无人驾驶汽车协同感知信息传输负载优化技术

吕品[1),2),3)] 李凯[1)] 许嘉[1),2),3)]* 李陶深[1),3)] 陈宁江[1),3)]

[1)](广西大学 计算机与电子信息学院, 南宁 530004)

[2)](广西多媒体通信与网络技术重点实验室, 南宁 530004)

[3)](广西高校并行分布式计算技术重点实验室, 南宁 530004)

摘 要 无人驾驶近年来成为了学术界和工业界的研究热点，无人驾驶汽车的环境感知则是其中的重要基础。仅通过提升无人驾驶汽车上的传感器数量和精度并不能完全消除车辆的感知盲区，因此无人驾驶汽车与路边基础设施进行协同环境感知越来越受到关注。通过车路协同感知，无人驾驶汽车的感知范围能够得到有效扩展，有助于消除感知盲区，对于提升无人驾驶的安全性具有重要意义。在各类环境感知信息中，摄像头拍摄的视频占有最重要的地位。然而，视频帧所包含的数据量较大，传输每个视频帧会导致网络负载过重，传输延迟增大，影响环境感知信息的时效性。本文提出了一种视频感知数据的传输负载优化方法，主要思想是路边基础设施把视频帧中的静态背景与动态前景进行分离，仅在初始时传输一次静态背景，其余每次仅传输动态前景信息，这样可以使得传输负载大幅降低。无人驾驶汽车将收到的静态背景图像与动态前景图像重新融合成视频帧，然后基于视频帧所反映的行车环境做出正确的驾驶决策。对于静态背景与动态前景的分离，本文提出了一种基于像素值计算的视频帧背景去除和降噪方法，能够快速地从视频帧中提取动态前景；对于静态背景与动态前景的融合，提出了一种基于生成对抗网络的视频帧生成方法，能够快速地把静态背景和动态前景融合成视频帧。通过在真实数据集上的测试可知，本文提出的方法能够在重要环境感知信息不丢失的前提下使传输负载降低 85%以上，感知信息处理时间降低 70%以上。这表明本文提出的方法能够高效地实现无人驾驶汽车与路边基础设施的协同环境感知，有助于构建更加安全的无人驾驶系统。

关键词 无人驾驶汽车；协同环境感知；深度学习；生成对抗网络；传输负载

中图法分类号 TP393

Cooperative Sensing Information Transmission Load Optimization for Automated Vehicles

LV Pin[1),2),3)] LI Kai[1)] XU Jia[1),2),3)] LI Tao-Shen[1),3)] CHEN Ning-Jiang[1),3)]

[1)](College of Computer, Electronics and Information, Guangxi University, Nanning 530004)

[2)]Guangxi Key Laboratory of Multimedia Communications Network Technology, Nanning 530004)

[3)](Guangxi Colleges and Universities Key Laboratory of Parallel and Distributed Computing, Nanning 530004)

Abstract Automated driving has become a research hot spot in both academic and industrial circles in recent years. Environment perception of automated vehicles is a fundamental technology in automated driving. However, only increasing sensors on the automated vehicle or improving the accuracy of the sensors cannot completely eliminate the blind area of environment sensing. Therefore, cooperative environment sensing

---

本课题得到国家自然科学基金(62062008, 62062006)、“广西八桂学者”专项经费、广西自然科学基金(2018JJA170194, 2018JJA170028, 2019JJA170045)资助. 吕品，博士，副研究员，中国计算机学会(CCF)高级会员（43480S）,主要研究领域为无线网络、群智感知.E-mail: lvpin@gxu.edu.cn.李凯，硕士研究生，中国计算机学会(CCF)学生会员（D8324G）,主要研究领域为人工智能、群智感知.E-mail: 1197094688@qq.com. 许嘉（**通信作者**），博士，副教授，中国计算机学会(CCF)高级会员（41578S）,主要研究领域为大数据分析与处理技术.E-mail: xujia@gxu.edu.cn. 李陶深，博士，教授，中国计算机学会(CCF)杰出会员（05005D）,主要研究领域为无线网络、协同计算.E-mail: tshli@gxu.edu.cn.陈宁江，博士，教授，中国计算机学会(CCF)高级会员（10254S）,主要研究领域为软件工程、协同计算.E-mail: chnj@gxu.edu.cn.

between automated vehicles and roadside infrastructure has attracted increasingly more attention. With the help of the cooperative environment sensing with roadside infrastructure, the sensing range of an automated vehicle is enlarged, which also promotes blind area elimination. Cooperative environment sensing is significant to improve the safety of automated driving. Among all kinds of environmental sensing information, the videos captured by cameras occupy the most important position. However, video frames contain a large amount of data. Transmitting each video frame leads to a heavy network traffic load and a long transmission delay, which affects the timeliness of environmental sensing information. In this paper, a video transmission load optimization framework is proposed. The main idea of the framework is that, the roadside camera separates the dynamic foreground from the static background in the video frame. It only transmits the static background once at the beginning; and in the following transmissions, only dynamic foreground in the video frames are transmitted, which reduces the transmission load greatly. After receiving dynamic foreground images, the automated vehicle fuses them with the previously received static background, and recover the video frames. Hence, the automated vehicle can make the correct driving decision based on the driving environment reflected by the recovered video frames. For dynamic foreground and static background separation, a pixel-based method is proposed to remove the background and reduce the noise quickly. With the help of the proposed method, the dynamic foreground is able to be extracted from the video frame in an efficient manner. For dynamic foreground and static background fusion, an approach based on generative adversarial network (GAN) is utilized in this paper to fuse dynamic foreground and static background into new video frames efficiently. With the confrontation between the generative model and the discriminative model, the quality of the recovered video frame improves. Through the performance evaluation on the real data set containing more than 43,000 images captured by roadside cameras, the following results are obtained. The framework proposed in this paper can reduce the transmission load by over 85% without lost in the key environmental sensing information, and also can reduce the environmental sensing information processing time by over 70%. Measurements on several metrics reveal that the quality of the fused image also outperforms other contrast methods. The results indicate that the proposed framework achieves efficient cooperative environment sensing for automated vehicles and roadside infrastructure, which is conducive to build a safer unmanned driving system.

Key words Automated vehicle; Cooperative environment sensing; Deep learning; Generative adversarial networks; Transmission load.

1 引言

随着人工智能技术的发展，无人驾驶汽车逐渐从愿景走向现实，成为学术界和工业界近年来的研究热点，各大传统汽车厂商和新兴科技公司都积极投身于无人驾驶汽车的研发之中。

在无人驾驶汽车相关技术中，环境感知是车辆自动做出各项行为决策和运动控制的基础。只有获得了充分、精确、可靠的环境感知信息，无人驾驶汽车才能做出安全、合理的驾驶决策。无人驾驶汽车依靠多种传感器（如摄像头、激光雷达、毫米波雷达等）进行环境感知。当前业界提升无人驾驶汽车环境感知能力的主要方法是安装数量更多、精度更高的传感器，然而这种方法并不能消除因障碍物遮挡而产生的感知盲区。因此，仅提升无人驾驶汽车的单体感知能力存在一定的局限性。采用群智协同环境感知的策略则可以突破上述局限。当一个区域对于一辆无人驾驶汽车来说是感知盲区，而这个区域对其他节点来说是可感知区域时，那么这辆无人驾驶汽车就可以从其他节点获取这个区域的感知信息，从而可以扩大自身的感知范围，消除感知盲区，实现非视距感知。由此可见，群智协同环境感知对于提升无人驾驶安全性具有重要意义[1]。

