

제목 : 멀티스펙트럴 영상 기반 연구 분석 및 동향

20110753 김 지 원

I. 서론

인공신경망 기술은 다양한 분야에 적용되며 그 쓰임새가 날로 높아지고 있다. 이러한 인공신경망 기술 발전의 주요 요인으로서는 ‘GPU의 비약적인 성능 향상’과 ‘다양하고 방대한 규모의 데이터셋의 공개’ 등을 꼽을 수 있다. 특히, 스탠포드 대학 Fei-Fei Li 교수 연구팀의 이미지넷(ImageNet)[1], 마이크로 소프트社의 MSCOCO[2] 등 다양하고 방대한 규모의 데이터셋이 공개되며 누구나 인공신경망 연구를 수행할 수 있게 됐다. 하지만 대다수의 대규모 공개 데이터셋은 RGB 영상으로만 구성됐다는 한계가 존재한다. 따라서 높은 인지 정확도를 요구하는 자율주행 분야에서는 RGB로 구성된 기존 공개 데이터셋의 한계를 극복하기 위해 다양한 센서 데이터를 포함하는 데이터셋을 속속히 공개하였다. 대표적으로는 튀빙겐(tuebingen) 대학 Andreas Geiger 교수 연구팀의 KITTI[3], 자율주행 업체 앵티브(Aptiv)에서 공개한 nuScene [4]등이 있으며, 이러한 데이터셋에는 RGB 영상 뿐만 아니라 라이다(LiDAR), 레이더(RADAR) 등 다양한 센서 데이터가 포함되어 있어 다중 센서 기반 인공신경망 연구가 가능해졌다.

이러한 흐름 속에서 우리나라 KAIST의 All-day Vision 연구팀은 열 에너지를 감지하는 열화상 카메라가 포함된 ‘KAIST Multispectral Pedestrian Detection Dataset[5]’을 처음으로 공개하며, 자율주행 분야에서 멀티스펙트럴 영상 기반 환경강인한 보행자 인식 연구의 포문을 열었다. 여기서 멀티스펙트럴 영상이란 가시광선 파장의 RGB 영상 적외선 파장의 열화상 영상을 융합하여 사용하는 것을 의미한다. 멀티스펙트럴 영상은 객체인식 및 분류에서 높은 정확도를 갖는 RGB 영상의 장점과 조도변화, 악조건등에서 강인한 객체 검출이 가능한 열화상 영상의 장점을 모두 갖추어 많은 자율주행 연구자들에게 각광받고 있다.

본 보고서에서는 인공신경망을 이용하여 멀티스펙트럴 영상 융합하고 영상 속 보행자를 검출하는 ‘멀티스펙트럴 영상 기반 보행자 인식’ 연구의 흐름을 살펴보고, 보행자 인식 외에 멀티스펙트럴 영상이 사용된 다양한 인공신경망 기반 연구들을 알아보고자 한다.

II. 멀티스펙트럴 영상 기반 보행자 인식 연구 동향

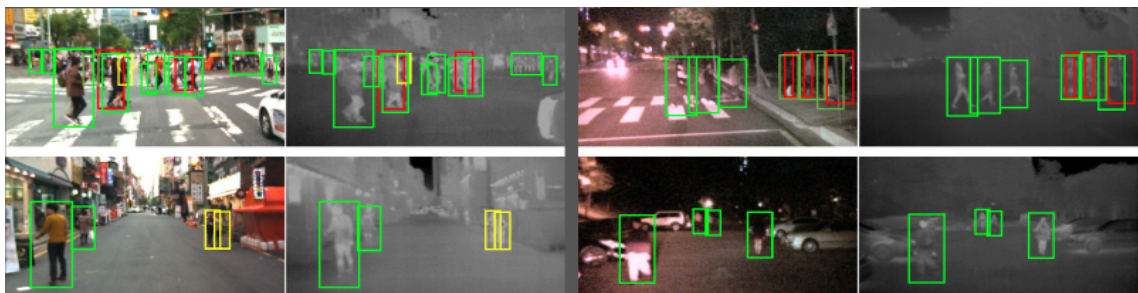


그림 1 카이스트 멀티스펙트럴 보행자 인식 벤치마크 데이터셋 예시

2015년 컴퓨터 비전 분야 최고 학술대회인 CVPR(Conference on Computer Vision and Pattern Recognition)에서 카이스트 All-day Vision 연구팀이 빔스플릿트를 이용하여 취득한

