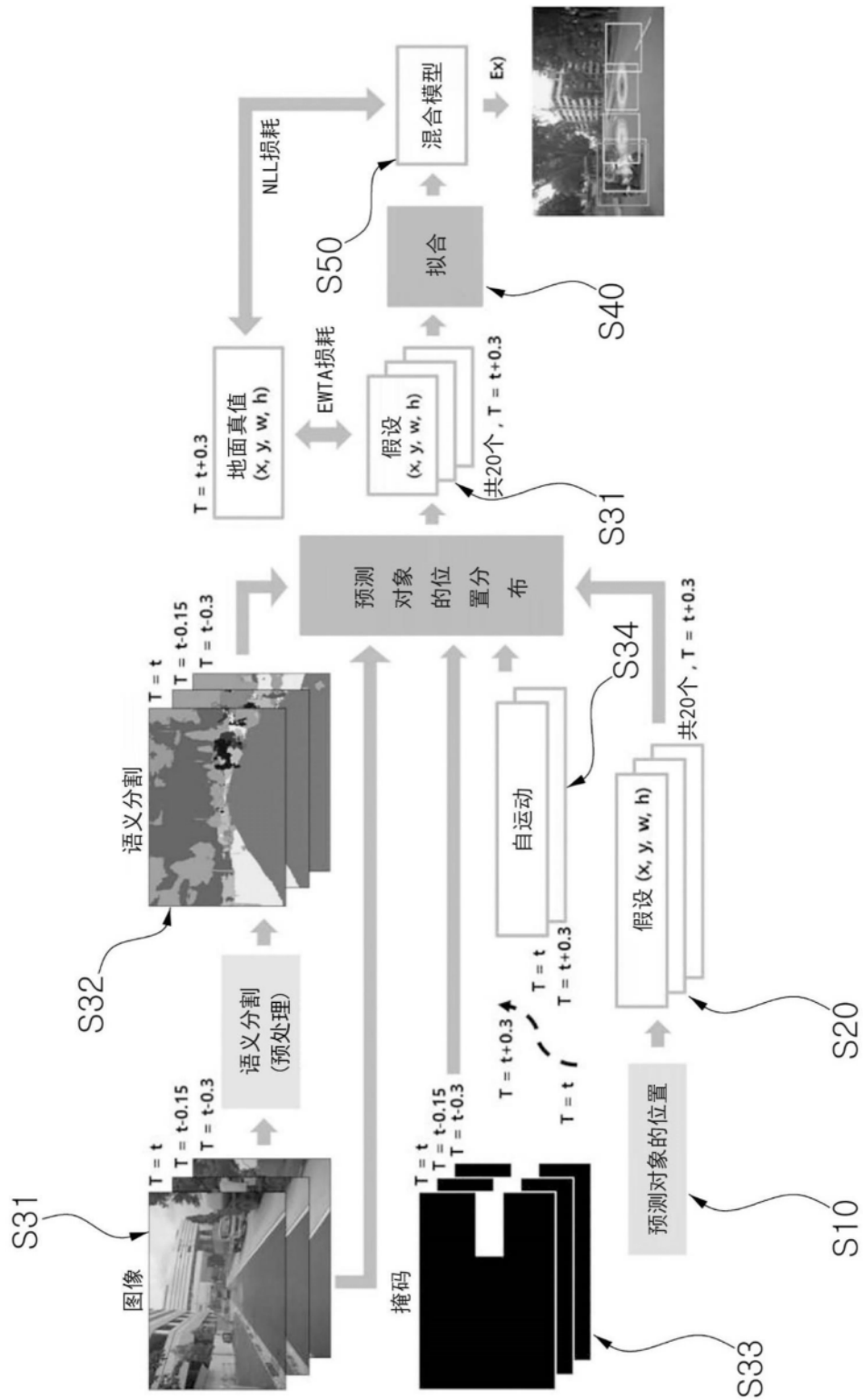


图9

