|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 研究目标 | 方法框架 | 数据集 | 主要贡献、创新点 | 性能评估 | 开源代码 | 未来展望 | 发表情况 |
| **1. TarDAL**  Jinyuan Liu等大连理工大学 | 红外和可见光图像融合，针对目标检测 | 双层优化公式，双对抗学习网络生成器和双鉴别器结构；  先做了两种图像的融合，然后输入到yolo v5网络中做目标检测 | M3FD基准数据集、TNO、Roadscene、Multispectral | **提出TarDAL网络结构和M3FD数据集**  1. \*\***双层优化**公式\*\*：提出了一个新的双层优化模型，用于同时解决图像融合和目标检测的问题。这种公式允许在融合过程中考虑目标检测的需求，从而生成既适合视觉检查又适合计算机感知的融合图像。  2. \*\***目标感知的双对抗学习网络**\*\*：设计了一个包含一个生成器和两个鉴别器的网络结构，其中生成器负责创建融合图像，而两个鉴别器（目标鉴别器和细节鉴别器）分别评估融合图像中的前景目标和背景细节。这种结构有助于保留红外图像中的目标结构信息和可见光图像中的纹理细节。  3. \*\***合作训练**策略\*\*：开发了一种合作训练策略，通过反向传播算法同时优化融合网络和检测网络的参数。这种策略不仅提高了融合和检测的性能，而且提高了训练效率。  4. \*\***多场景多模态基准数据集**（M3FD）\*\*：构建了一个包含多种环境、照明、季节和天气条件下的红外和可见光图像对的全面数据集。 | 目标检测的mAP | [https://github.com/dlut-dimt/TarDAL](https://github.com/dlut-dimt/TarDAL。) | 没有提到未来工作。数据集扩展、算法优化、鲁棒性提升 | 2022  CVPR |
| 2. CDDFuse  Zixiang Zhao等西安交通大学 | 多模态图像融合，包括红外-可见光和医学图像 | 双分支Transformer-CNN架构；Restormer、Lite Transformer、INN块；  两阶段训练，第1阶段训练Encoder, Decoder，重建自身图像。（自编码器）  第2阶段，生成融合图像。 | MSRS、RoadScene、TNO等 | **提出CDDFuse融合方法**  1. \*\***双分支Transformer-CNN**架构\*\*：通过双分支结构分别处理全局和局部特征，以更有效地捕捉和融合多模态图像中的信息。  2. \*\***相关性驱动**的特征分解\*\*：相关性驱动的损失函数，增强低频特征（模态共享信息）之间的相关性，同时减少高频特征（模态特定信息）之间的相关性，从而实现更有效的特征分解和融合。  3. \*\***全局和局部**特征融合\*\*：CDDFuse通过特定的融合层将提取的全局和局部特征进行融合，以生成最终的融合图像。有助于在保留细节的同时，整合来自不同模态的互补信息。  4. \*\*广泛的应用场景\*\*：CDDFuse不仅适用于红外和可见光图像融合，还可用于医学图像融合等其他多模态图像融合任务，显示了其良好的泛化能力和应用范围。  5. \*\*统一的测量基准\*\*：CDDFuse提出了一个统一的测量基准，用于评估和比较不同融合方法在下游任务（如目标检测和语义分割）中的性能 | 8矩阵评估指标，评估融合后的图像的质量 | [https://github.com/Zhaozixiang1228/MMIF-CDDFuse](https://github.com/Zhaozixiang1228/MMIF-CDDFuse。) | 没有提到未来工作。 | 2023 CVPR |
| **3. MFFENet**  Wujie Zhou等浙江科技大学 | RGB-T城市道路场景的**语义分割**；专注于提高复杂光照条件下的语义分割性能 | MFFENet包含两个编码器、一个特征融合层和一个多标签监督层。该网络通过多尺度特征融合和增强机制，能够在RGB数据质量受损的情况下，准确解析RGB-T城市道路场景。 | MFNet和PST900 | 提出了一种新颖的多尺度特征融合和增强网络（MFFENet），用于准确解析RGB-热成像城市道路场景。   * 通过在编码器的末端添加一个紧凑版的空洞空间金字塔池化（ASPP）模块（CASPP），提取高层全局语义信息。 * 探索了在**多个阶段**进行RGB和热成像特征的交叉模态融合，而不是仅在低层或高层进行一次融合。 * 提出了一种**空间注意力**机制模块（SAMM），它能够更加关注前景区域，使MFFENet能够强调前景对象。 * 引入了**多标签监督**来优化MFFENet的参数。 | 融合图像生成和目标检测的平均精度（mAP） | **未公开代码** | 提高运行速度：通过使用更高效和强大的编码器  探索二值边界图监督：研究使用二值边界图作为监督信息来改进边界重建的性能  多模态数据的有效利用  半监督学习方法：由于获取语义分割的地面真实标签是一个繁琐且耗时的过程，未来的工作可以采用半监督学习方法来解决语义分割问题，而不完全依赖于标记数据。 | 2022  IEEE TRANSACTIONS ON MULTIMEDIA |