카이스트 멀티스펙트럴 보행자 벤치마크 데이터셋[5]을 공개하며, 멀티스펙트럴 영상 기반 보행자 인식 연구는 시작됐다. 데이터셋이 공개된 다음 해인 2016년, Jingjing Liu는 영국에서 열린 머신 비전 학회인 BMVC(British Machine Vision Conference)에서 당시 최고 성능을 달성한 인공지능망 기반 객체 검출 알고리즘인 Faster R-CNN을 기반으로 멀티스펙트럴 영상 기반 보행자 검출 알고리즘[6]을 제안하였다. 해당 논문은 다양한 특징 레벨 (Feature-level)에서 멀티스펙트럴 영상 특징의 융합을 비교하였고, 중간 레벨에서의 특징 융합(Halfway Fusion)이 가장 좋은 성능이 나타남을 실험적으로 나타냈다.

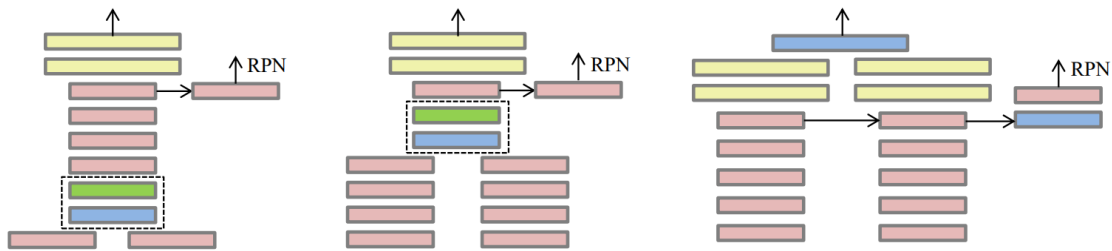


그림 2 Jingjing Liu가 제안한 Faster R-CNN 기반 멀티스펙트럴 영상 융합 모델 아키텍처

2018년 Dayan Guan는 멀티스펙트럴 영상을 융합할 때, 입력된 영상의 조도 환경에 따라 유동적인 가중치를 계산하는 보조 네트워크[7]를 제안하였다. 이를 통해 조도 환경이 풍부한 상황에서는 RGB 영상, 조도 환경이 열악한 상황에서는 열화상 영상의 가중치를 더 많이 부여하여 보행자 인식 성능을 크게 향상시켰다.