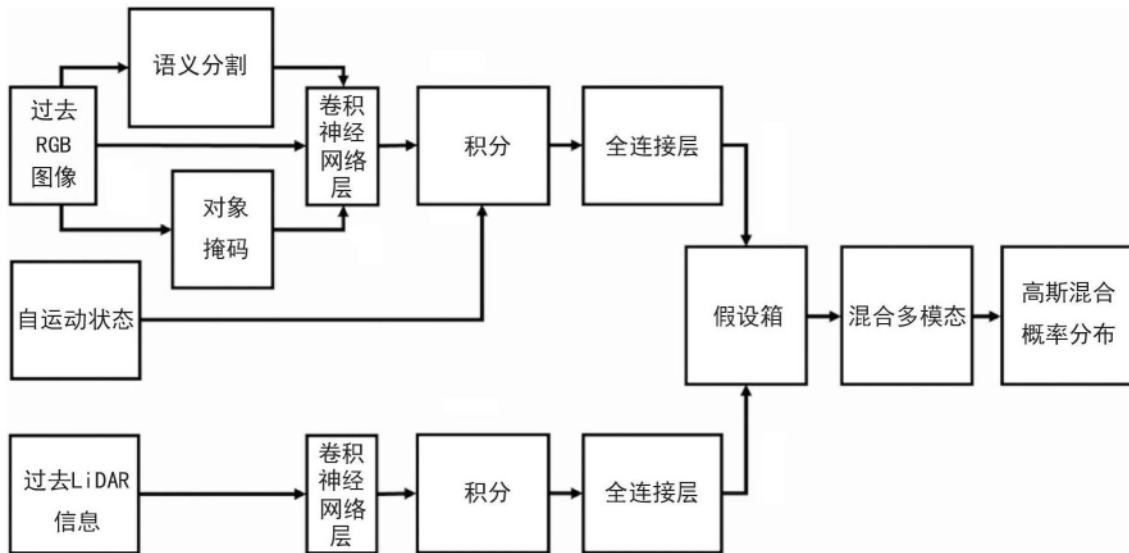


图10

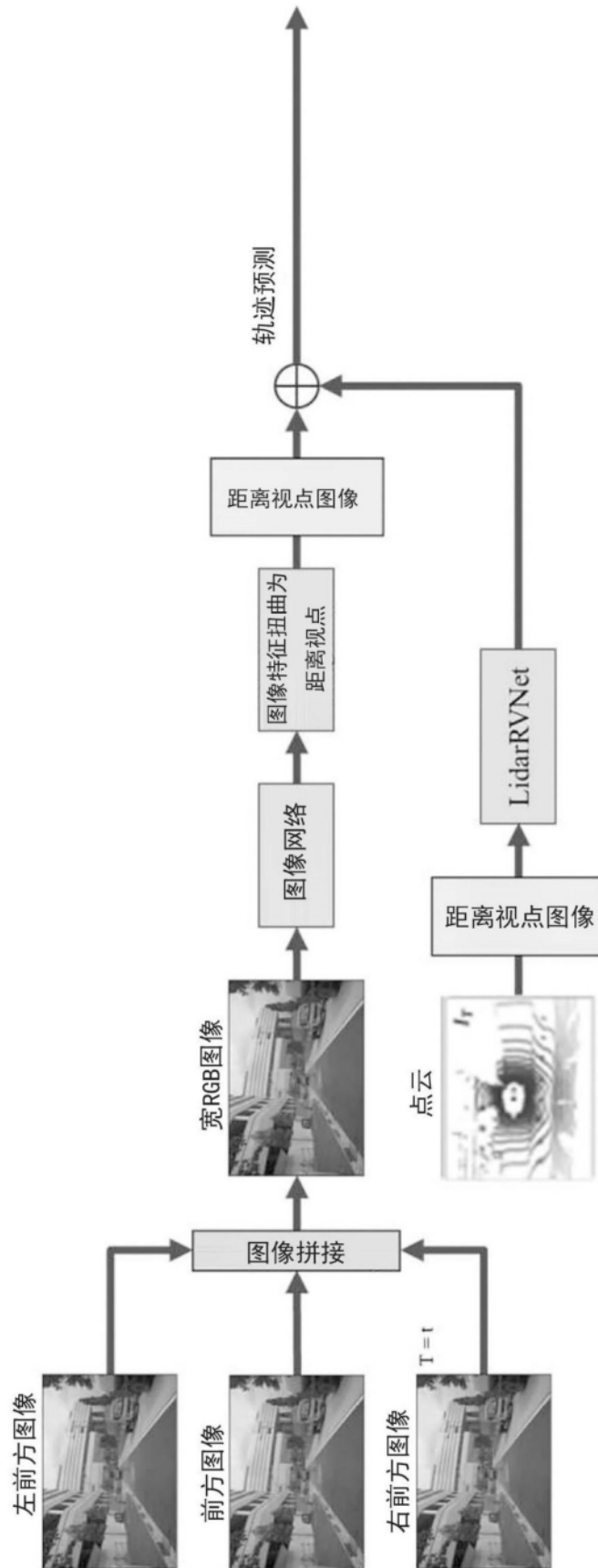