与其他感知数据相比，摄像头拍摄的视频数据所包含的环境信息往往更加丰富和直观，对于环境感知具有更重要的作用，百度、特斯拉等公司甚至研发了基于纯视觉感知数据的无人驾驶汽车。因此，在进行协同环境感知时，视频数据是无人驾驶汽车与其他感知节点共享的主要数据类型。在实际

应用场景中，道路监控摄像头往往具有固定的安装位置、稳定的电源供应、广阔的拍摄视野，因此非常适合作为无人驾驶汽车的协同感知节点。如图 1 所示，路边摄像头把拍摄到的视频数据发送给无人驾驶汽车，就能帮助车辆扩大自身的感知范围，根据环境情况及早做出安全、合理的驾驶决策。

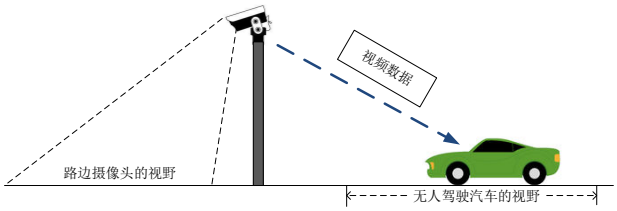

图 1 协同感知示意图

然而，随着摄像头分辨率不断提高，摄像头每秒钟所产生的视频数据量急剧增长。一个高清摄像头每秒产生的数据量可达几十兆比特，而车辆或路侧单元通常会安装多个摄像头以覆盖各个方向，使得每个节点产生的数据量更是成倍增长。现有的车载网通信技术，如车辆专用短程通信技术（DSRC）、3G/4G 等，很难支持如此巨大的传输负载；特别是在高速移动时，车辆能获得的有效传输速率会更低。即使采用容量更大的 5G 网络进行传输，当一个区域内有很多车辆时，为这些车辆传输大量视频数据也会使得网络负载过重，进而造成传输延迟增大，不利于环境感知数据的时效性，同时也会影响其他网络应用的正常运行。因此，网络传输负载受限成为阻碍无人驾驶汽车进行群智协同环境感知的重要因素。

为了降低协同环境感知数据的传输负载，本文提出了一种基于深度学习的传输负载优化方法。该方法的主要思想是，协同环境感知数据的发送方将视频帧中的静态背景与动态前景相分离，静态背景只需在初始时传输一次，对于之后的每个视频帧，仅传输其中的动态前景数据；无人驾驶汽车在收到动态前景数据后，将其与静态背景数据重新融合成视频帧，并基于视频帧所反映出的环境信息做出正确的驾驶决策。与传统方法不同，本文提出的方法不是传输每个完整的视频帧，而是传输其中发生动态变化的部分，这样就可以使得网络负载大幅降低，有利于保证环境感知信息传输的时效性。通过在真实数据集上的实验可知，使用这种方法能够在不丢失对驾驶决策起作用的环境感知信息的基础上，将视频图像数据的传输负载降低 85%以上。

本文的主要贡献总结如下：

（1）提出了一种降低协同环境感知信息传输负载的方案，通过在发送端对视频帧中静态背景与动态前景进行分离，在接收端再对两者进行融合，可以使得传输负载大幅降低；

（2）针对如何快速分离视频帧中静态背景和动态前景的问题，提出了一种基于像素值计算的视频帧背景去除和降噪方法，能够快速地从视频帧中提取动态前景；

（3）针对如何快速融合视频帧中静态背景和动态前景的问题，提出了一种基于生成对抗网络的视频帧生成方法，能够快速地把静态背景和动态前景融合成视频帧；

（4）在真实数据集上进行了测试，结果表明本文提出的方法不会丢失对驾驶决策起作用的环境感知信息，并且能够把传输负载降低 85%以上。

本文后面的部分安排如下：第 2 节对相关工作进行了总结；第 3 节和第 4 节分别对环境图像数据中静态背景和动态前景的分离方案和融合方案进行描述；第 5 节对本文方法进行了实验评估，并对实验结果进行了分析；第 6 节对全文进行了总结。

2 相关工作

2.1 面向无人驾驶的协同环境感知

群智感知是指以普通用户的移动设备作为基本感知单元，大量感知单元通过移动互联网进行有意识或无意识的协作，实现感知任务分发与感知数据收集，完成大规模的、复杂的社会感知任务[2,3]。群智感知已经在智慧城市[4]、环境监测[5]、智能交通[6]、公共安全[7]等领域都有了不少研究工作。受群智感知思想的启发，无人驾驶汽车协同环境感知已经开始受到关注[8]，即无人驾驶汽车通过与其他车辆和路边基础设施共享环境感知数据，使得无人驾驶汽车的环境感知能力获得提升。但与传统群智感知问题不同的地方在于，在无人驾驶汽车协同环境感知场景下，摄像头、激光雷达等传感器带来的数据量更大，并且无人驾驶汽车对感知数据的实时性和可靠性有着更为严格的要求，而车联网环境又具有显著的异构性和动态性，这使得已有的群智感知机制并不能很好地满足无人驾驶汽车的独特需求。

一些研究人员针对无人驾驶汽车的激光雷达感知数据提出了不同的压缩技术。例如，首先将激

光雷达的点云数据组织为二维图像阵列，然后使用传统图像压缩技术[9]、聚类技术[10]或深度学习技术[11]对图像进行压缩。由于基于纯视觉的无人驾驶汽车成为重要的发展方向，因此本文主要关注以摄像头拍摄的图像数据作为无人驾驶汽车环境感知信息来源的应用场景。

对于图像数据，H.265 编码技术[12]可以利用帧内预测编码和帧间预测编码来降低视频图像空间冗余和时间冗余，从而实现视频图像的数据压缩。H.265 中的编码帧包括 I 帧、P 帧和 B 帧，I 帧为帧内编码帧，P 帧为当前帧与前一帧（I 帧或 P 帧）的差别，B 帧为双向预测编码帧。然而，帧间编码具有依赖性，一旦 I 帧或 P 帧在传输过程中出错或丢失都会导致后续的帧出错，不适用于丢包率较高的车载网络环境。

本文针对无人驾驶汽车协同环境感知这一应用场景进行研究，提出了将视频图像中的静态背景与动态前景相分离的策略，能够大幅降低传输负载，与已有研究工作[1,8,12]有着显著的不同。

2.2 视频图像静态背景与动态前景的分离和融合

由于本文提出的传输负载优化方法涉及视频帧中静态背景与动态前景的分离和融合，以下分别从这两个方面对相关工作进行总结。

静态背景与动态前景的分离是许多计算机视觉任务（如目标跟踪、人群分析等）的关键步骤，近年来深度学习技术被越来越多地应用于这个领域。文献[13]使用卷积神经网络从给定的视频序列中进行背景构造和前景信息提取。文献[14]考虑了视频的时间连续性，将三维卷积应用于视频的最新帧，追踪视频序列的时间变化，实现了端到端的背景减除。文献[15]利用多尺度的全卷积网络提升模型学习能力，大大提高了前景检测准确性。上述背景减除方案都仅考虑了前景物体的大致形状，而对前景物体的细节方面刻画不够精准。