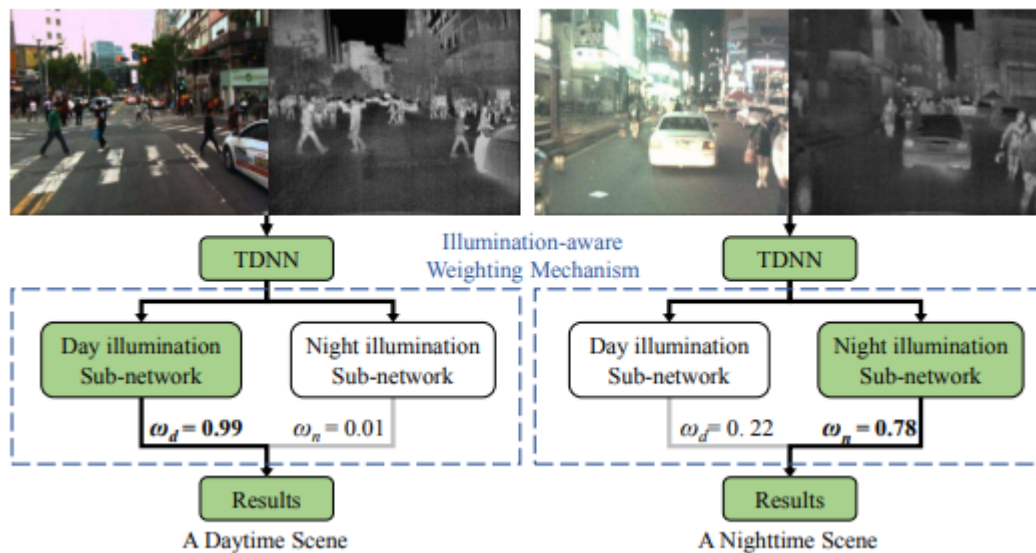


그림 3 Dayan Guan이 제안한 조도 환경에 따라 다른 가중치를 계산하는 보조 네트워크

2018년 열린 BMVC 학회에서 Chengyang Li는 기존 멀티스펙트럴 영상 기반 보행자 인식 연구에서는 RGB 영상과 열화상 영상의 전역 특징 레벨에서의 융합을 고려하고 있는 점의 문제를 제기하며, Faster R-CNN의 RPN(Region Proposal Network)에서 추출된 관심영역

(Region of Interest, ROI)에 대해서 지역 특징 레벨에서의 특징도 함께 융합하는 알고리즘 [8]을 제안하였다. 또한 해당 연구에서는 기존 카이스트 멀티스펙트럴 보행자 벤치마크 데이터셋의 어노테이션(Annotation)의 부정확성을 이야기하며 정확도를 개선한 새로운 어노테이션을 공개하였다.

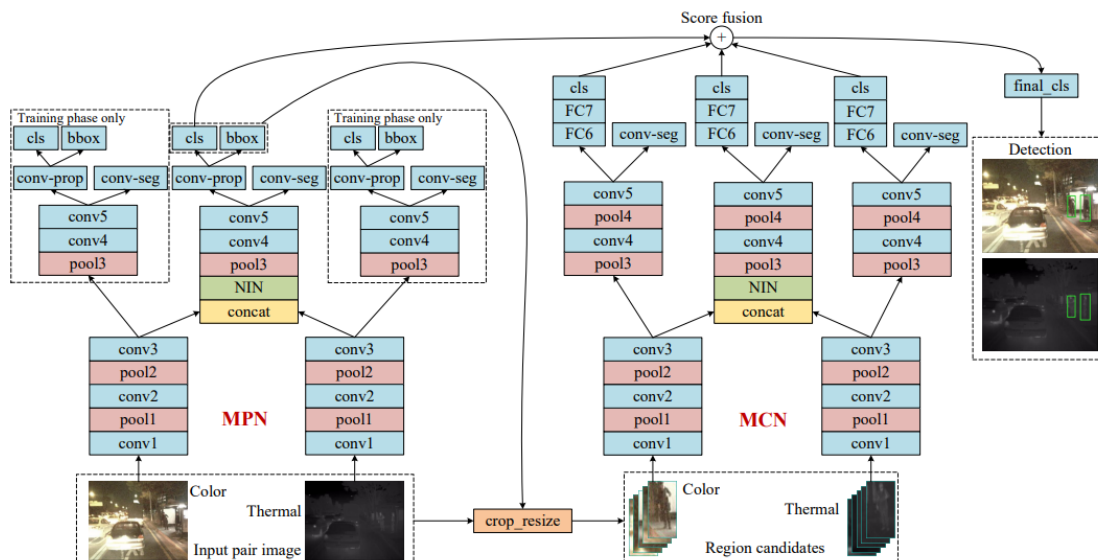


그림 4 Chengyang Li이 제안한 전역-지역 특징 레벨 융합을 통한 보행자 인식 네트워크

2019년 서울에서 열린 국제컴퓨터비전학회 ICCV(International Conference on Computer Vision)에서 Lu Zhang은 멀티스펙트럴 영상을 취득하는 과정에서 발생하는 오정렬(Miss alignment) 문제를 처음으로 제기하며, 약한 오정렬 문제를 해결하기 위한 Region Feature Alignment (RFA) 모듈[9]을 제안하였다.