图11

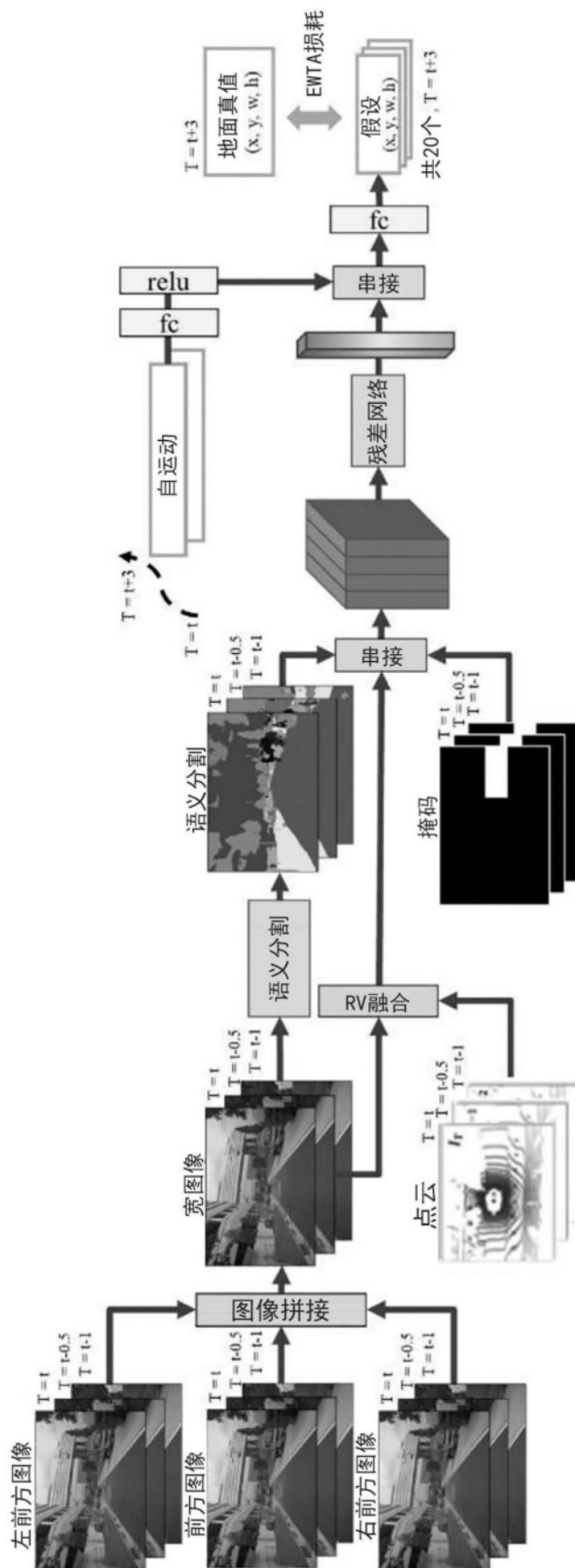