在静态背景与动态前景融合方面，近年来深度学习技术已被成功应用于图像融合领域，主要包括红外与可见光图像融合、医学图像融合和多焦点图像融合等。文献[16]首次将卷积神经网络引入图像融合领域，提出了一种可用于多焦点图像融合的卷积网络，展示了卷积神经网络在图像融合领域中的潜力。文献[17]在文献[16]的基础上将卷积神经网络进一步引入医学图像融合领域，视觉质量和客观评估方面都可以取得令人满意的结果。文献[18]将三维卷积神经网络引入泛锐化处理，生成高分辨率高光谱图像。文献[19]在泛锐化问题中引入残差网络，取得了更好的结果。文献[20]利用生成对抗网络[21]的思想处理可见光与红外线的问题，融合的图像更好地保留了所需的信息。文献[22]利用文献[23]提出的密集连接卷积神经网络结构进行可见光与红外线融合，充分利用了中间层所获得的信息。为了更好地满足多任务的需求，通用的网络模型被人们提出，在有监督学习和无监督学习方面都取得了优异的表现。IFCNN[24]是最新提出的通用有监督图像融合模型，以卷积神经网络为基础。随着输入图像的不同，模型可以选择不同的融合规则。利用预训练好的 Resnet 网络[25]良好的特征提取能力和与之相关的感知损失函数，IFCNN 在不同的任务中获得了比以往模型更好的表现。DIF[26]是关于通用无监督图像融合的最新研究成果，为各类缺少标记的无监督学习任务提供了新的思路。与 IFCNN 相同，DIF 同样使用卷积神经网络作为模型构造的基础。在进行图像融合的过程中，DIF 以生成与高维输入图像具有相同对比度的输出图像作为目标。为了使模型的融合结果保留更多的原始图像细节，DIF 将结构张量引入损失函数，重新考虑了局部对比度的概念。在定量和定性评估方面，DIF 都优于各类任务的最新技术。

由于在无人驾驶应用场景中环境感知信息的时效性和准确度要求更高，因此本文提出了更加快速的静态背景和动态前景的分离与融合方法，更加适合无人驾驶汽车协同环境感知应用场景。

3 静态背景与动态前景的分离

系统中的信息传输如图 2 所示。路边摄像头将拍摄到的原始图像发送给路侧单元中的计算模块，计算模块将前景和背景图像进行分离，通过传输模块发送给无人驾驶汽车。无人驾驶汽车将收到的前景和背景图像通过环境构建模块进行融合，并且结合自身摄像头拍摄的图像，形成环境感知信息。根据这些环境感知信息，驾驶决策模块将做出车辆控制决策，交由车辆控制模块实施。

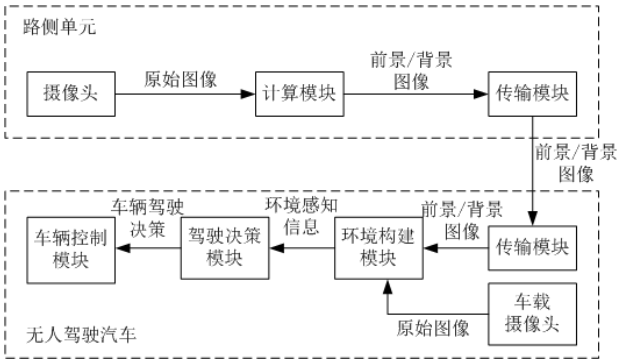

图 2 信息传输示意图

为了降低视频数据的传输负载，本文采用了“动静分离”的传输方法，即把图像静态背景与动态前景进行分离，分别进行传输。图 3 显示了传输每一帧视频数据的传统方法与动静分离的传输方法的不同。动静分离的传输方法在初始时传输一次环境图像的静态背景，之后就仅传输环境图像中动态前景，这样能够避免静态背景数据的重复传输，从而大幅降低传输负载。

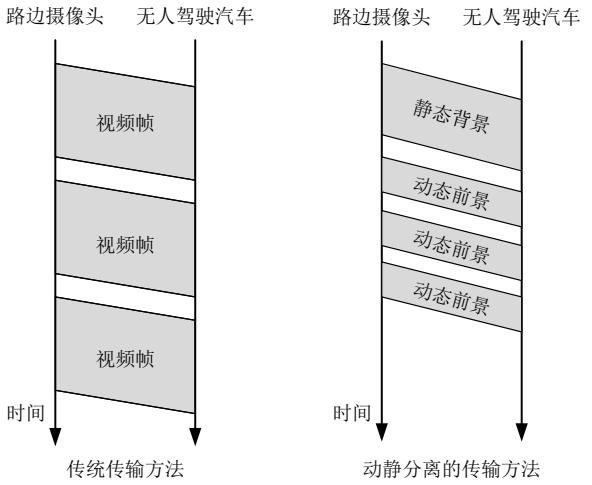

图 3 动静分离的传输方法与传统传输方法对比

已有的研究工作[13-15]均需要较长时间的训练和运行时间。为了保证环境感知数据处理的实时性，本文采用了更为高效的静态背景与动态前景分离方法，步骤如下：

（1）路边摄像头首先拍摄一张视野内无移动物体时的图像，作为静态背景图像。

（2）由于光照强度会随时间发生变化，摄像头实时拍摄的图像与之前所拍摄背景图像的光照条件可能不同。如果直接进行背景减除，会造成减除背景后的图像存在较多噪音。为了降低光照变化对图像背景减除带来的影响，路侧单元需要对摄像头拍摄的实时图像与静态背景图像中的每个像素按公式(1)进行灰度归一化预处理：

$$x_t = (\frac{x - x_{\min}}{x_{\max} - x_{\min}}) * 255 \tag{1}$$

其中 x 为本次拍摄的图像像素灰度值，$x_{\min}$ 为图像矩阵中灰度最小值，$x_{\max}$ 为图像矩阵中灰度最大值，x_t 代表经过灰度归一化预处理后的像素灰度值。因此无论图像的光照条件有何不同，处理后的图像灰度都被统一到[0, 255]这个范围内，从而方便进一步的处理和匹配。

（3）将经过第（2）步处理的背景图像和实时图像进行相似度比较。比较的过程为：首先，对背景图像和实时图像按照相同的规格划分成多个区域，然后比较对应区域的相似度。如果两个对应区域相似，则说明该区域内的图像为背景，因此需要去除该区域内的图像信息，即将该区域内的像素值都置为 0；如果两个对应区域不相似，则说明实时图像中该区域内包含前景物体，因此需要保留。相似度计算方法如公式(2)所示：

$$p = \frac{\mathrm{cov}(x, x_b)}{\sigma_x \sigma_{x_b}} = \frac{\mathrm{E}[(x - \mu_x)(x_b - \mu_{x_b})]}{\sigma_x \sigma_{x_b}} \tag{2}$$

$$\sigma_x = \sqrt{\sum_{i=1}^{n} (x_i - \mu_x)^2} \tag{3}$$

$$\sigma_{x_b} = \sqrt{\sum_{i=1}^{n} (x_{bi} - \mu_{x_b})^2} \tag{4}$$

其中，x_b 和 x 分别是背景图像和实时图像中对应区域像素矩阵转换成的向量，$\mathrm{cov}(x, x_b)$ 为两个向量的协方差，μ_x 和 μ_{x_b} 分别是 x 和 x_b 的均值，σ_x 和 σ_{x_b} 分别是 x 和 x_b 的标准差。标准差的计算分别如公式(3)、公式(4)所示。

使用上述方法后，路边摄像头就可以从拍摄到的视频帧中快速分离出动态变化的前景图像用于传输。与直接传输整个视频帧相比，用这种方法所需传输的数据量大幅降低，可以有效降低网络负载，并且提升了环境感知数据的时效性。