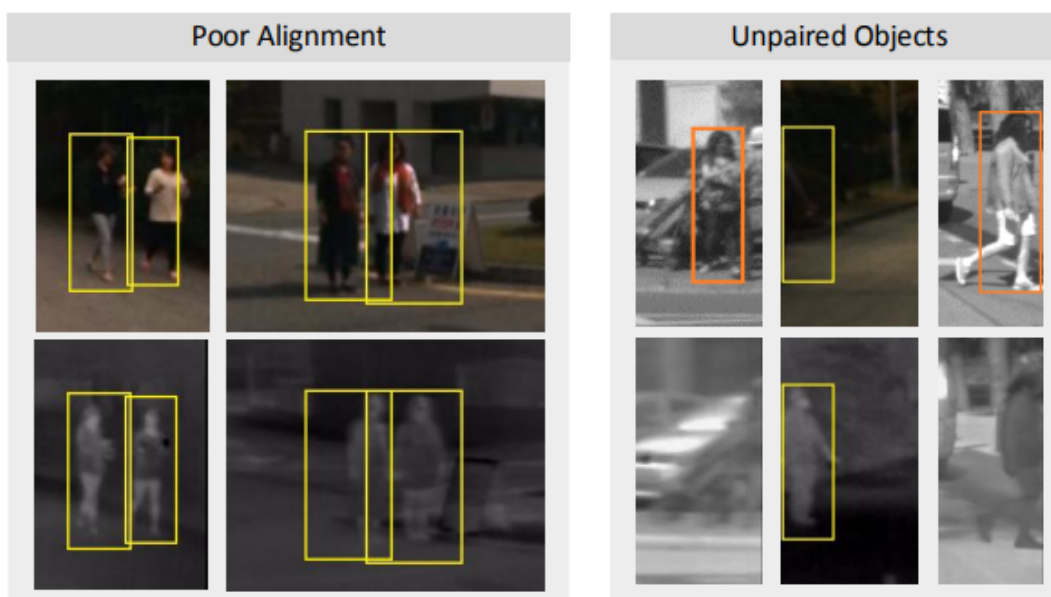


그림 5 Lu Zhang이 제기한 멀티스펙트럴 영상에서의 오정렬 예시

2020년 CVPR, ICCV와 함께 세계 3대 컴퓨터 비전학회인 ECCV(European Conference on Computer Vision)에서 Kailai Zhou 는 멀티스펙트럴 영상에서 RGB 영상과 열화상 영상이 다른 파장대 영역에서 취득돼 나타나는 모달리티 불균형 문제(Modality imbalance problem)를 해결하기 위한 Differential Modality Aware Fusion(DMAF) 모듈[10]을 제안하였다. 그림 6에서는 DMAF를 적용하여 모달리티의 밸런스 문제를 줄여 두 영상간 유사한 피쳐맵이 나타남을 확인할 수 있다.