图12

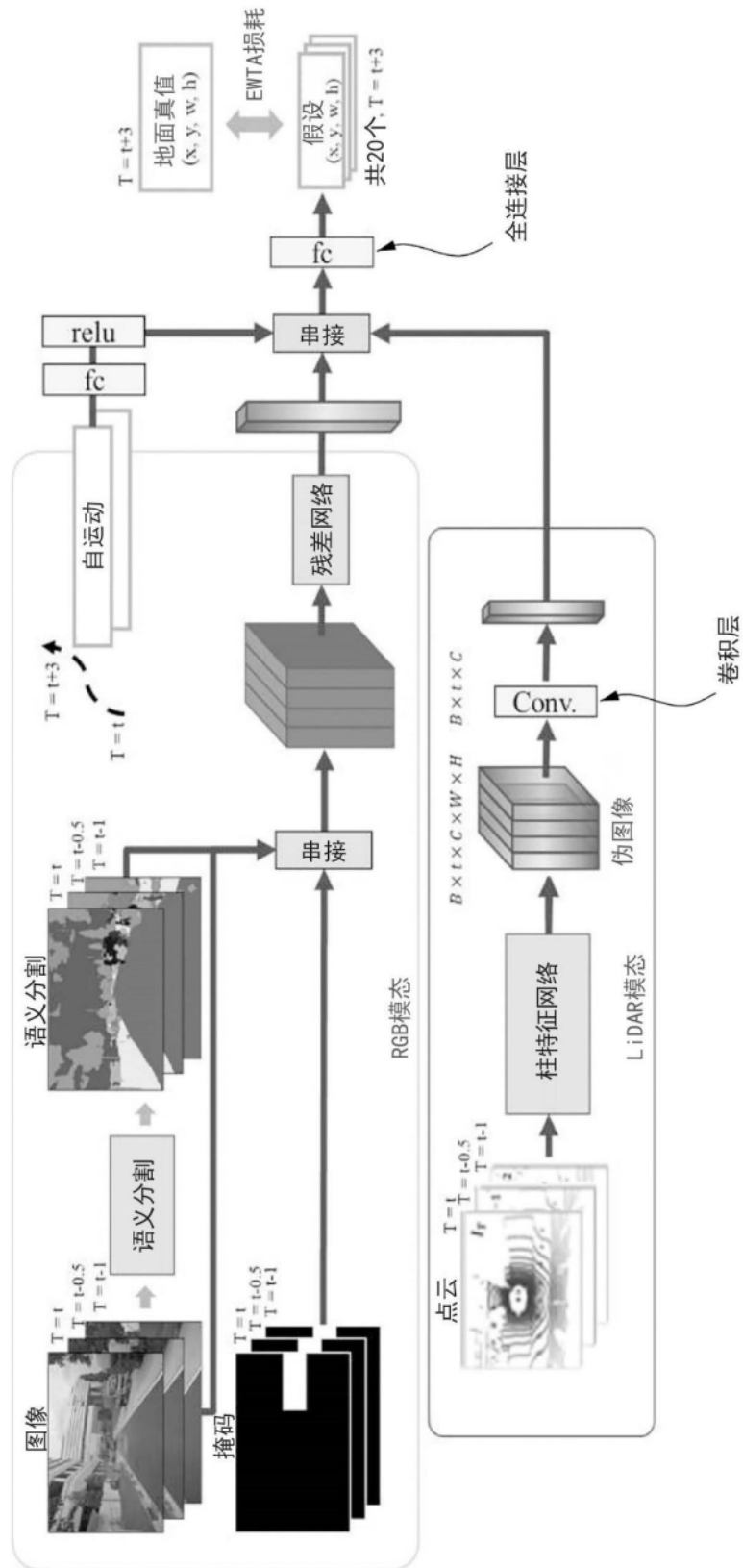


图13