| **4.**  **MTANet**  Wujie Zhou等浙江科技大学 | 提高RGB-T城市场景理解的准确性和鲁棒性，特别是在不利光照条件下 | 双编码器结构、分层多模态融合、高级语义模块、多级融合模块、多任务模块、解码器结构、损失函数 | MFNet和PST900数据集 | 1. 分层多模态融合（Hierarchical Multimodal Fusion）：提出了一种新的融合策略，通过在网络的**不同层次上融合**RGB和热成像特征，更有效地结合了两种模态的互补信息。  2.多级融合模块（Multi-Level Fusion Module, MLFM）：设计了一个模块来融合不同层次的特征，包括低级、中级和高级特征，以保留更多的空间和语义细节。  3.多任务模块（Multi-Task Module, MTM）：引入了一种基于边缘保持、二值和语义监督的多任务学习策略，通过多个损失函数来共同优化网络，以提高分割精度和边界检测能力。  4.深度监督（Deep Supervision）：通过在网络的多个阶段引入深度监督，利用不同层次的特征图进行训练，增强了网络的学习能力和泛化性。  5.端到端训练：整个网络结构支持端到端的训练，可以直接从输入的RGB和热成像图像到最终的语义分割结果。 | **场景分割**后的可视化、mIoU等指标、模型的参数量、训练速度等 | **未公开代码** | 推理速度仍有待提高，特别是在嵌入式平台上的应用；  边界检测改进：MTANet提供的分割对象边界不够锐利，未来的工作将探索使用快捷连接来整合低级和高级特征图，以产生更精确的边界和细节。  信息选择机制：虽然RGB-T图像在某些情况下比其他类型的图像更具信息量，但对于温度相似的对象效果不佳。  突出最有信息量的数据：将最大化利用鉴别机制来突出最富有信息量的数据，以进一步提升城市场景理解的性能。 | 2023  IEEE TRANSACTIONS ON MULTIMEDIA |
| 5.  CGFNet | 提出一种新的跨模态融合网络（CGFNet），用于RGB-T（红绿蓝-热成像）**显著性目标检测**，特别是在光照不足或黑暗条件下 | 网络包含跨尺度交替引导融合（CSAGF）模块、指导融合模块（GFM）和跨引导融合模块（CGFM），以实现高效的跨模态特征融合和解码。 | VT821  VT1000  VT5000 | * GFM通过单模态作为主要引导，其他模态作为辅助，实现充分的跨模态融合。 * CGFM作为主要解码块，通过跨共享的跨层次增强（CLE）和全局辅助增强（GAE）实现多级特征的有效集成。 * CSAGF模块用于挖掘高层次语义信息并提供全局上下文支持。 |  |  |  | 2022  IEEE Transactions on Circuits and Systems for Video Technology |
| 6.  ECFFNet | 提出一种有效且一致的特征融合网络（ECFFNet），用于结合RGB图像和热成像进行**显著性目标检测**，以提高在复杂背景下的检测性能 | 网络采用双流结构，包括有效跨模态融合模块（ECFM）、双边反转融合模块（BRFM）和多级一致融合模块（MCFM）。 |  | * BRFM执行前景和背景信息的双边融合，以准确提取显著对象边界。 * MCFM在不同层级之间融合特征，以获取互补信息并生成最终的显著性图。 * ECFM通过改进的DIASPP块融合RGB和热成像模态的特征。 |  |  |  | 2022  IEEE Transactions on Circuits and Systems for Video Technology |
| 7.  IGT（Illumination-guided RGB-T object detection with transformers） | 提出一种**基于transformer的**目标检测框架，该框架旨在提高自动驾驶系统中目标检测的准确性和鲁棒性，尤其是在不同光照条件下 | 基于transformer的双流结构、Swin-Transformer-Tiny、照明基础子网络（IBSN）、照明引导特征校正模块（IFRM）、照明引导特征融合模块（IFFM） | FLIR数据集特别针对自动驾驶应用。多光谱数据集：为自动驾驶车辆准备了7512对RGB红外图像。 | 1.照明引导的框架：提出了一个新的基于transformer的RGB-T目标检测框架，该框架利用照明信息来指导多光谱特征的校正和融合过程。  2.新颖的融合机制：引入了照明引导特征融合模块（IFFM），使用一种新颖的基于照明的损失函数来促进多光谱特征融合，并生成照明不变的融合特征。  3.多光谱(近红外、中红外、远红外)融合的改进：重新思考了多光谱融合方法，特别是在不同光照条件下如何有效地利用多光谱特征之间的互补信息 | mAP50：这是在IoU（交并比）阈值为50%时的平均精度（mean Average Precision）。这是一个常用的评估目标检测模型性能的指标，特别是在自动驾驶领域。 |  |  | 2023  KNOWLEDGE-BASED SYSTEMS |