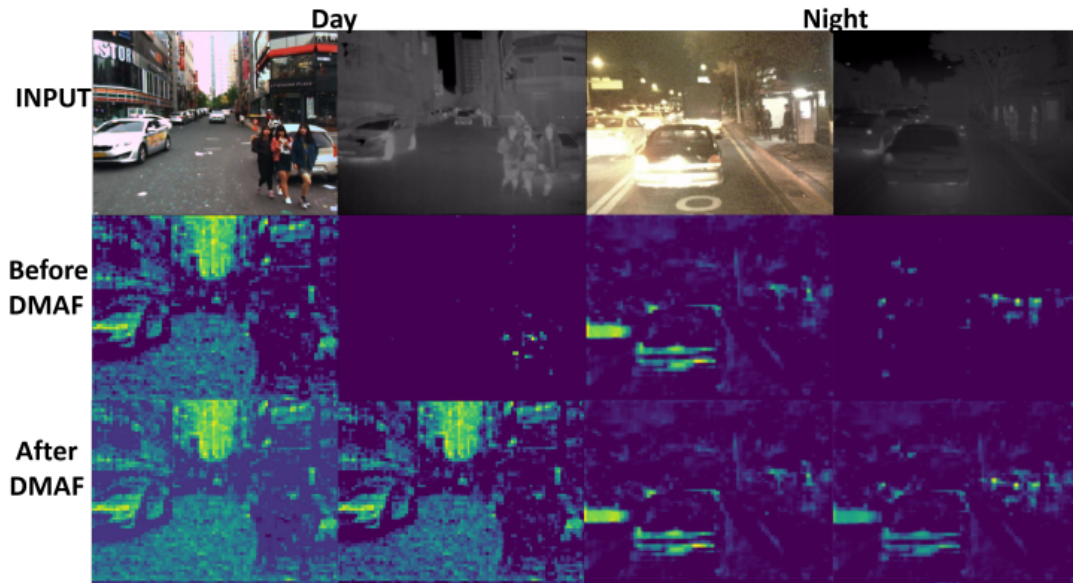


그림 6 Kailai Zhou가 제안한 DMAF를 적용 전/후 특징맵 비교

2021년 세계 최고 권위를 자랑하는 로보틱스 학회 중 하나인 IROS에서 세종대학교 RCV 연구팀은 기존 연구가 가정하고 있는 RGB 영상과 열화상 영상이 온전히 오버랩된다는 부분에 대해서 문제를 제기하였다. 실제 멀티스펙트럴 영상을 취득시 RGB, 열화상 센서의 화각 및 베이스라인의 차이로 인하여 온전히 오버랩된 멀티스펙트럴 영상을 취득하기 어려우며, 이러한 상황에서 기존 제안된 알고리즘들은 제 기능을 발휘하기 어려움을 지적하였다. 따라서 해당 연구에서는 이를 해결하기위해 멀티레이블을 적용한 새로운 멀티스펙트럴 영상 기반 보행자 인식 알고리즘[11]을 제안하였다.



그림 7 기존 연구와의 비교 (순서) MSDS, AR-CNN, MBNet, 제안하는 연구, Ground Truth

앞서 설명한 연구 이외에도 다양한 학회 및 저널에 멀티스펙트럴 영상 기반 보행자 인식을 위한 알고리즘들이 제안됐으며, 지금도 멀티스펙트럴 영상을 이용하여 조도 변화 및 악조건의 기상환경에서도 강인한 보행자 인식을 통해 자율주행차의 신뢰성을 높이기 위한 연구가 활발히 진행되고 있다.

III. 멀티스펙트럴 영상 기반 보행자 인식 외 연구 동향

II 장에서는 멀티스펙트럴 영상 기반 보행자 인식 연구들에 대해 소개하였다면, III 장에서는 보행자 인식 이외에 다양한 멀티스펙트럴 영상 기반 인공지능 연구들에 대해 소개하고자 한다.

1. 멀티스펙트럴 영상 기반 의미론적 영상분할 연구

객체검출(Object Detection)과 더불어 의미론적 영상분할(Semantic Segmentation)은 컴퓨터비전에서 가장 활발히 이뤄지는 연구분야이다. 따라서 멀티스펙트럴 영상을 기반으로 의미론적 영상분할을 진행하는 연구들도 활발히 이뤄지고 있다. 2019년 Robotics and Automation Letters(RA-L)에는 멀티스펙트럴 영상을 사용하여 조도 변화에 강인한 의미론적 영상분할을 수행하는 RTFNet(RGB-Thermal Fusion Network)[12]가 제안됐다. 해당 네트워크는 인코더-디코더로 구성된 전체 네트워크에서 열화상 영상의 특징을 각 스테이지마다 RGB 영상 특징에 Element-wise Summation를 이용하여 융합을 진행하였다. 이를 통해 조도가 어두운 환경에서도 강인한 의미론적 분할이 가능함을 나타냈다.

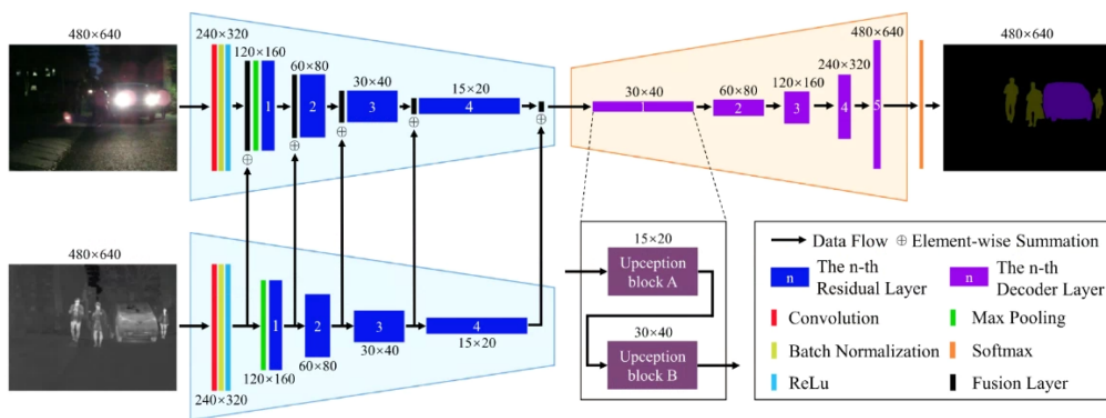


그림 8 멀티스펙트럴 영상 기반 의미론적 영상분할 네트워크 (RTFNet)