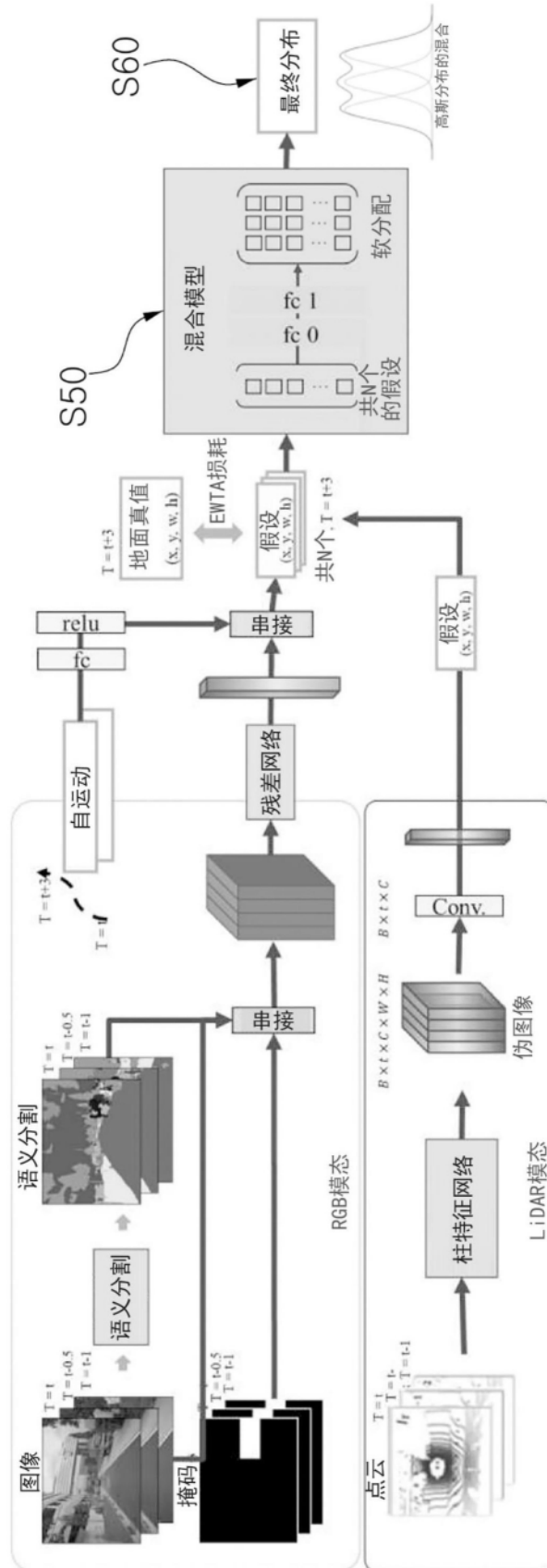


图14