| 8.  Infrared and Visible Image Fusion Technology and Application: A Review | 这篇文章是关于红外和可见光**图像融合**技术及其应用的**综述** | 传统算法：包括多尺度变换、稀疏表示和子空间方法。  深度学习技术：包括自动编码器（AE）、卷积神经网络（CNN）和生成对抗网络（GAN）。 | MFNet 、哈佛医学院的脑部图像数据集、Landsat 和 Moderate Resolution Imaging Spectrometer等 | 技术与应用综述：文章全面回顾了主流的红外和可见光图像融合技术及其在机器人视觉、医学成像、农业遥感和工业缺陷检测等领域的应用现状。  评估指标分析：将图像融合方法的评估指标分为主观评估和客观评估，并分析了当前主流技术的属性，提供了对不同图像融合技术的评估和比较。  问题与挑战讨论：指出了红外和可见光图像融合过程中存在的问题，如融合图像的准确性不高、有不可恢复的像素丢失等，并讨论了提高图像融合技术性能的可能方向。  实验分析：通过选取代表性的图像融合方法和评估指标，对红外和可见光图像融合技术进行了定性和定量的测试分析，提供了实验结果的图形化展示。 | 文章将图像融合方法的评估指标分为主观评估和客观评估，分析了当前主流技术的属性，并进行了比较和展望。 | 综述文章，无代码 | 未来研究方向：对图像融合技术的未来发展方向进行了展望，提出了几个潜在的研究领域，包括数据集生成、神经网络构建、损失函数设计、评估指标和应用领域的挑战。 | 2023  SENSORS |
| 9.  Infrared pedestrian detection using improved Unet  Wei 等 | 一种改进的**红外行人检测方法**，该方法结合了改进的UNet和YOLO网络，并通过共享可见光域信息来提高检测精度 | 使用改进的UNet和YOLO网络架构，该架构能够从可见光域融合信息。 利用红外和可见光数据，提出了一种端到端的红外行人检测方法，该方法可以借用可见光成像中的行人特征来调整UNet参数，提高使用相对较小的红外数据集的红外行人检测性能 | FLIR、M3FD和LLVIP | 研究贡献：   * 提出了一种基于改进UNet和可见光预训练骨架的高效端到端红外行人检测方法。提出了改进的注意力UNet用于生成伪彩色行人图像。 * 提出的UNet + YOLO网络架构具有通用能力，可以与多种YOLO系列模型结合使用，包括但不限于YOLOv5和YOLOv7，用于红外行人检测。 * 该方法证明了其在实际道路场景中准确检测行人的能力，强调了其在实际部署中的潜力。 | FLIR、M3FD和LLVIP上，所提出的方法达到了87.7%、88%和97.5%的平均精度值 mAP。实时性能：在边缘计算设备上部署，并使用双光手持摄像机进行测试，达到了25.6 FPS的实时检测速度 |  | 数据增强技术：研究和开发专门针对红外行人检测的数据增强技术，以模拟不同的环境条件，如不同的天气状况或照明变化，从而提高模型的泛化能力。  不同遮挡条件下的性能：评估和改进算法在不同遮挡条件下的性能，如行人之间的遮挡、烟雾遮挡等。 | 2023  Measurement |
| 10.  RGB-T Salient Object Detection via Fusing Multi-Level CNN Features  Qiang Zhang等西电大学 | 提出一种新的**多模态显著目标检测网络**，该网络能够融合RGB和热红外图像的信息，以提高在各种挑战性条件下的检测性能 | 使用一个基本网络（如VGG-16）作为骨干网络，分别从RGB和热红外图像中提取粗略特征。  设计了几个相邻深度特征组合（ADFC）模块来提取每个单模态输入图像的多级细化特征。  使用多分支组融合（MGF）模块来捕获跨模态特征，通过融合ADFC模块的每一级特征。  引入一个联合注意力引导的双向消息传递（JABMP）模块，通过整合MGF模块的多级融合特征来进行显著性预测。 | RGB-thermal dataset  、Grayscale-thermal dataset、MSRA-B dataset | 研究贡献：   * 提出了一个端到端的基于CNN的RGB-T显著目标检测方法，该方法是**首个为RGB-T显著目标检测建立的端到端CNN**。 * 提出了ADFC模块来提取单模态图像特征，通过使用多个ADFC模块提取输入图像的多级特征。 * 提出了MGF模块来代替简单的特征连接，以捕获RGB-T图像对之间的跨模态补充信息，并减少网络参数的数量。 * 为了有效地捕获融合特征中的跨层补充信息，引入了JABMP模块进行最终的显著性预测。 | 在多个公共RGB-T显著目标检测数据集上的性能优于现有方法，特别是在照明条件差、背景复杂和对比度低等挑战性条件下。 |  | 包括将显著性检测器应用于工业应用，如图像分类、目标跟踪和实例级对象检索。 | 2020  IEEE Transactions on Image Processing |