2. 멀티스펙트럴 영상 데이터셋의 부족을 해결하기 위한 연구

다양한 환경에서 강인성을 확보하기 위해 멀티스펙트럴 영상은 수많은 컴퓨터비전 분야에 적용되고있지만, 멀티스펙트럴 영상 연구는 적은 데이터셋으로 인해 RGB 영상 기반 연구 대비

범용성이 낮다는 한계가 존재한다. 최근 이러한 문제를 해결하기 위한 연구들이 진행되고 있다. 2019년 국제저명학회 ICRA(International Conference on Robotics and Automation)에서 세종대학교 RCV 연구팀은 RGB 영상과 열화상 영상의 중간 도메인 이미지(Multispectral Domain Invariant Image, MDII)를 만드는 네트워크[13]를 제안하였다. 이를 통해 RGB 영상 기반 연구에 직접적으로 열화상 영상을 사용할 수 있도록 두 영상차이를 줄이고, 검색기반 위치인식 분야에 적용하여 만들어진 중간 도메인 영상의 효율성을 입증하였다.

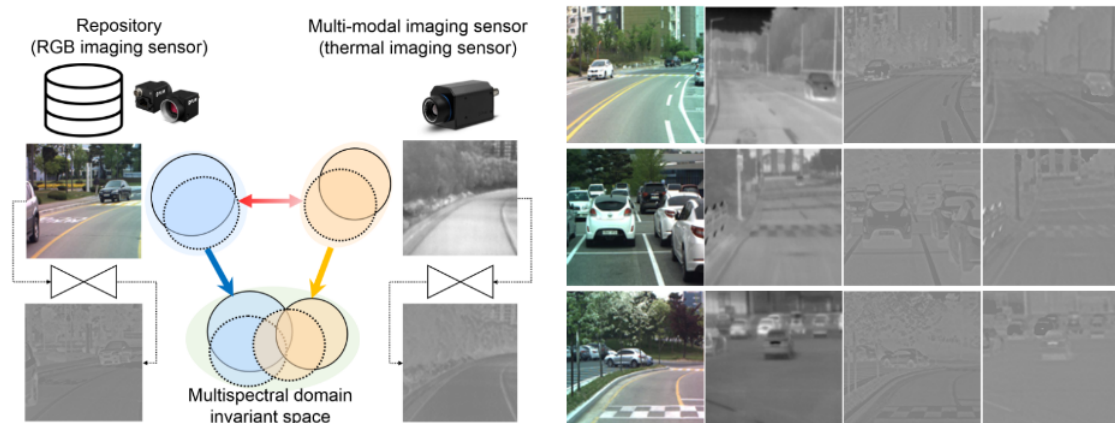


그림 9 좌) MDII 제작을 위한 프로세스, 우) 중간 도메인 생성 영상 (3,4열)

멀티스펙트럴 영상 기반 의미론적 영상분할 연구에서도 열화상 영상의 데이터셋 부족을 극복하기 위해 2021년 RA-L에서 카이스트 RCV 연구팀은 비지도학습 기반 도메인 적응 방법(Multi-Spectral Unsupervised Domain Adaptation, MS-UDA)[14]을 제안하였다. 해당 연구는 풍부한 RGB 영상을 기반으로 학습된 의미론적 영상분할 네트워크를 이용하여 멀티스펙트럴 영상에서 어노테이션 없이 의미론적 영상분할 네트워크를 학습한다. 이를 통해 멀티스펙트럴 영상에서의 데이터셋 부족 문제를 해결할 수 있음을 실험을 통해 입증하였다.