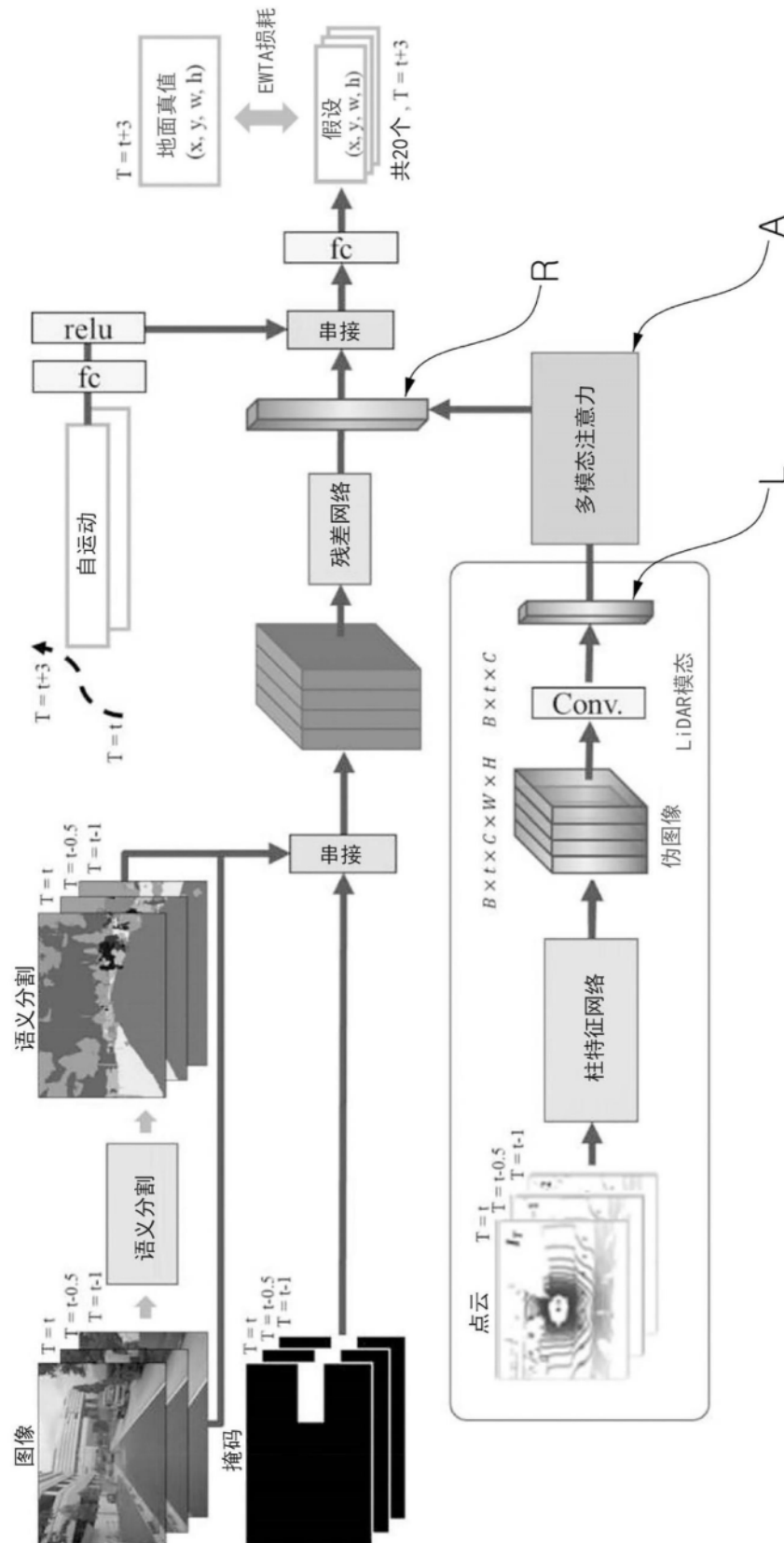


图15

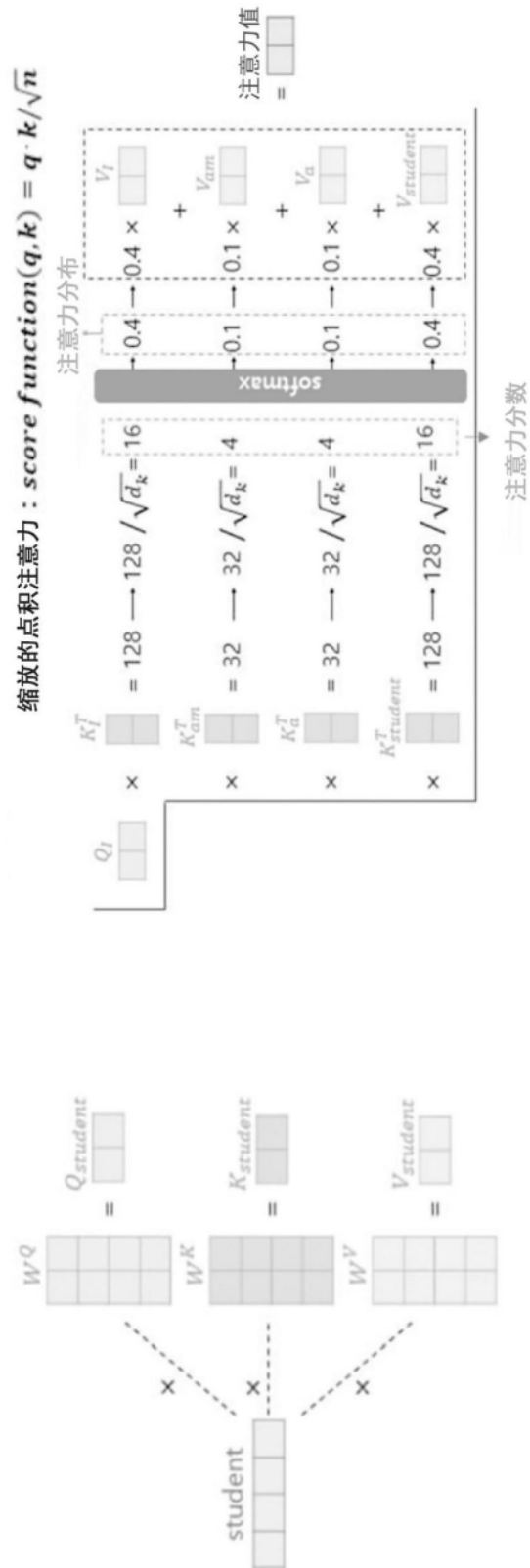