4 静态背景与动态前景的融合

无人驾驶汽车收到动态变化的前景图像数据

后，需要把前景图像与背景图像重新融合成完整视频图像，有助于无人驾驶汽车判断前景图像所代表物体的相对位置，从而做出正确的驾驶决策。

本文设计了一个基于生成对抗网络（Generative Adversarial Networks, GAN）的前景图像与背景图像的融合机制。该机制包括生成模型和判别模型两个部分，对于判别模型还需设计梯度约束以帮助模型进行深度学习。考虑到无人驾驶对时延和精度的高要求，分别从两个方面进行设计：一是利用注意力机制对关键信息的关注和对噪音的抑制，结合生成对抗网络的思想帮助提升网络融合精度；二是利用密集卷积神经网络对特征图的复用，降低网络的深度，减少融合所需的时间。结合 WGAN-GP[27] 的思路，提出了对抗性背景融合模型：FWGAN。

本文出现的符号如表 1 所示。

表 1 本文出现的符号

符号	含义	初始值
r_i	神经网络第 i 层感受野大小	1
	（输入图像作为第 0 层）	
s_i	第 i 卷积层的步长大小	1
t_i	第 i 卷积层的有效步长	—
k_i	第 i 卷积层的卷积核大小	—
κ	生成器对抗性损失和内容性损失平衡系数	1
λ	内容性损失中的信息损失和结构性损失平衡系数	100
ρ	对真实图像和生成图像采样范围进行插值采样的系数	0.3
θ	判别器梯度的范围	1
ω	生成器参数	—
δ	判别器参数	—
α	RMSProp 优化器参数	0.9
∇	梯度	—

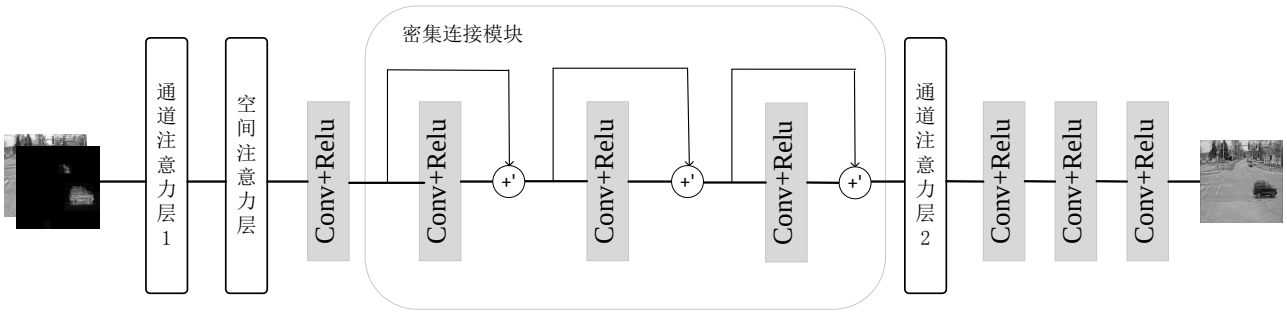

图 4 生成模型结构图

4.1 生成模型

生成模型模拟了人类视觉对两张透明度不同图像的叠加过程：背景图像不透明，前景图像透明度高且空白区域较多。视觉会将背景图像整体内容作为基底，忽视前景图像中的空白区域，将其中的关键信息与背景图像叠加，获得最终视觉效果。生成模型的结构如图 4 所示，分别由 2 个通道注意力层、1 个空间注意力层、密集连接模块和普通卷积层构成。每一个卷积层后使用 ReLu 作为激活函数。

针对背景图像与前景图像所包含信息量差距较大的特性，利用通道注意力层 1 和空间注意力层[28]对输入的双通道图像进行直接处理。通道注意力层 1 在通道层面给包含信息量更多的背景图像赋予更大的权重，空间注意力层则对前景图像给予更多的关注，因此生成模型能够在特征提取过程中将注意力更快地集中到关键信息。通道注意力层 1 的结构如 5 所示，空间注意力层的结构如图 6 所示。

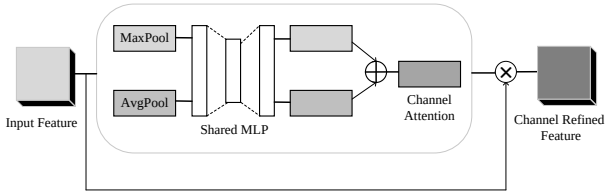

图 5 通道注意力层 1 结构图

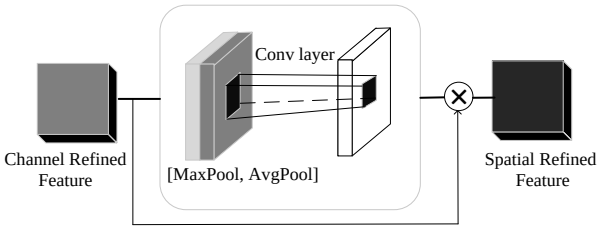

图 6 空间注意力层结构图

密集连接模块对提取的特征图进行复用，不仅帮助降低模型的深度，还通过特征复用给与图像边缘像素点更多参与计算的机会，拓展有效感知范围，提升模型的精度。在卷积计算中，图像边缘像素点参与运算次数小于图像中央像素点，这会影响融合图像边缘清晰度。密集连接模块通过复用多尺度的特征层，增加了图像边缘像素点参与卷积运算的次数，增强融合图像清晰度。

通道注意力层 2 结合特征图之间的信息依赖[29]，帮助模型对不同阶段获得的特征图进行权重分配，其结构如图 7 所示。

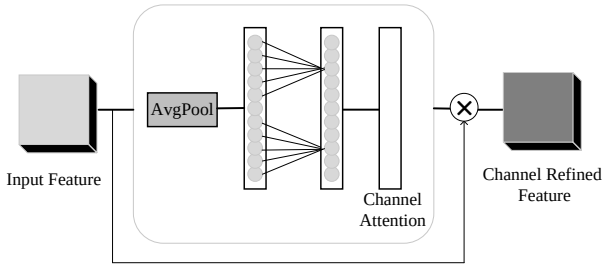

图 7 通道注意力层 2 结构图

在通道注意力层 2 之后，还需经过 3 个普通卷积层对特征图做进一步处理，以完成对背景图像和前景图像的融合。

4.2 判别模型

上述生成模型需要学习图像数据的深层特征并能够重构图像；与此不同，FWGAN 的判别模型本质是一个二分类网络，根据输入图像产生为真或假的判别结果，对生成模型进行反馈。判别模型的结构如图 8 所示。在图像分辨率为 256×256 时，判别器由 6 个卷积层构成，卷积核大小均为 4×4，输入层步长设为 4，输出层步长设为 1，中间层步长设为 2。判别器最后一个卷积层使用 Sigmoid 作为激活函数来完成二分类任务，其余卷积层均使用 LeakyReLu 激活函数，这是因为 LeakyReLu 函数能够帮助判别模型更好地学习数据特征[30]。

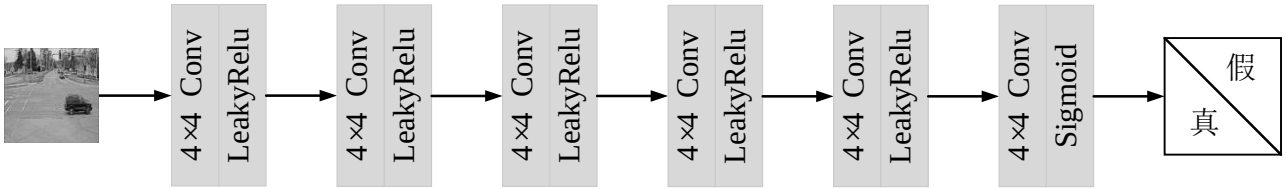

图 8 判别模型结构图

判别模型中卷积层的数量由感受野的大小决定。感受野是指输入中对当前层产生影响的区域大小[31]，计算方式如公式(5)所示。

$$r_i = (k_i - 1) * s_i + r_{i-1} \tag{5}$$

$$s_i = s_{i-1} * t_i \tag{6}$$

公式(6)为有效步长计算计算公式，其中 r_i 为第 i 层感受野的大小，输入层是第 1 层，初始 r_0 为 1；s_i 为第 i 层的有效步长，初始 s_0 为 1；k_i 为第 i 卷积层的卷积核大小；t_i 为第 i 层卷积层步长的大小。