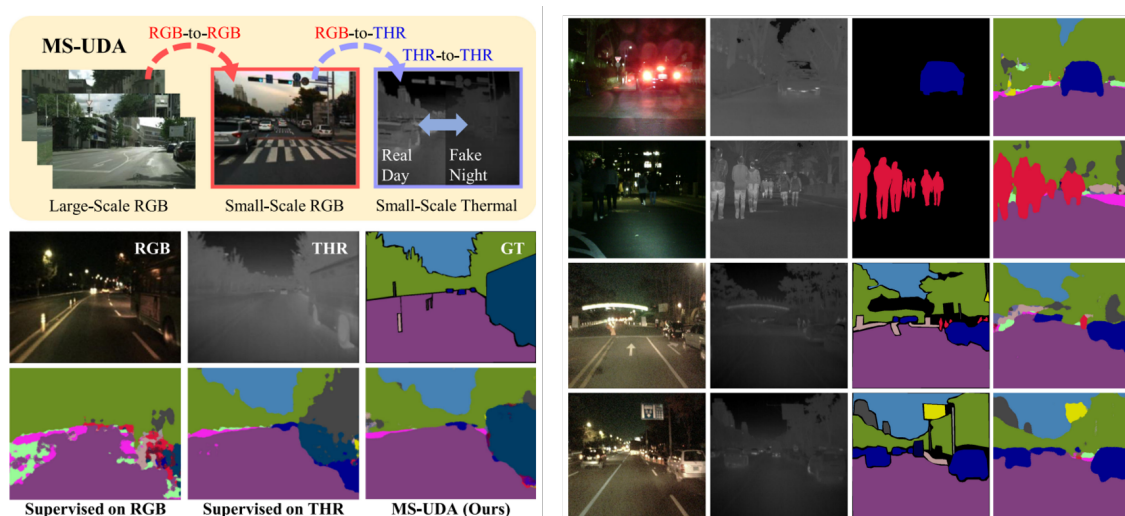


그림 10 좌) MS-UDA 프레임워크, 우) 정성적 결과 (RGB, Thermal, GT, Predicted)

IV. 결론 및 시사점

본 보고서에서는 멀티스펙트럴 영상 기반 연구들을 시간별로 분석하였다. 멀티스펙트럴 영상 기반 연구는 세계 3대 컴퓨터 비전학회(CVPR, ICCV, ECCV), 세계 최고 권위를 자랑하는 로봇틱스 학회(ICRA, IROS) 등 다양한 저명 학회에서 논문이 매년 발표될 정도로 많은 연구가 이뤄졌으며, 해를 거듭할수록 멀티스펙트럴 영상을 활용하는 연구자 풀이 넓어지고 있는 추세이다. 대한민국이 멀티스펙트럴 영상 기반 연구의 포문을 연 만큼 앞으로도 멀티스펙트럴 영상 연구에 대한 지속적인 관심이 필요하다.

참고문헌

- [1] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.
- [2] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.
- [3] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." The International Journal of Robotics Research 32.11 (2013): 1231-1237.
- [4] Caesar, Holger, et al. "nusenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [5] Hwang, Soonmin, et al. "Multispectral pedestrian detection: Benchmark dataset and baseline." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [6] Liu, Jingjing, et al. "Multispectral deep neural networks for pedestrian detection." arXiv preprint arXiv:1611.02644 (2016).
- [7] Guan, Dayan, et al. "Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection." Information Fusion 50 (2019): 148-157.
- [8] Li, Chengyang, et al. "Multispectral pedestrian detection via simultaneous detection and segmentation." arXiv preprint arXiv:1808.04818 (2018).
- [9] Zhang, Lu, et al. "Weakly aligned cross-modal learning for multispectral pedestrian detection." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [10] Zhou, Kailai, Linsen Chen, and Xun Cao. "Improving multispectral pedestrian detection by addressing modality imbalance problems." Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XVIII 16. Springer International Publishing, 2020.
- [11] Kim, Jiwon, et al. "MLPD: Multi-Label Pedestrian Detector in Multispectral Domain." IEEE Robotics and Automation Letters 6.4 (2021): 7846-7853.
- [12] Sun, Yuxiang, Weixun Zuo, and Ming Liu. "Rtfnet: Rgb-thermal fusion network

for semantic segmentation of urban scenes." IEEE Robotics and Automation Letters 4.3 (2019): 2576-2583.

[13] Han, Daechan, et al. "Multispectral Domain Invariant Image for Retrieval-based Place Recognition." 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020.

[14] Kim, Yeong-Hyeon, et al. "MS-UDA: Multi-Spectral Unsupervised Domain Adaptation for Thermal Image Semantic Segmentation." IEEE Robotics and Automation Letters 6.4 (2021): 6497-6504.