图16

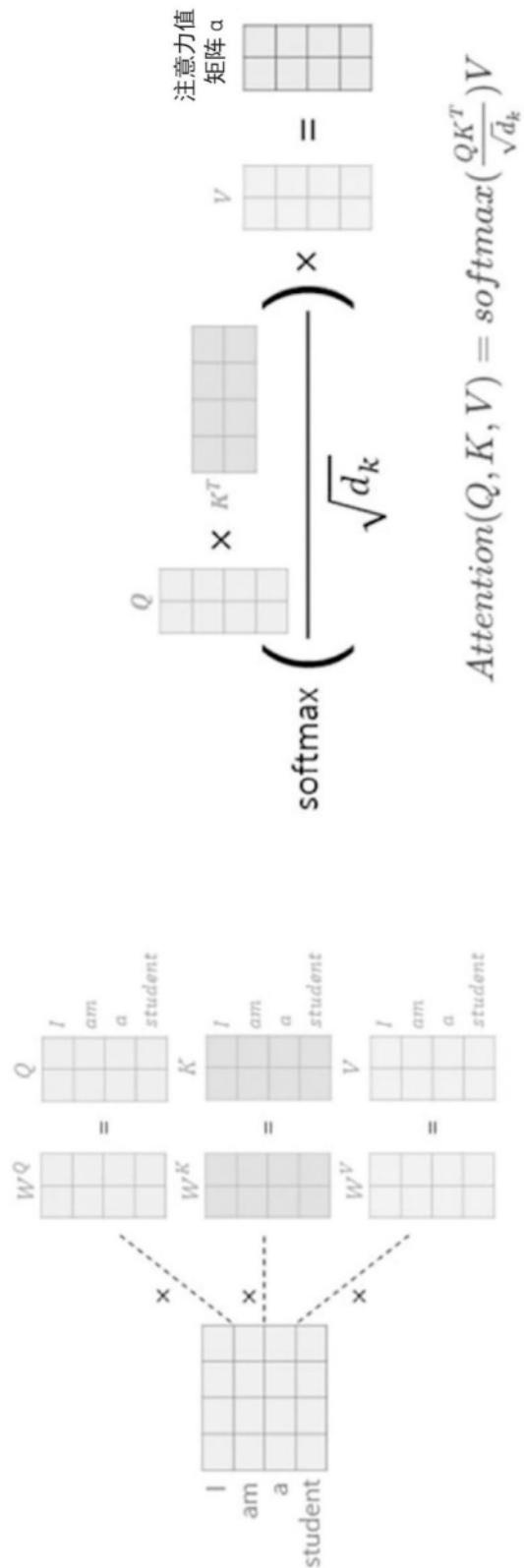


图17