在图像分辨率为 256×256 时，可以计算得出感受野大小为 376×376。由于输入图像像素小于感受野的大小，所以判别模型是合理的。在图像分辨率为 128×128 和 512×512 的情况下，判别模型中卷积层的数量分别为 5 和 7。

4.3 损失函数

FWGAN 的损失函数包括生成模型的损失函数 L_G 和判别模型的损失函数 L_D 两部分。

生成模型的损失函数 L_G 反映了生成模型的训练目标，其中包括如公式(7)所示的两部分：

$$L_G = V_{FWGAN}(G) + \kappa L_{content} \tag{7}$$

$V_{FWGAN}(G)$ 为生成模型和判别模型之间的对抗性损失，计算方法如公式(8)所示：

$$V_{FWGAN}(G) = \min(-\sum_{\bar{x}\sim P_g}[D(\bar{x})]) \tag{8}$$

其中，$\bar{x}$ 为所得融合图像样本域 P_g 的随机采样，$D(\bar{x})$ 为判别模型对融合图像采样的判别结果。

$L_{content}$ 表示融合图像和真实图像的内容性损失差异，参数 κ 用于平衡对抗性损失和内容性损失。内容性损失包括图像信息损失和结构性损失两部分，如公式(9)所示：

$$L_{content} = L_{pixel} + \lambda L_{ssim} \tag{9}$$

L_{pixel} 代表融合图像和真实图像的像素差，作为衡量图像整体损失的指标；L_{ssim} 代表融合图像和真实图像的结构性差异[32]，作为衡量图像结构性损失的指标。参数 λ 用于平衡信息损失和结构性损失。L_{pixel} 和 L_{ssim} 的定义如公式(10)和(11)所示：

$$L_{pixel} = \sum_{i=1,j=1}^{n}(\bar{x}_{i,j} - x_{i,j})^2 \tag{10}$$

$$L_{ssim} = 1 - \frac{(2\mu_x\mu_{\bar{x}} + c_1)(2\sigma_{x\bar{x}} + c_2)}{(\mu_x^{\ 2} + \mu_{\bar{x}}^{\ 2} + c_1)(\sigma_x^{\ 2} + \sigma_{\bar{x}}^{\ 2} + c_2)} \tag{11}$$

其中，x 为真实图像样本域 P_r 的随机采样；公式(10)中，$\bar{x}_{i,j}$ 为融合图像 $\bar{x}$ 在点 (i, j) 处的像素大小，$x_{i,j}$ 为真实图像 x 在点 (i, j) 处的像素大小，n 为图像大小；公式(11)中 μ_x、$\mu_{\bar{x}}$ 分别为 x 和 $\bar{x}$ 的平均值，c_1 和 c_2 是用来维持稳定的常数。

判别模型的损失函数 L_D 反映了判别模型的训练目标，其定义如公式(12)所示：

$$L_D = \min(E_{\bar{x}\sim P_g}\left[D(\bar{x})\right] - E_{x\sim P_r}\left[D(x)\right] + \theta E_{\hat{x}\sim P_{\hat{x}}}[(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2 - 1)^2]) \tag{12}$$

其中，$\hat{x}$ 为融合图像与真实图像之间区域的随机插值采样，如公式(13)所示：

$$\hat{x} = \rho x + (1-\rho)\bar{x} \quad \rho \in uniform[0,1] \tag{13}$$

$\theta E_{\hat{x}\sim P_r}[(\left\|\nabla_{\hat{x}}D(\hat{x})\right\|_2 - 1)^2$ 为梯度惩罚项，将判别模型梯度约束在固定范围，以保证训练的稳定。

4.4 训练流程

在 WGAN[33] 中，批归一化（Batch Normalization）[34]被用来帮助网络的训练。通过加入可训练参数对数据进行批归一化处理，可以规范神经网络层的输入分布，从而加快神经网络的训练速度。但是，批归一化将判别模型问题的形式从单个输入映射到单个输出更改为一批输入映射到一批输出。公式(12)中的梯度惩罚项要求对每个样本独立地施加梯度惩罚，与批归一化的批量处理方案冲突。因此在判别模型中批归一化层无法使用。在优化算法的选择方面，实验证明 RMSProp 算法能够比 WGAN-GP[27]所使用的 Adam 算法取得更好的结果，因此本文选择了 RMSProp 算法。训练流程如算法 1 所示。

算法 1. FWGAN 算法.

输入：前景图像样本 x_t，真实样本 x，背景图像样本 x_b，内容损失系数 λ，梯度约束项系数 θ，参数为 ω 的生成模型 G_ω，参数为 δ 的判别模型 D_δ，学习率 l，RMSProp 超参 α，批处理大小 m，判别器更新次数 n

输出：判别器分类结果

1. WHILE ω has not converged DO
2. FOR t = 1, ..., n DO
3. FOR i = 1, ..., m DO
4. get background picture $x_b \sim P_b$
5. get foreground picture $x_t \sim P_t$
6. get real picture $x \sim P_r$
7. get a random number $\rho \in uniform[0,1]$
8. $\bar{x} \leftarrow G_\omega(x_t, x_b)$
9. $\hat{x} \leftarrow \rho x + (1-\rho)\bar{x}$
10. $L_D \leftarrow [D_\delta(\bar{x}) - D_\delta(x) + \theta(\left\|\nabla_{\hat{x}}D_\delta(\hat{x})\right\|_2 - 1)^2]$
11. END FOR
12. $\delta \leftarrow \mathrm{RMSPRop}\left(\nabla_\delta \frac{1}{m}\sum_{i=1}^{m} L_D, \delta, l, \alpha\right)$
13. END FOR
14. $\omega \leftarrow \mathrm{RMSPRop}\left(\nabla_\omega \frac{1}{m}\sum_{i=1}^{m} -D_\delta\left(G_\omega(x_t, x_b)\right), \omega, l, \alpha\right)$
15. END WHILE

(a) (b) (c) (d)

(e) (f) (g)

图 9 图像分离与融合对比示例。(a)是真实的拍摄图像（包含前景与背景）；(b)是真实的拍摄图像（只包含背景）；(c)是从(a)中去除(b)后得到的前景图像；(d)是使用本文方法将(b)与(c)融合形成的图像；(e)是不使用生成对抗网络融合得到的图像；(f)是使用文献[24]中的有监督学习算法 IFCNN 融合得到的图像；(g)是使用文献[26]中的无监督学习算法 DIF 融合得到的图像。

(a) (b) (c)

图 10 夜间光照条件下的背景去除示例。(a)是真实的拍摄图像（包含前景与背景）；(b)是真实的拍摄图像（只包含背景）；(c)是从(a)中去除(b)后得到的前景图像。

5 性能评估

5.1 实验设置

本文使用 NVIDIA 公司发布的 CityFlow[35]数据集作为训练和测试用数据集，其中共包含 15 种不同场景。训练集包含 43264 张图像，测试集包含 1952 张图像，分别采用分辨率为 128×128、256×256 和 512×512 三种图像尺寸对模型性能进行验证。

在模型训练过程中，使用 RMSProp 作为模型优化器。设置衰减为 0.9，学习率为 0.001，每一个批次大小 m=16。生成模型损失函数中的 λ 取值范围较广，较大的 λ 取值能够帮助更快的收敛。实验中将 λ 取值设置为 100，能够帮助实验取得较好的结果。判别模型的 θ 取值为 1。

实验中，以具有 16GB 内存的 Intel Core 7700 CPU 作为路侧单元配置，以 Tesla T4 16G RAM 和 Intel Xeon Gold 6230 作为车辆单元配置。

5.2 实验结果与分析

5.2.1 图像分离与融合的视觉效果对比

图 9 展示了对视频帧中的静态背景和动态前景进行分离和融合的视觉效果。图 9(a)是真实的拍摄图像（包含前景与背景）；图 9(b)是真实的拍摄图像（只包含背景）；图 9(c)是子图(a)减子图(b)后提取出的前景图像；图 9(d)是使用本文方法将子

图(b)与子图(c)融合形成的图像；图 9(e)是不使用生成对抗网络融合出的图像；图 9(f)是使用有监督学习算法 IFCNN[24]融合得到的图像；图 9(g)是使用无监督学习算法 DIF[26]融合得到的图像。

通过对比图 9(a)和图 9(d)可以看到，使用本文方法融合前景与背景得到的图像与原始图像在视觉效果上极为接近，对于图 9(d)中黑色圆框标记的远处物体也能够很好地还原出来，使得对驾驶决策有影响的环境信息不被丢失。

通过对比图 9(d)和图 9(e)可知，如果不使用生成对抗网络，融合得到的图像中车辆轮廓仍然完整，对阴影部分的拟合效果也较好，但对于黑色圆框标记的远处物体清晰度差于生成对抗网络的表现。说明本文方法使用生成对抗网络对于提升图像融合效果有帮助。

通过对比图 9(d)和图 9(f)可知，IFCNN 对背景的拟合较好，但对从动态前景图像融合的效果较差，圆框标记的远处物体也没有能够恢复出来，不利于无人驾驶汽车从融合得到的图像中进行环境物体识别。

通过对比图 9(d)和图 9(g)可知，DIF 融合得到的图像中，静态背景和动态前景都不够清晰，会对无人驾驶汽车的环境感知造成不利影响。

为了验证不同光照条件下本文提出的背景减除方法的有效性，在实验中还选取了路边摄像头夜间拍摄的图像进行了测试。图 10(a)是夜间条件下包括前景与背景的图像，图 10(b)是夜间条件下只包含背景的图像，图 10(c)是从图 10(a)中去除图 10(b)得到的前景图像。从图 10 中可以看出，本文提出的背景去除方法能够适应较大范围的光照条件，具有较好的通用性。

5.2.2 前景物体保留率

根据本文提出的方法，视频图像需要经过静态背景与动态前景的分离与融合。在这个过程中，如果出现重要前景物体丢失的情况，则可能会对无人驾驶汽车的安全行驶造成不利影响，因此需要对动态前景物体的保留率进行测试与统计。

在路边摄像头拍摄的画面中（如图 11 所示），画面下方区域距离摄像头较近，其中的前景物体较大且相对清晰；画面上方区域距离摄像头较远，其中的前景物体较小且相对模糊。因此将视频图像分为两个区域：将图像上方四分之一的区域称为“远区域”，将图像下方四分之三的区域称为“近区域”。其中，近区域前景物体对于无人驾驶汽车的驾驶决策影响更大，远区域前景物体的影响则较小，因此分别对近区域和远区域内前景物体的保留率分别进行统计。

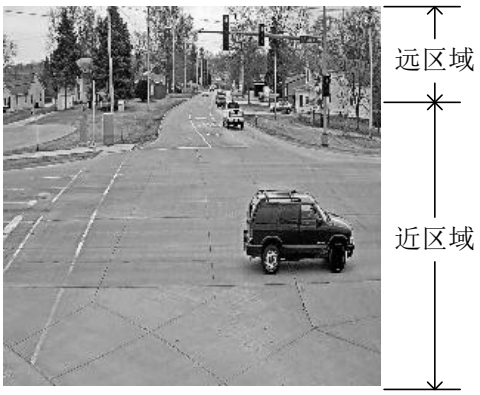

图 11 视频图像近区域与远区域划分示意图

经统计，如图 12 所示，在使用本文方法对视频图像进行背景与前景分离和融合后，近区域的前景物体保留率为 100%，远区域的前景物体保留率约为 83.3%。由此可见，本文方法能够保证近区域内的前景物体不会出现丢失，从而能够保证无人驾驶汽车环境感知的可靠性。远区域前景物体虽然有 17%左右的丢失率，但由于距离无人驾驶汽车较远，不会对车辆的安全行驶造成不利影响。

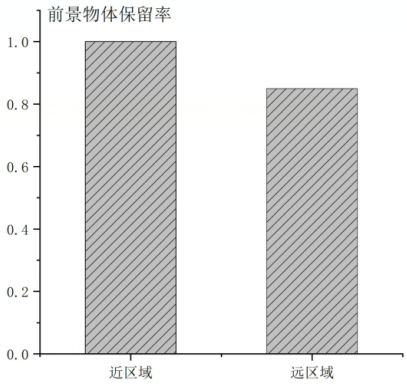

图 12 近区域与远区域的前景物体保留率

5.2.3 传输负载

按照传统的传输方法，路边摄像头不会对拍摄的图像进行处理，直接把类似于图 9(a)的每个视频帧向无人驾驶汽车进行传输。而本文提出的方法在图 9(a)所示的视频帧中去除如图 9(b)所示的静态背景，提取得到如图 9(c)所示的动态前景。在生成的前景图像中，除了前景物体部分，背景部分都具有相同的像素值（值为 0）。使用 JPEG 格式存储

这种前景图像时，能够有效压缩图像文件的大小，从而降低传输前景图像的数据量。因此，使用本文方法传输协同环境感知数据时，将首先传输如图 9(b)所示的静态背景一次，之后对于每个视频帧，仅传输如图 9(c)所示的动态前景图像，使得传输负载大幅降低。

实验中，图像尺寸为 256×256 时，每帧原始图像与每帧背景图像的大小约为 50KB，而去除背景的前景图像平均大小约为 7KB，如图 13 所示。

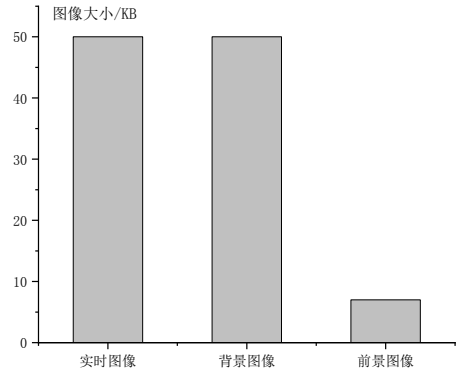

图 13 图像文件的大小对比

由于道路监控摄像头的拍摄覆盖距离通常为 200 米，假设道路被路边摄像头完全覆盖，则路边至少每 200 米就有一个摄像头。若无人驾驶汽车以 20 米/秒的速度行驶，则每个摄像头为该车辆的服务时间为 10 秒。当路侧单元以 10 帧/秒的速率向无人驾驶汽车传输图像时，则在服务时间内一共需要传输 100 帧视频图像。若直接传输原始图像，传输的数据量为 50KB×100=5000KB；而使用本文方法的传输数据量为 50KB+7KB×100=750KB，仅为 5000KB 的 15%。若服务时间内路侧单元向无人驾驶汽车传输更多的视频帧，使用“动静分离”的方法将使传输负载降低的比例更大。因此，本文方法能够将传输负载降低 85%以上。

5.2.4 感知信息处理时间

按照本文的方法，一帧视频图像从拍摄完成到交付给无人驾驶汽车的环境构建模块需要经历三个阶段，分别是前景与背景分离阶段、前景图像传输阶段、前景与背景融合阶段。

在分离阶段，对于尺寸为 256×256 的图像，对一帧图像进行前景与背景分离所需的时间约为 4.6ms，而已有研究工作[13-15]对一帧图像进行背景去除所需的时间一般在 50ms 以上。

在传输阶段，若使用车辆专用短程通信技术（DSRC）进行传输，传输速率为 6Mbps，那么传输一帧前景图像所需的时间约为 9.3ms；而传输一帧未经处理的原始图像所需时间约为 66.7ms。

在融合阶段，使用 FWGAN 将前景与背景融合成一帧图像所需的时间为 4.6ms。由此可知，使用本文方法处理感知信息时，一帧图像在三个阶段共需耗时 18.5ms，约为传统方法的 27.7%（如图 14 所示）。从上述结果可以看出，本文方法能够有效降低环境感知信息处理时间，更能满足无人驾驶汽车对环境感知的实时性要求。

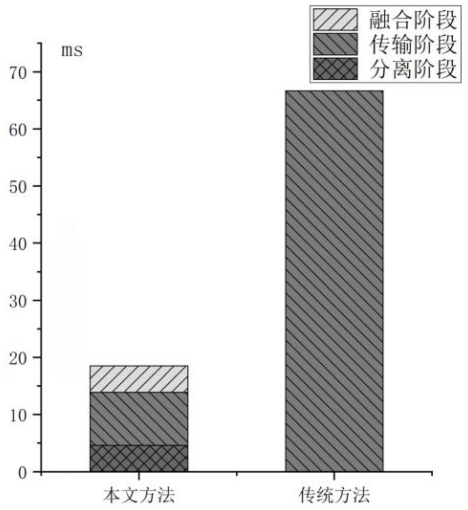

图 14 感知信息处理时间对比

5.2.5 图像融合质量的定量分析

本文对融合得到的图像与原始图像进行对比，用于评估模型融合表现的指标包括：

- 衡量图像结构相似度的 SSIM 指数，该指数的值越接近 1 说明融合图像与原始图像越相似；
- 衡量图像失真程度的 UQI[36]指数，该指数的值越大说明图像失真程度越低；
- 基于视觉信息保真度提出的衡量融合图像质量的指标 VIFF[37]，值越大说明融合表现越好；
- 衡量融合图像与原始图像相似程度的皮尔逊相关系数（Pearson correlation coefficient, PCC）[38]，该数值越大说明图像融合效果越好。

以上几类指标完整地考虑了融合图像保留细节信息、结构信息及失真效果的能力。

以下实验中分别对比了本文方法（FWGAN）、本文方法但不使用生成对抗网络（FW-Net）、基于有监督学习的 IFCNN[24]、基于无监督学习的 DIF[26]这四种方法在上述指标上的表现。

对于 SSIM 指标（如图 15 所示），在图像分辨率为 128×128 时， FW-Net 的融合表现最好，FWGAN 的表现次于 FW-Net，都高于 DIF 和 IFCNN；在图像分辨率为 256×256 和 512×512 时，FWGAN 和 FW-Net 的融合表现近似，仍明显优于 DIF 和 IFCNN，说明本文所提出的方案（无论是否使用对抗思想）在保持图像整体结构方面具有优势。

对于 UQI 指标（如图 16 所示），在三种图像分辨率下，FWGAN 的表现略优于 FW-Net，FWGAN 和 FW-Net 的表现明显优于另外两种方案，说明本文提出的方案融合图像时能够达到更低的失真程度，在使用对抗思想的情况下达到最优。

对于 VIFF 指标（如图 17 所示），在图像分辨率为 128×128 时，FWGAN 的表现优于 FW-Net；三种图像分辨率下，FWGAN 和 FW-Net 均优于另外两种方案，说明本文提出的方案能获得较高的视觉信息保真度，在使用对抗思想时达到最优。

对于 PCC 指标（如图 18 所示），在图像分辨率为 128×128 时，FWGAN 的表现最好，FW-Net 的表现与 FWGAN 近似；在图像分辨率为 256×256 时，FWGAN 的表现明显优于 FW-Net。在三种图像分辨率的情况下，FWGAN 和 FW-Net 的融合表现均优于另外两种方案。

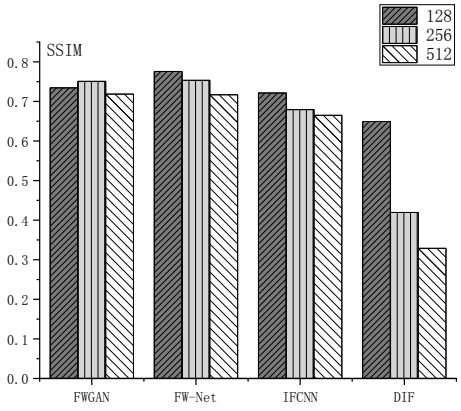

图 15 在 SSIM 指标上的对比

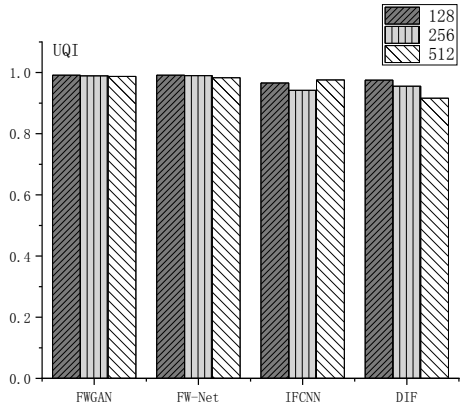

图 16 在 UQI 指标上的对比

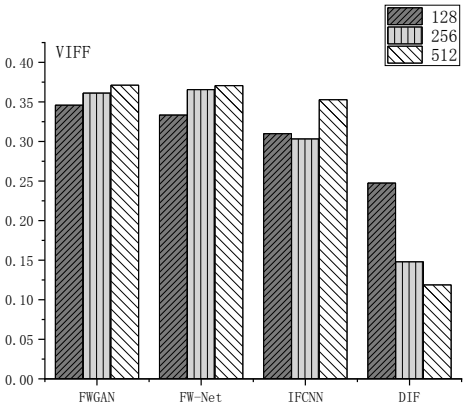

图 17 在 VIFF 指标上的对比

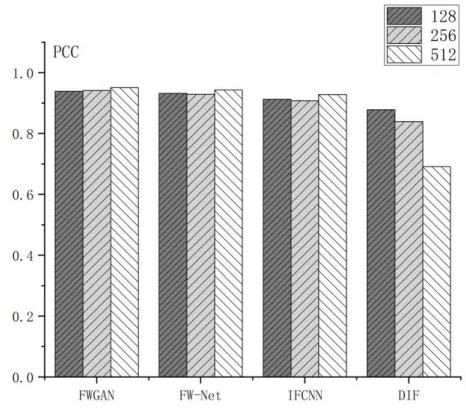

图 18 在 PCC 指标上的对比

综合考虑上述四种指标可以得出结论：本文提出的方法具有最佳的融合图像质量。与 IFCNN 和 DIF 方法相比，FWGAN 更充分地考虑了背景图像和前景图像的数据特性，利用注意力机制对关键信息赋予更高的权重，因此更适合无人驾驶汽车进行环境感知。

6 总结

协同环境感知对于无人驾驶技术的发展具有重要意义，但是受到网络容量的制约。本文提出了一种协同环境感知信息的传输负载优化方法，通过把视频帧中的静态背景和动态前景相分离，可以使静态背景在初始时只传输一次，之后仅需传输动态前景数据，达到了大幅降低传输负载的目的。无人驾驶汽车使用生成对抗网络将动态前景与静态背景重新融合成视频帧，并能够基于视频帧反映出的行车环境信息做出正确的驾驶决策。在真实数据集上的实验证明了本文提出方法的有效性，能够促进面向无人驾驶汽车的协同环境感知技术的进一步发展。

参 考 文 献

[1] Hobert L, Festag A, Llatser I, et al. Enhancements of V2X communication in support of cooperative autonomous driving. IEEE Communications Magazine, 2015, 53(12): 64-70.