| 11.  RGBT Salient Object Detection: A Large-Scale Dataset and Benchmark  Zhengzheng Tu等安徽大学 | 一种新的RGB-热红外（RGBT）**显著目标检测**方法，以及一个大规模的RGBT数据集VT5000的构建 | 提出了一个强大的基线方法，该方法提取每种模态的多级特征，并通过注意力机制聚合这些特征，以实现准确的RGBT显著目标检测。 为了解决显著目标边界模糊的问题，还使用了边缘损失来细化边界。 | VT5000、VT821、VT1000、NJU2K、NLPR、STERE等 | VT5000数据集：贡献了一个名为VT5000的RGBT图像数据集，包含5000对空间对齐的RGBT图像对，并带有地面真实标注。VT5000数据集考虑了11种挑战，收集了不同场景和环境中的图像，以探索算法的鲁棒性。  提出了一种新的深度CNN架构，作为RGBT显著目标检测的强大基线方法。特别是，提出了使用卷积块注意力来选择性地收集RGB和热红外分支的特征，以及使用边缘损失来指导网络学习更多细节，以保留显著目标的边界。 |  |  | 包括深度学习方法的进一步探索、基于属性的特征表示、无监督和弱监督RGBT SOD以及无需对齐的方法 | 2023  IEEE Transactions on Multimedia |
| **12. Semantic perceptive infrared and visible image fusion Transformer** | 一种称为SePT的**图像融合**新方法，特别关注红外和可见光图像融合 | * 所提出的SePT方法将基于CNN的局部特征提取与基于Transformer的长程依赖学习和语义建模模块相结合。 * 设计了两个基于 Transformer 架构的语义建模模块，用于处理高级语义信息。 * 融合结果由局部特征、长程依赖和语义特征组合而成。 * SePT与其他融合算法进行了评估，并在定性和定量比较中表现出卓越的性能。 | TNO数据集和RoadScene数据集 | **主要贡献**：  提出了SePT，该融合网络通过CNN和变换器块实现局部特征提取、长期依赖性建模和语义信息学习的整合。  设计了两个语义信息建模模块，一个将源图像的浅层特征映射到深层语义，另一个在不同的感受野中学习深层语义信息。  通过与其他复杂的融合算法进行定性和定量的比较评估，证明了SePT的优越性。 |  |  | 未来的工作将致力于通过设计有效的与语义相关的损失函数来提高融合网络的效率，并考虑多通道图像的固有特性，设计具有高泛化性能的通用融合网络。 | 2024  Pattern Recognition |
| **13. Siamese Network for RGB-D Salient Object Detection and Beyond** | 一种新的RGB-D**显著目标检测**（SOD）框架——联合学习和密集合作融合（JL-DCF）架构 | * Siamese网络是一种用于学习两个相似输入的通用特征表示和距离（或相似性）度量的网络 |  | **主要贡献**：  提出了JL-DCF框架，该框架通过Siamese网络利用RGB和深度视图之间的共性，进行联合学习。  JL-DCF在七个挑战性数据集上平均提高了约2.0%的最大F-度量，显著提高了现有最先进模型的性能。JL-DCF不仅适用于RGB-D SOD任务，还可直接应用于其他多模态检测任务，如RGB-T（热红外）SOD和视频SOD，显示出其鲁棒性和通用性。 |  | https://github.com/kerenfu/JLDCF/ | **语义相关损失函数的设计**：尽管当前的工作通过网络设计来学习语义信息，但在损失函数的改进上投入较少。未来的工作将包括设计有效的与语义相关的损失函数，以指导网络训练过程。  **多通道图像的融合网络**：目前的工作主要考虑灰度图像的融合，这使得将成果推广到多通道图像融合领域存在一定难度。 | 2021  IEEE Transactions on Pattern Analysis and Machine Intelligence |