[2] Ma H, Zhao D, Yuan P. Opportunities in mobile crowd sensing. IEEE Communications Magazine, 2014, 52(8): 29-35.

[3] Guo B, Wang Z, Yu, Z, et al. Mobile crowd sensing and computing: the review of an emerging human-powered sensing paradigm. ACM Computing Surveys, 2015, 48(1): 1-31.

[4] Wang J, Wang Y, Zhang D, et al. Crowd-powered sensing and actuation in smart cities: current issues and future directions. IEEE Wireless Communications, 2019, 26(2): 86-92.

[5] Vahdat-Nejad H, Asef M. Architecture design of the air pollution mapping system by mobile crowd sensing. IET Wireless Sensor Systems, 2018, 8(6): 268-275.

[6] Qiu H, Chen J, Jain S, et al. Towards robust vehicular context sensing. IEEE Transactions on Vehicular Technology, 2018, 67(3): 1909-1922.

[7] Simoens P, Xiao Y, Pillal P, et al. Scalable crowd-sourcing of video from mobile devices//Proceedings of ACM MobiSys, Taipei, China, 2013: 139-152.

[8] Wei S, Yu D, Guo C, et al. Survey of connected automated vehicle perception mode: from autonomy to interaction. IET Intelligent Transport Systems, 2019, 13(3): 495-505.

[9] Beek P V. Image-based compression of LiDAR sensor data. Electronic Imaging, 2019,43(7): 1-7.

[10] Sun X, Ma H, Sun Y, et al. A novel point cloud compression algorithm based on clustering. IEEE Robotics and Automation Letters, 2019, 4(2): 2132-2139.

[11] Tu C, Takeuchi E, Carballo A, et al. Point cloud compression for 3D LiDAR sensor using recurrent neural network with residual blocks// Proceedings of the IEEE International Conference on Robotics and Automation,Montreal, Canada, 2019: 3274-3280.

[12] Gary J. S, Jens-Rainer O, Woo-Jin H, et al. Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 22(12): 1649-1668.

[13] Babaee M, Dinh D. T, Rigoll G. A deep convolutional neural network for video sequence background subtraction. Pattern Recognition, 2018, 76: 635-649.

[14] Sakkos D, Liu Heng, Han Jun-Gong, et al. End-to-end video background subtraction with 3D convolutional neural networks. Multimedia Tools Applications, 2018: 23023–23041.

[15] Zeng D, Zhu M. Background subtraction using multiscale fully convolutional network. IEEE Access, 2018, 6:16010-16021.

[16] Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network. Information Fusion, 2017, 36: 191-207.

[17] Liu Yu, Chen Xun, Cheng J, et al. A medical image fusion method based on convolutional neural networks//Proceedings of the 20th International Conference on Information Fusion, Xi'an, China, 2017: 1-7.

[18] Giuseppe M, Davide C, Luisa V, et al. Pansharpening by convolutional neural networks. Remote Sensing, 2016, 8(7): 594.

[19] Rao YZ, He L, Zhu JW. A residual convolutional neural network for pan-shaprening//Proceedings of the International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 2017: 1-4.

[20] Ma JY, Yu W, Liang PW, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion. Information Fusion, 2019, 48: 11-26.

[21] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.

[22] Liu H, W XJ. DenseFuse: a fusion approach to infrared and visible images. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.

[23] Huang G, Liu Z, Laurens V, et al. Densely connected convolutional networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2261-2269.

[24] Zhang Y, Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network. Information Fusion. 2020, 54: 99-118.

[25] He K, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition.//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778.

[26] Jung H, Kim Y, Jang H, et al. Unsupervised deep image fusion with structure tensor representations. IEEE Transactions on Image Processing, 2020, 19: 3845-3858.

[27] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein GANs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 5769–5779.

[28] Woo S, Park J, Lee J.Y, et al. CBAM: convolutional block attention

module//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 3-19.

[29] Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.

[30] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks//Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico ,2016.

[31] Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks//Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 4905–4913.

[32] Zhou W, Bovik A. C, Sheikh H. R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[33] Arjovsky M, Chintala S, Bottou, et al. Wasserstein GAN//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 214-223.

[34] Ioffe S, Szegedy S. Batch normalization: accelerating deep network training by reducing internal covariate shift//Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015: 1-9.

[35] Zheng T, Naphade M, Liu M, et al. CityFlow: a city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 8789-8798.

[36] Hossny M, Nahavandi S, Creighton S. Information measure for performance of image fusion. Electronics Letters, 2008, 44(18): 1066-1067.

[37] Han Y, Cai Y, Cao Y, et al. A new image fusion performance metric based on visual information fidelity. Information Fusion, 2013, 14(2): 127-135.

[38] Wu W, Xu Y. Correlation analysis of visual verbs' subcategorization based on Pearson's correlation coefficient//Proceedings of the International Conference on Machine Learning and Cybernetics, Qingdao, China, 2010: 2042-2046.

LV Pin, Ph.D., associate researcher. His research interest include wireless networks, crowd sensing, etc.

XU Jia, Ph.D., associate professor. Her research interest include big data analysis and processing.

LI Tao-Shen, Ph.D., professor. His research interest include wireless networks and cooperative computing.

LI Kai, M. S. candidate. His research interest include artificial intelligence and crowd sensing.

CHEN Ning-Jiang, Ph.D., professor. His research interest include software engineering and cooperative computing.

Background

Automated driving is a current research hot-spot in the world. Due to the limitation of sensors, blind sensing area is inevitable for automated vehicles. Hence, cooperative environment sensing is an effective way to eliminate the blind sensing area and improve the safety of automated driving. Among all kinds of environmental sensing information, the video captured by camera occupies the most important position. However, video frames contain a large amount of data. Transmitting each video frame leads to heavy network load and increased transmission delay, which affects the timeliness of environmental sensing information. In this paper, a video transmission load optimization method is proposed. The main idea of the method is that, the transmitter separates the dynamic foreground from the static background in the video frame, and transmits the static background once at the beginning and only dynamic foreground in the following transmissions, which reduces the transmission load greatly. Using generative adversarial network, the automated vehicle fuses the static background and dynamic foreground into video frames again, and then makes the correct driving decision based on the driving environment reflected by the video frames. Through the performance evaluation on the real data set, it can be seen that the method proposed in this paper can reduce the transmission load by over 85% without lost in the environmental sensing

information, which lays the foundation for the promotion and application of cooperative environment sensing for automated vehicles.

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant Nos. 62062008 and 62062006, the special funds for Guangxi BaGui Scholars, the Guangxi Natural Science Foundation under Grant Nos. 2018JJA170194, 2018JJA170028, and 2019JJA170045.