| **14. SwinNet: Swin Transformer Drives Edge-Aware RGB-D and RGB-T Salient Object Detection** | 一种新的RGB-D和RGB-T**显著目标检测**（SOD）模型——SwinNet | * SwinNet包含双流Swin Transformer编码器、空间对齐和通道重校准模块、边缘感知模块和边缘引导的解码器。这些组件共同工作以提取和融合多模态特征，并在边缘信息的指导下锐化显著目标的轮廓。 | VT821、VT1000、  VT5000、  NLPR、  NJU2K 、  STERE 、  DES、  SIP、  DUT | 1. **新颖的SOD模型**：提出了一个名为SwinNet的新型显著目标检测（SOD）模型，该模型适用于RGB-D和RGB-T任务，并且基于Swin Transformer骨干网络构建。 2. **跨模态特征融合**：通过设计空间对齐和通道重校准模块，实现了不同模态特征之间的优化，增强了每种模态的显著内容。 3. **边缘引导的解码器**：引入了边缘引导的解码器，在边缘信息的指导下，实现了跨模态特征的层间融合，生成了更清晰的显著目标轮廓。 |  | https://github.com/liuzywen/SwinNet | 尽管SwinNet在准确性上取得了进步，但也带来了速度的降低。未来的工作将讨论轻量级设计。 | 2022  IEEE Transactions on Circuits and Systems for Video Technology |
| **15. Unified Information Fusion Network for Multi-Modal RGB-D and RGB-T Salient Object Detection** | 多模态**显著目标检测**（SOD）框架——统一信息融合网络（MMNet），用于RGB-D和RGB-T图像的显著目标检测 | * 文章提出了一个统一的端到端框架MMNet，用于同时分析RGB-D和RGB-T SOD任务。该框架包括跨模态多阶段融合模块（CMFM）和双向多尺度解码器（BMD）。 | VT821、VT1000、  VT5000、  NLPR、  NJUD、  LFSD、  STERE、  SSD、  DUT-RGBD | 1. **统一的端到端框架**：提出了一个名为MMNet的统一端到端框架，能够同时处理RGB-D和RGB-T显著目标检测（SOD）任务。 2. **跨模态多阶段融合模块（CMFM）**：设计了一个新颖的CMFM，灵感来源于人类视觉系统（HVS）中的视觉色彩阶段理论。CMFM通过特征响应阶段和对抗性组合阶段，探索不同模态特征的有效性、互补性和准确性。 3. **双向多尺度解码器（BMD）**：引入了BMD，通过双向融合结构合理地结合多尺度层次的跨模态特征，以提高RGB-D和RGB-T SOD性能。 |  |  | **扩展到RGB-D-T SOD**：文章提到了RGB-D和RGB-T SOD任务，但并未详细探讨RGB-D-T（结合深度和热红外信息）的SOD任务。未来的研究可能会探索如何有效地结合RGB、深度和热红外三种模态的信息来进行SOD。 | 2022  IEEE Transactions on Circuits and Systems for Video Technology |
| 16. Fixated Object Detection Based on Saliency Prior in Traffic Scenes | 关于在交通场景中基于显著性先验的凝视对象检测的研究 | * 对象检测模块（Object Detection Module, ODM）：基于YOLOv5-L * 显著区域引导模块 * 显著性引导策略（Saliency Guidance Strategy）： * 损失函数 * 特征融合 * 网络训练：在训练过程中，首先预训练SRGM，然后使用预训练得到的权重来初始化FOD-Net模型，并继续训练直到验证集上的平均精度（mAP）最大化。 | 自建数据集：ETFOD（Eye-Tracking Based Fixated Object Detection dataset）这个数据集是基于多名驾驶员的眼动追踪信息建立的，用于作为未来研究凝视对象检测的基准 | 主要贡献：   * 提供了一个新的基于驾驶员凝视的凝视对象检测数据集（ETFOD），作为研究驾驶员视角下交通对象检测的基准。 * 提出了一个基于显著性先验的凝视对象检测模型，称为FOD-Net。该模型利用预测的显著区域作为显著性先验来指导检测与驾驶任务密切相关的凝视对象，从而提高检测精度。 * 实验结果表明，FOD-Net在所提出的数据集上达到了78.4%的mAP值，模型参数小，优于其他最先进的模型。 |  |  |  | 2024 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |