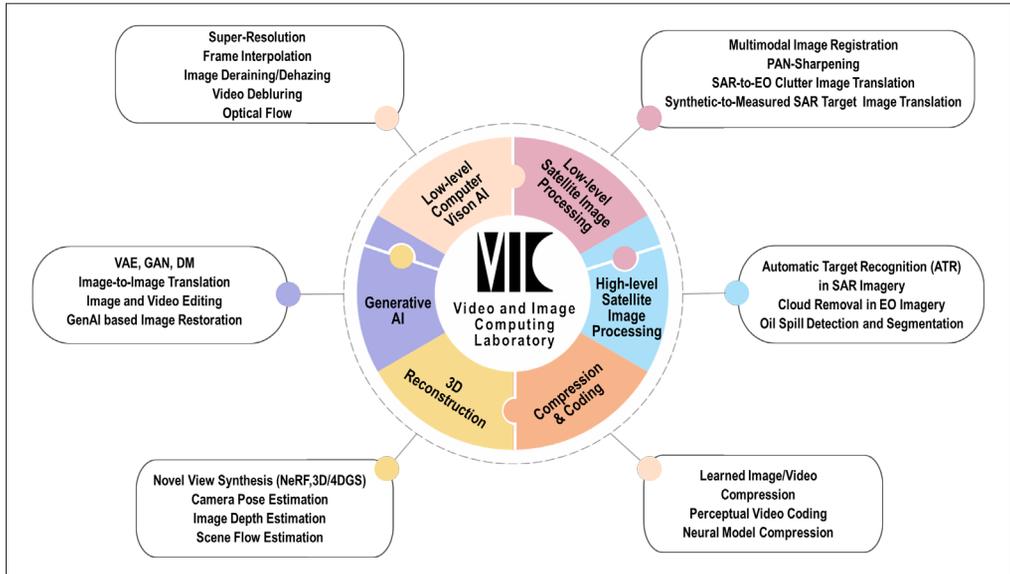


KAIST VICLAB

김문철 교수 / KAIST 전기및전자공학부

I . KAIST VICLAB(Video and Image Computing Lab) 현황

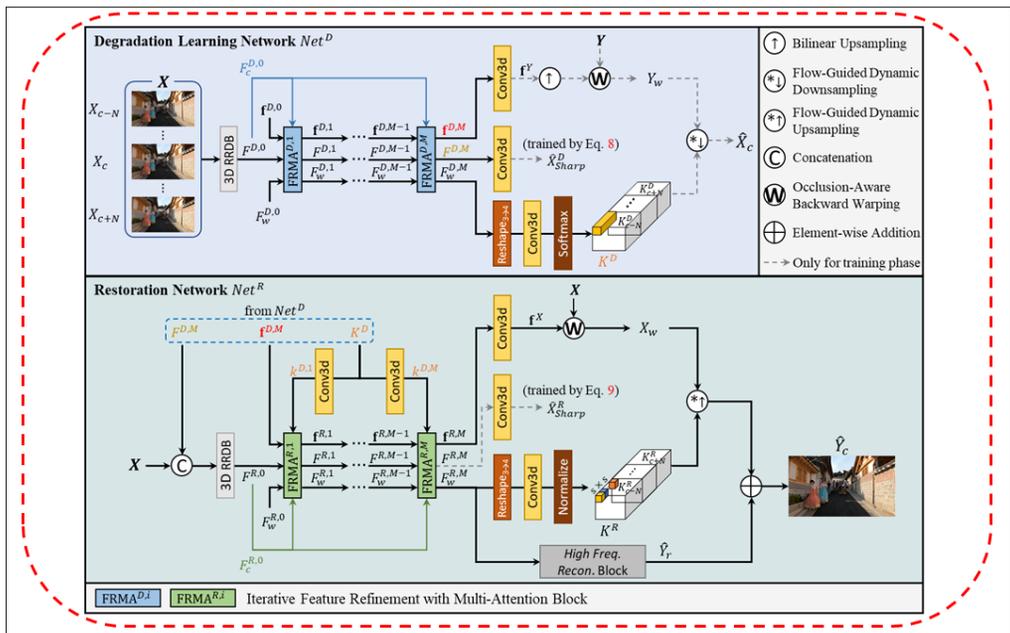
VICLAB은 영상 처리 및 컴퓨터 비전 분야에서 다양한 연구를 활발히 수행하고 있으며, 현재 IITP 지원 “SW 스타랩” 및 “ICT 기초 연구실” 과제와 연구재단 지원 “개인 기초 중견 연구” 과제 및 산업체 공동 연구 등을 수행해 오고 있다. 최근 크게 3가지 분야 연구에 집중해 오고 있으며, 첫째는 자연 영상 생성, 편집 및 복원 분야로서 generative video editing, video colorization, super-resolution, frame interpolation, SDR-to-HDR inverse tone mapping, image in-painting, depth estimation, image deraining, image dehazing, video motion deblurring, generative restoration of old photos 등이 있으며, 두 번째 연구 분야로는 3D 영상 복원 및 재현 분야로서 optical flow estimation, camera pose estimation, dynamic neural radiance field(NeRF) and Gaussian splatting learning of video for novel view synthesis 등이 있으며, 세 번째는 위성 영상 처리 및 분석으로서 PAN sharpening, super-resolution and cloud removal of Electro-Optical(EO) images, super-resolution, detection and classification of Synthetic Aperture Radar(SAR) image targets, SAR-to-EO image-to-image translation learning 등의 연구 주제를 다루고 있다. 각 연구 분야에서 도출된 성과를 국제 최우수 학회인 CVPR, ICCV, ECCV, NeurIPS, ICLR, AAAI 등에 발표하고 IEEE 저널 등 국제 유수 학술지에 논문을 게재해 오고 있다. 나아가, 연구실 소속 학생들은 이러한 성과를 바탕으로 Google, Adobe Research, Meta 등과 같은 Global Big-tech 기업에서 연구 인턴 등을 통해 글로벌 연구 수행 역량을 높여가고 있다. 본 KAIST VICLAB 소개에서는 위 3개 연구 분야에서 가장 최근에 대표적으로 수행된 세부 연구 주제들에 대해 간략히 소개하고자 한다.



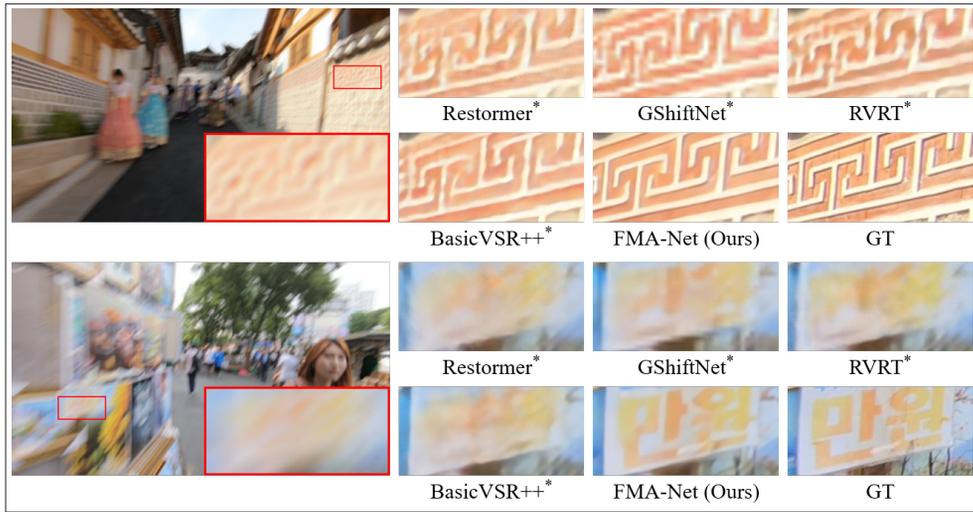
<그림 1> KAIST VICLAB 연구 분야

1. Joint Video Super-Resolution and Deblurring 연구

저해상도의 비디오로부터 고해상도 비디오를 복원하는 비디오 초해상화 방법은 최근까지 많이 연구되어 왔다. 일반적인 딥러닝 기반 비디오 초해상화 모델은 선명한 저해상도 비디오에 대해 학습되었기 때문에, 움직임에 의한 흐림



<그림 2> Joint Video Super-Resolution and Deblurring, CVPR 2024 (Oral)

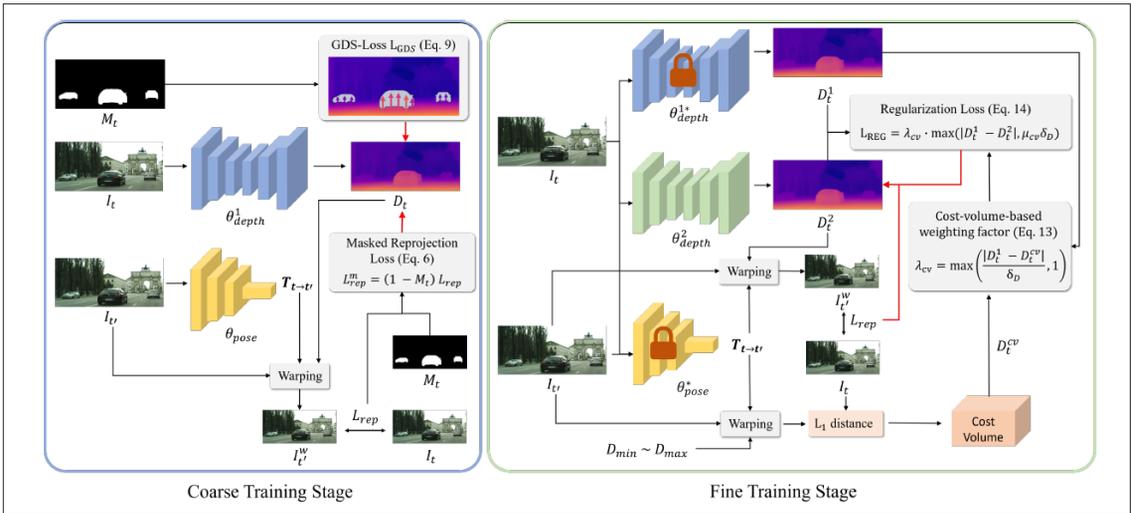


<그림 2-1> Joint Video Super-Resolution and Deblurring 결과 (Project Page: <https://kaist-viclab.github.io/fmanet-site/>)

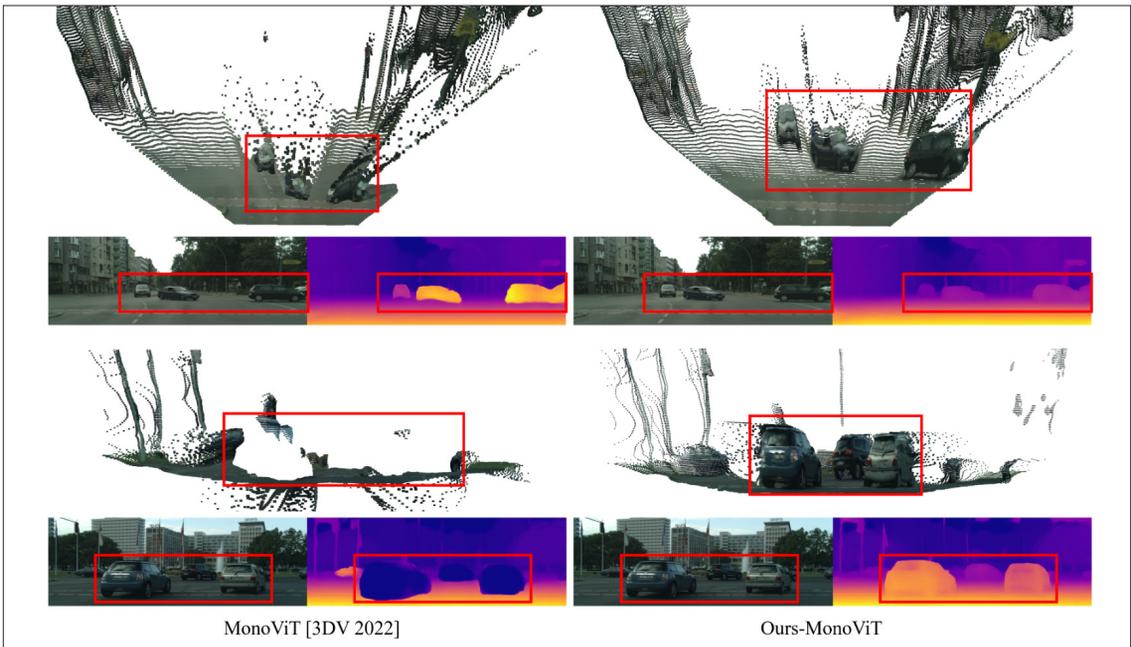
(motion blur) 왜곡이 포함된 비디오에 대해 효과적으로 성능을 발휘하지 못한다. 실제 상황에서 획득되는 대부분의 비디오는 물체 혹은 카메라의 움직임에 의해 발생한 움직임 흐림에 의한 화질 저하를 동반하기 때문에, 이를 해결하기 위해 비디오 초해상화 및 움직임 흐림 제거를 동시에 처리하는 연구가 최근 활발히 이루어지고 있다. 본 연구실은 딥러닝 기술을 활용하여 움직임 흐림이 포함된 저해상도 비디오에서 선명한 고해상도 비디오를 복원하는 기술을 연구해 오고 있다. 이는 단순히 해상도를 높이는 것을 넘어, 영상의 시간적 일관성을 유지하면서 움직임 흐림을 효과적으로 제거하는 복합적인 문제를 해결하는 것을 목표로 한다. 본 연구의 최근 성과는 CVPR 2024에 구두(Oral) 논문(제출 논문 중 0.72% 선정)으로 발표되었다.

2. Self-supervised Monocular Depth Estimation 연구

2D 영상에서 3D 장면을 이해하기 위한 중요한 정보인 깊이는 센서를 통해 얻기 어렵기 때문에 RGB 영상으로부터 영상의 깊이(image depth) 정보를 예측하는 연구가 활발히 진행되고 있다. 이때 단안(Monocular) RGB 영상에서 두 프레임 간의 시차(disparity)를 이용해 깊이를 학습하는 자기지도 단안 깊이 예측(Self-supervised Monocular Depth Estimation) 기술은 매우 정밀한 깊이 정보를 얻을 수 있기에 실용적인 응용 가능성이 매우 크다. 그러나 정밀한 깊이 정보를 얻는데 있어 가장 어려운 점 중의 하나는, 동적 물체(dynamic objects)가 실제 깊이에 따른 시차와 다른 시차를 발생시켜 정확한 깊이 값 학습에 방해가 된다는 점이다. 이에 본 연구실에서는 지면 접지 사전 정보(ground-contacting prior)를 활용한 coarse-to-fine 학습 기법을 사용하여 동적 물체 영역에 대해 정확한 깊이 학습을 유도하는 방법을 연구하였다. 본 연구에서는 동적 물체의 전반적인 깊이 정보를 그와 인접한 지면의 깊이를 활용하여 학습한 뒤, 점진적으로 세부적인 깊이 정보를 정교하게 학습한다. 본 연구는 다양한 종류의 딥러닝 네트워크에 적용이 가능하여 동적 물체에 대한 정확한 깊이 정보가 필요한 다양한 분야에서 활용될 가능성이 높다. 본 연구 성과는 CVPR 2024에 발표되었다.



<그림 3> Self-supervised Monocular Depth Estimation, CVPR 2024

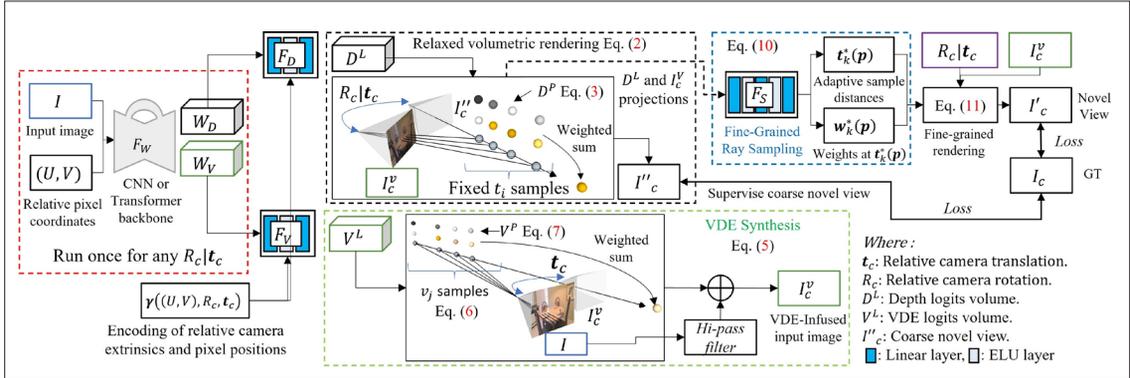


<그림 3-1> Self-supervised Monocular Depth Estimation 결과 (Project Page: https://kaist-viclab.github.io/From_Ground_To_Objects_site/)

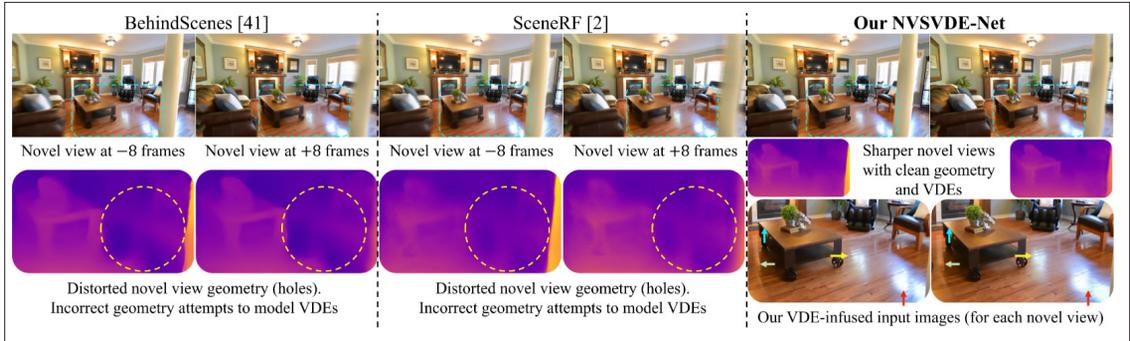
3. 시점 의존 효과를 고려한 시점 합성 연구

주어진 RGB 영상에서 깊이 정보를 이해하여 새로운 시점(view point) 영상을 합성하는 기술(NVS: Novel View Synthesis)은 최근에 많이 연구되고 있는 주제이다. 시차(disparity) 정보를 이용하여 깊이 정보를 학습하는 Self-supervised Image Depth Estimation(SIDE)을 이용했을 때, 반사율이 높은 물체에서 반사되는 부분은 적절한

깊이 정보가 예측되기 어렵고, 이는 결국 부정확한 시점 합성(view synthesis)으로 이루어지게 된다. 이와 같은 현상을 관찰자의 시점에 따라 대상이 다르게 인식되는 시점 의존 효과(VDE: view-dependent effect)라 한다. 일반적인 VDE 효과는 반사광이나 투명한 물체를 볼 때 두드러진다. 본 연구실에서는 VDE를 고려한 정확한 시점 합성을 위하여 카메라 움직임을 활용한 방법을 연구하고 있다. 카메라 움직임을 모델링하여 이를 시점 합성에 반영하면 정확한 깊이 정보 학습 및 시점 합성을 유도할 수 있다. 본 연구는 현실감을 높이는 증강 현실, 가상 현실 구현 등 다양한 3D 응용 기술에 적용될 수 있다. 본 연구 성과는 CVPR 2024에 발표되었다.



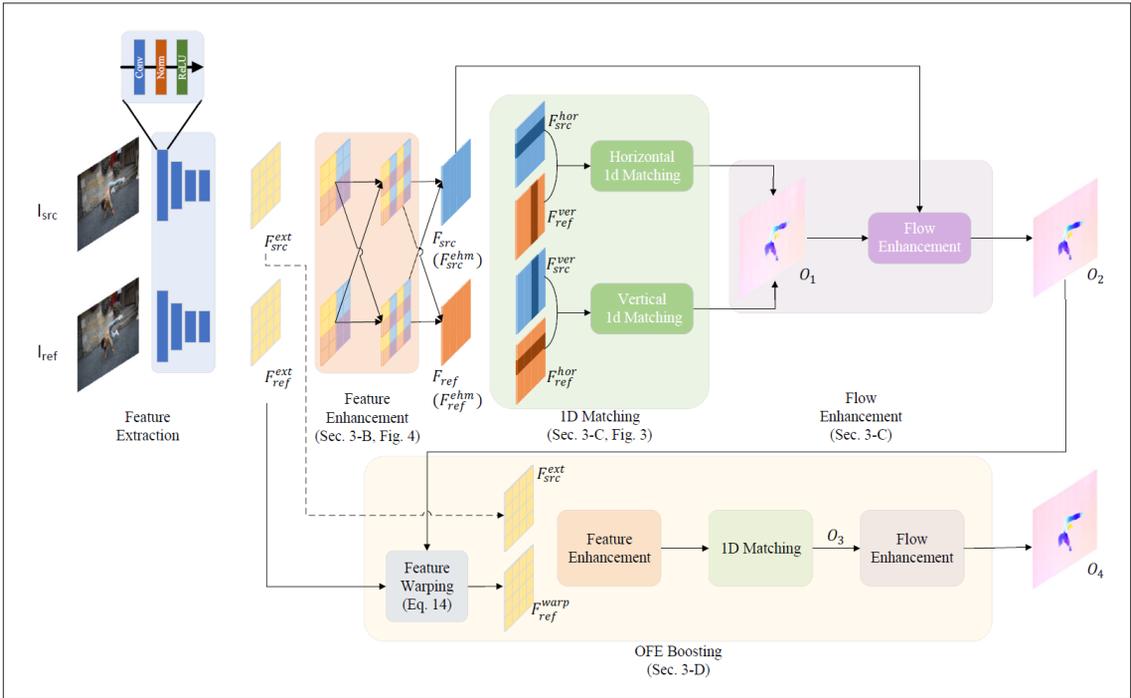
<그림 4> 시점 의존 효과를 고려한 시점 합성, CVPR 2024



<그림 4-1> 시점 의존 효과를 고려한 시점 합성 결과 (Project Page: <https://kaist-viclab.github.io/monovde-site/>)

4. 저복잡도 광학 흐름 예측(optical flow estimation) 연구

광학 흐름(optical flow)은 연속된 영상에서의 고해상도, 프레임 보간, 영상 생성, 행동 인식 등 다양한 분야에 활용되는 정보이다. 최근 광학 흐름 예측 연구는 복잡한 트랜스포머 구조 및 GRU(Gated Recursive Unit)의 재귀적 추론으로 높은 성능을 달성하였다. 하지만, 이러한 방식을 실제 응용에 활용하려면 큰 메모리와 높은 연산 복잡도를 필요로 하는 한계가 존재한다. 본 연구실에서는 이러한 큰 메모리와 높은 연산 복잡도를 획기적으로 줄이기 위해, 1차원 매칭 및 다중축 트랜스포머를 활용해 입력 영상 해상도의 제공에 비례하던 기존 연구들의 공간 복잡도와 GRU의 재귀적 추론으로 인한 높은 연산 복잡도를 크게 낮추었다.



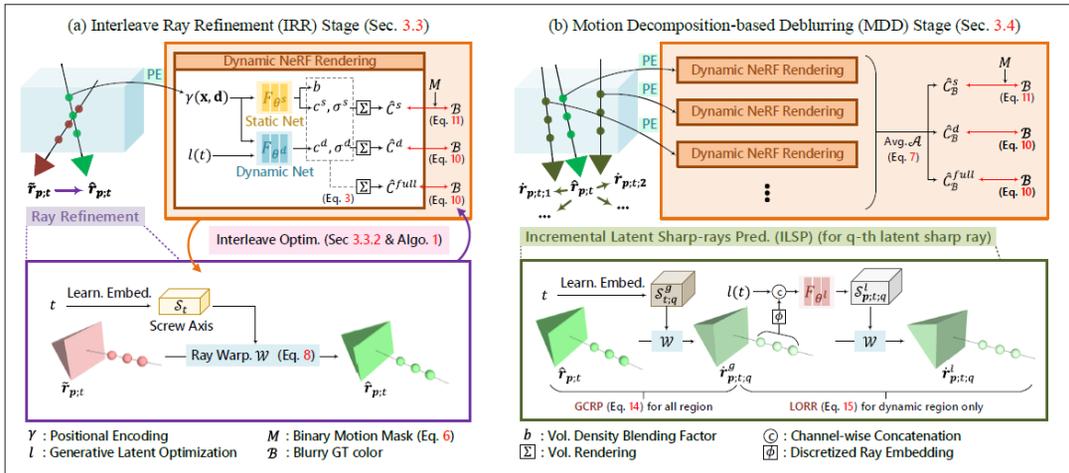
<그림 5> 저복잡도 optical flow estimation, IEEE ACCESS, 2024



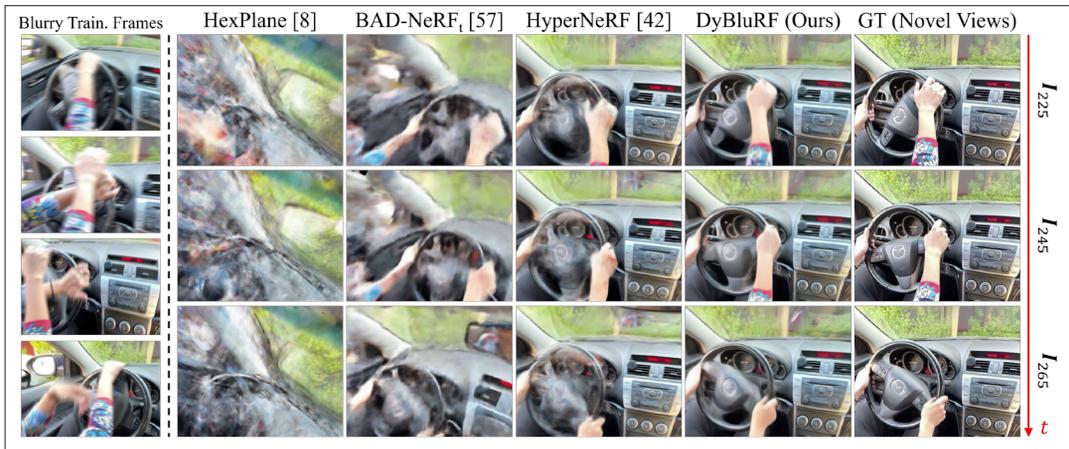
<그림 5-1> 저복잡도 optical flow estimation 결과

5. Dynamic Deblurring Neural Radiance Fields 연구

Novel View Synthesis(NVS)는 임의의 시점 및 시간에서 시각적으로 자연스러운 비디오 프레임을 생성하여 몰입감 있는 시청 경험을 제공할 수 있다. 비디오 NVS의 경우 노출 시간 동안 객체나 카메라의 움직임으로 인해 발생하는 움직임 흐림(motion blur)이 존재하기 때문에 정확한 시점의 합성에 어려움이 있다. 이를 해결하기 위해 본 연구실에서는 움직임 흐림을 고려하여 비디오 NVS를 최적화하는 기술을 최초로 개발하였다. 해당 기술은 크게 두 단계로 이루어져 있으며, 1단계에서는 움직임 흐림으로 인해 발생하는 카메라 오차를 최적화하고, 2단계에서는 이를 바탕으로 각 객체와 카메라의 움직임 흐림을 모델링하여 선명한 색상(RGB) 값을 추정한다. 본 문제를 풀기 위해 움직임 흐림이 포함된 NVS 데이터셋을 새롭게 구성하였고, 이에 대해 최신 비교 모델과 비교한 결과 모두 우수한 성능의 NVS 결과를 도출하였다.



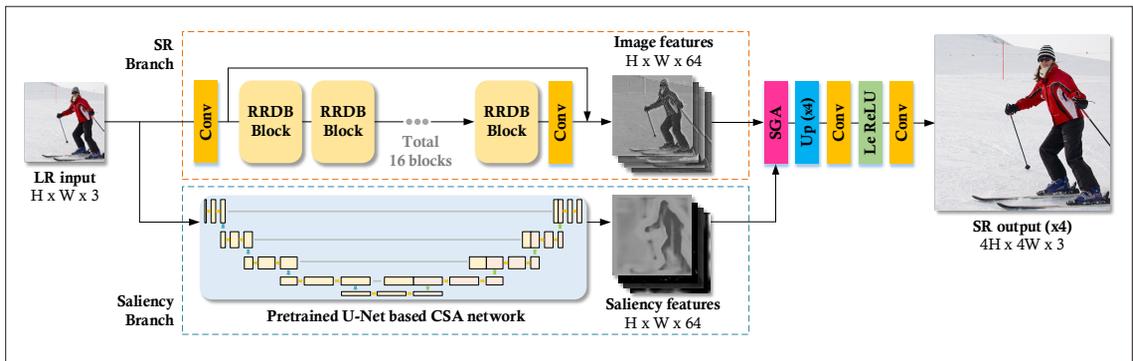
<그림 6> 움직임 흐림 입력 비디오에 강건한 NeRF 기반 Novel View Synthesis



<그림 6-1> 움직임 흐림 입력 비디오에 강건한 NeRF 기반 Novel View Synthesis 결과 (Project page: <https://kaist-viclab.github.io/dyblurf-site/>)

6. Saliency-Guided Image Super-Resolution 연구

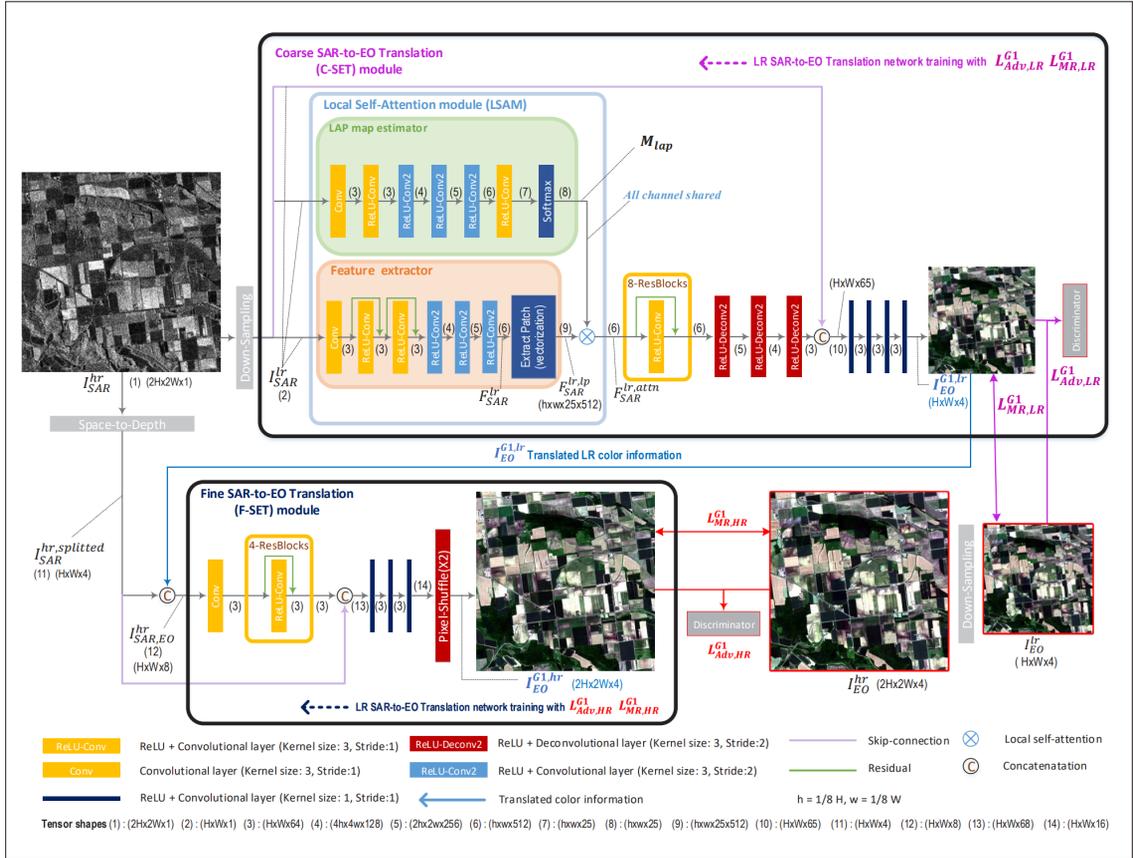
저해상도의 영상으로부터 고해상도의 영상을 복원하는 이미지 초해상화(Super-Resolution, SR)는 컴퓨터 비전 분야의 주요 과제이다. 최근 딥러닝 기술의 발전과 함께 다양한 기법들이 제안되고 있으며, 생성형 모델 기반의 기법들이 각광을 받고 있다. 하지만 이런 방법들은 고주파 성분의 질감과 구조를 복원해내는 동시에 불필요한 구조까지 과도하게 생성하여 artifact를 유발하고 원본 영상과의 유사도를 떨어뜨린다는 문제점이 있다. 따라서 본 연구에서는 일반적으로 사람들이 영상을 볼 때 중요하게 인식하는 부분에 집중한다는 점에 착안하여, 영상 내 유표(salient) 객체에 집중하는 이미지 초해상화 모델(Saliency-Guided Image Super-Resolution, SGSR)을 개발하였다. 영상의 모든 영역에 대하여 동등한 품질로 초해상화 복원을 학습하는 기존의 방법들과는 달리, 제안하는 모델은 영역의 유표도에 따라 차별적인 초해상화를 수행한다. 이를 위해, 객체 영역에 대한 새로운 유표도 스코어를 이용한 학습 방식을 제안한다. 또한, saliency-guided attention(SGA)과 saliency-weighted loss를 사용하여 주목할 만한 영역의 이미지 품질을 극대화할 수 있도록 하였다. 결과적으로 차별적 생성 네트워크를 통해 초해상화된 영상은 시각적으로 더 좋은 화질을 나타내는 효과를 가지며, 제안하는 SGSR은 자연 영상 SR 분야에서 유표도를 이용하여 차별적 결과를 도출하려는 최초의 시도이다.



< 그림 7 > Saliency-Guided Image Super-Resolution 기술, ICIP 2023

7. Segmentation-Guided SAR-to-EO Translation 연구

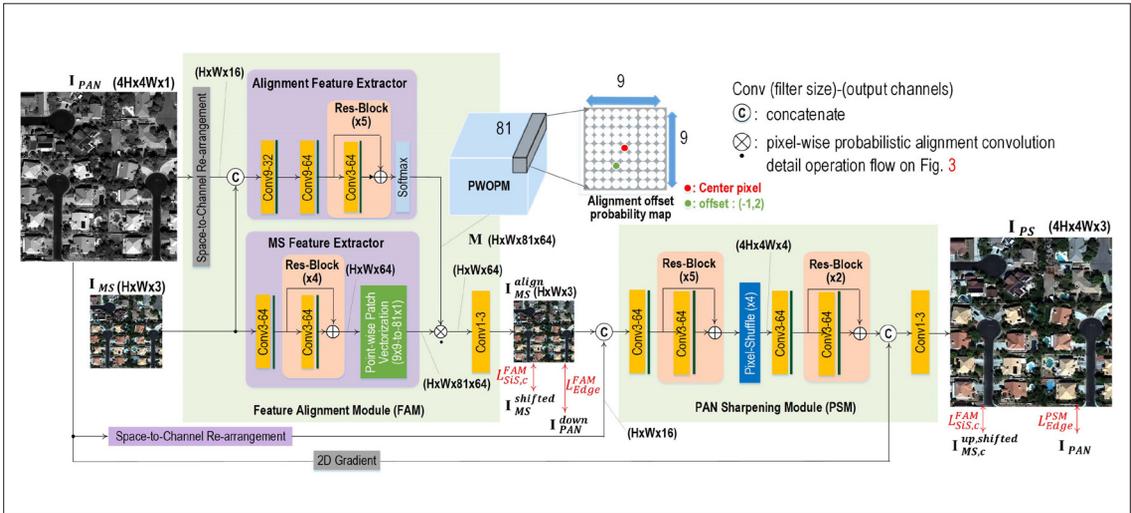
최근 합성 개구 레이더(Synthetic Aperture Radar, SAR) 이미지를 분석하는 연구가 감시 분야, 군사 작전 및 환경 모니터링에 필수적으로 사용되고 있다. SAR 이미지에는 Speckle noise가 존재하기 때문에 해석에 어려움이 발생한다. SAR 이미지의 선명도를 높이기 위해, 본 연구실에서는 광학(Electro-Optical, EO) 영상의 세그멘테이션 레이블 정보를 활용하여 안정적으로 이미지를 변환하는 새로운 SAR-to-EO 이미지 변환(SET) 네트워크인 SGCL-SET 연구를 수행하였다. 본 연구에서는 사전 학습된 세그멘테이션 네트워크로 방대한 세그멘테이션 레이블을 생성하여 SET 학습에 활용한다. SGCL-SET은 세그멘테이션 및 레이블 정보를 사용하기 때문에 변환이 어려운 영역에 대해서도 효과적으로 훈련될 수 있다. KOMPSAT 위성 영상 데이터셋에서 수행한 실험을 통해, SGCL-SET은 9개의 이미지 품질 평가 지표에서 이전의 모든 방법들을 큰 차이로 능가했음을 보인다.



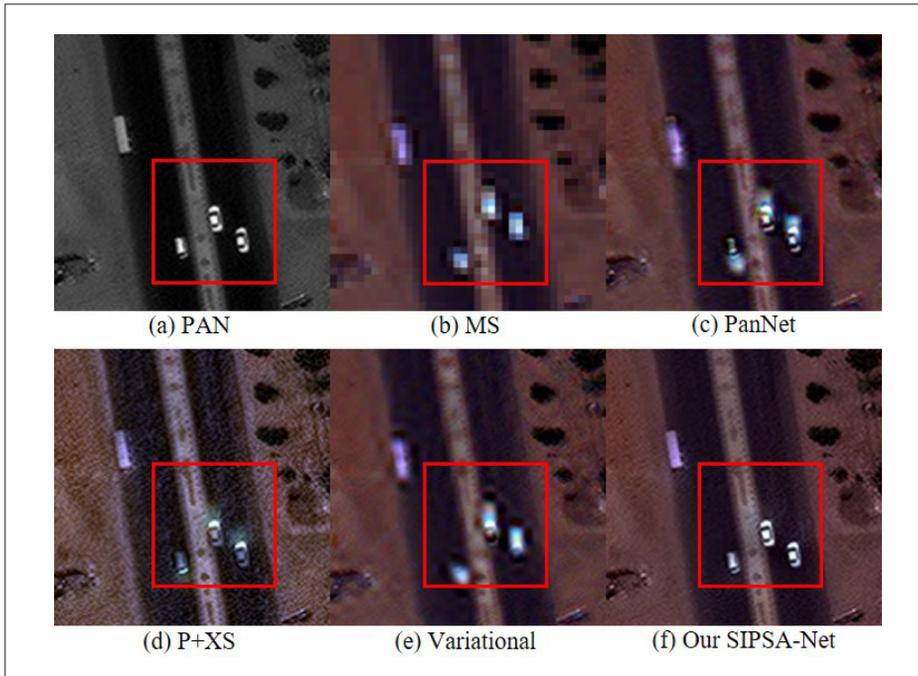
< 그림 9> 컨텍스트 인지적 생성 모델 기반 SAR-to-EO 변환, IEEE TGRS, 2023

9. 위성 영상 Pan-Sharpening 연구

팬샤프닝(Pan-Sharpening)은 고해상도(HR) 팬크로매틱(PAN) 이미지와 해당 저해상도(LR) 멀티스펙트럴 (MS) 이미지를 결합하여 고해상도 멀티스펙트럴(HR-MS) 및 팬샤프닝 이미지를 생성하는 과정이다. 그러나 서로 다른 센서 위치, 특성 및 획득 시간 차이로 인해 PAN 및 MS 이미지 쌍은 비정렬(misalignment) 문제를 겪는다. 이러한 비정렬된 PAN-MS 이미지 쌍으로 훈련된 기존의 딥러닝 기반 방법들은 팬샤프닝 결과물에 이중 경계선(double-edge) 및 흐림(blur)과 같은 다양한 아티팩트를 일으킨다. 본 연구실에서는 팬샤프닝을 위해 이동 물체 영역의 큰 불일치(misalignment)를 학습 과정에서 극복하고자 이동 물체 정렬 매칭을 통한 쉬프트 불변(shift-invariant) Pan-sharpening 네트워크(SIPSA-Net)를 제안하였다. SIPSA-Net은 서로 다른 PAN 및 MS 도메인 간에도 하나의 특징맵을 다른 특징맵에 맞출 수 있는 feature align module을 포함한다. 팬샤프닝된 이미지의 더 나은 정렬을 위해, 잘 정렬된 MS 이미지로 스펙트럼 손실을 최적화하는 것과 동일한 효과를 가지는 shift-invariant spectral loss를 제안하였다. 실험 결과, SIPSA-Net은 시각적 품질과 정렬 측면에서 최신의 방법들과 비교하였을 때 현저한 개선을 이루며 팬샤프닝된 이미지를 생성하였다.



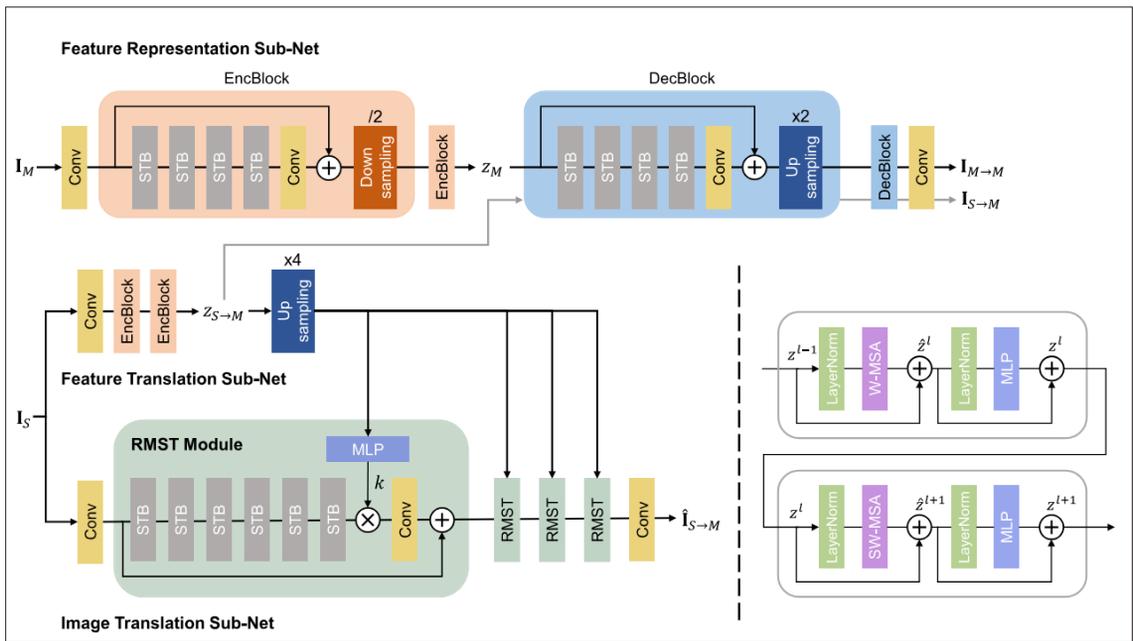
<그림 10> 위성 영상 Pan-Sharpening 연구, CVPR 2021



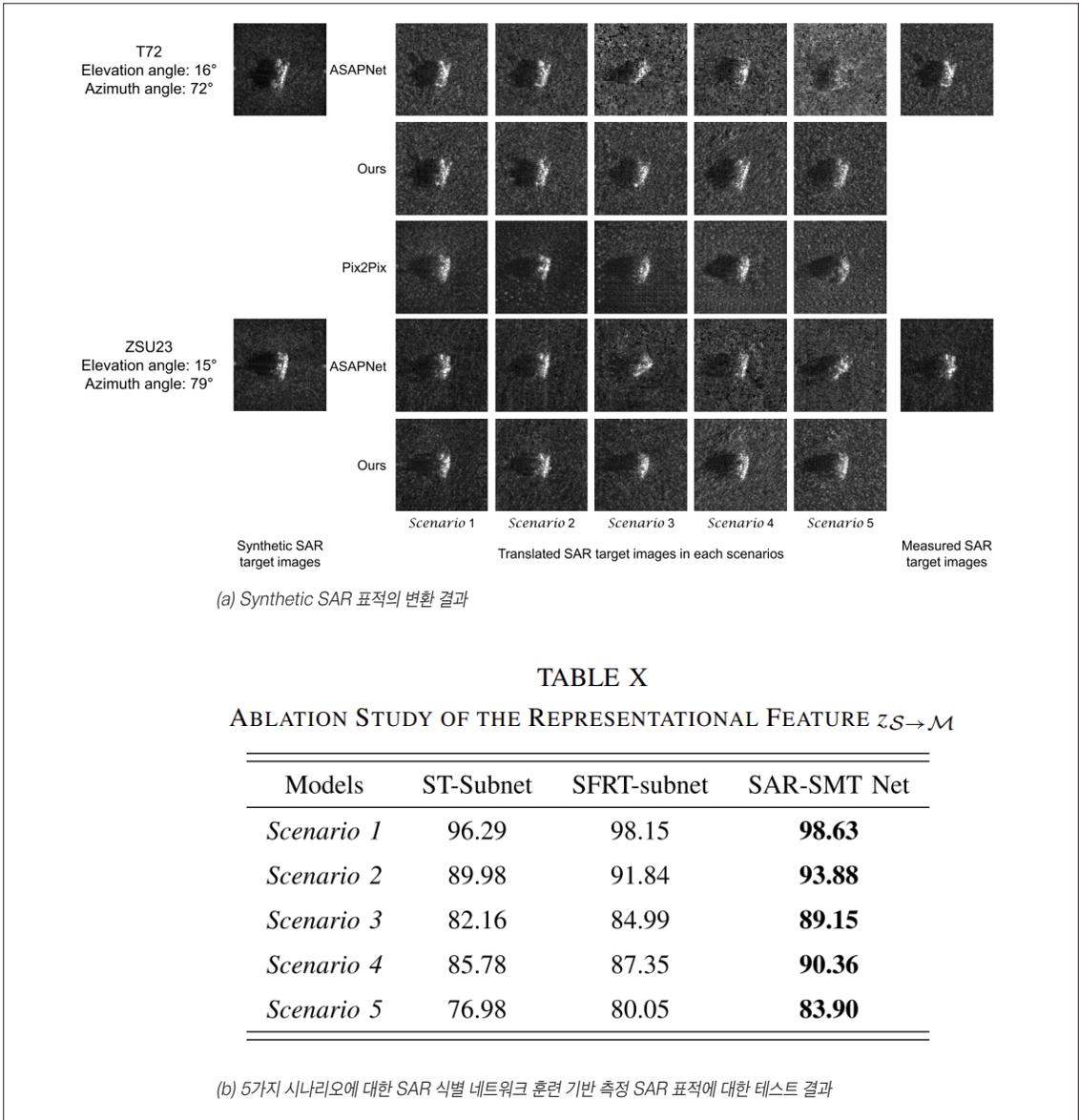
<그림 10-1> 위성 영상 Pan-Sharpening 연구 결과

10. 생성 모델 기반 합성 SAR 영상 변환 연구

최근 SAR 영상을 분석하여 활용하는 연구들이 활발히 진행되고 있는데, 고성능의 인공지능 네트워크를 학습시키기 위해 다양한 SAR 영상을 대량으로 수집하는 것은 비용 문제로 인해 쉽지 않다. 따라서 이러한 데이터 부족 문제를 해결하기 위한 다양한 방법들이 시도되고 있는데, 그 중 대표적인 방법이 CAD 모델링을 통해 SAR 표적 영상을 생성하는 것이다. 이를 통해 다양한 종류의 SAR 표적 영상을 쉽고 빠르게 생성할 수 있다. 그러나 CAD 모델링의 한계로 인해 합성된 SAR 표적 영상이 실제로 측정된 SAR 표적 영상과 많은 차이가 있어 이를 직접 딥러닝 모델 학습에 사용하기에는 많은 문제점이 존재한다. 따라서, 본 연구실에서는 CAD로 모델링된 합성 SAR 표적 영상을 실제 SAR 표적 영상과 유사하게 변환하는 SAR 표적 영상 변환 네트워크인 SAR-SMT Net을 연구하였다. SAR-SMT Net은 합성 SAR 표적 영상 입력에 대해 실제 SAR 표적 영상의 잠재적 표현 특성을 추정하여, 입력된 합성 SAR 표적 영상의 컨텍스트와 산란 특성을 실제 SAR 표적 영상 값에 맞게 충실히 조정한다. 또한 본 연구에서는 SAR 표적 영상 변환 성능 결과를 검증할 수 있는 5가지 도전적인 실험 시나리오를 제안하여 SAR-SMT Net이 기존 방법들보다 합성 SAR 영상을 충실하게 변환할 수 있음을 보였다.



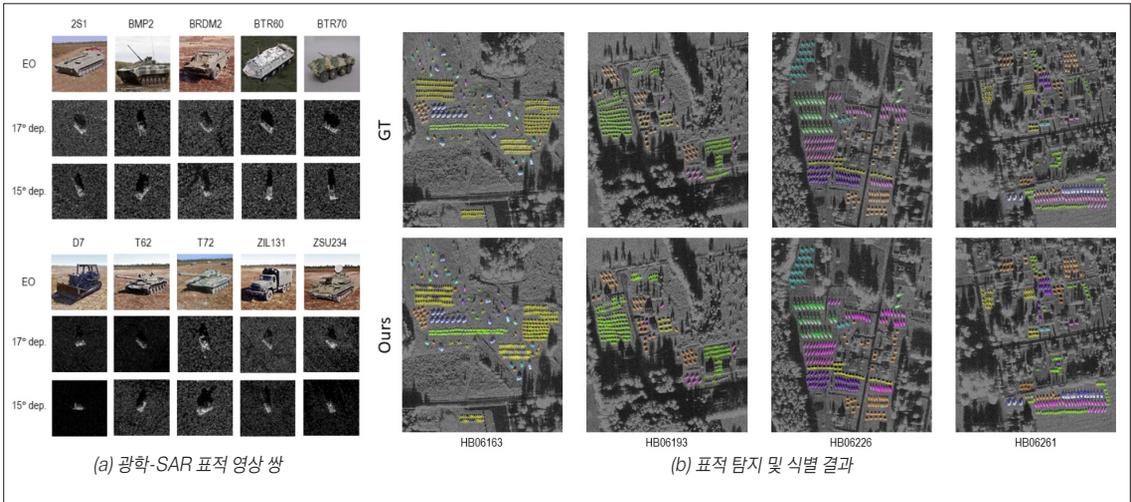
< 그림 11 > 생성 모델 기반 합성 SAR 영상 변환 연구, IEEE TGRS, 2023



<그림 11-1> 합성 SAR 표적 영상 변환 결과 및 식별 성능 결과

11. 열악한 탐지 및 식별 환경에서의 SAR 표적 탐지 및 식별 연구

SAR 영상은 날씨 조건 및 시간에 영향을 거의 받지 않고 원하는 지역(또는 영역)에 대한 영상을 획득하는 것이 가능하여 감시 및 정찰을 위한 용도로 많이 활용되고 있다. SAR 영상에 대해 표적 탐지를 수행하면 EO(Electro-Optical) 영상 장비를 통해 획득한 영상에 대해 수행하는 것보다 높은 일관성을 유지하면서 탐지 결과를 얻을 수 있다. 이러한 특성으로 SAR 영상을 통한 표적 탐지는 국방 분야에서 높은 중요성을 가진다. 본 연구실에서는 열악한 SAR 표



<그림 12> 열악한 조건에서의 SAR 표적 탐지 및 식별 결과, 한국군사과학기술학회지 2022.10

적 탐지 및 식별 환경에서 높은 성능의 표적 탐지 및 식별을 수행하기 위해 딥러닝 기반 표적 탐지 및 식별 알고리즘인 YOLOv4를 활용한 네트워크를 제안하였다. 특히, 특징 지도의 각 요소가 표적 탐지 성능에 미치는 중요도를 학습하는 어텐션 기능을 YOLOv4 구조의 네트워크에 추가하여 확장하였다. 본 연구를 진행하기 위해 열악한 SAR 표적 탐지 및 식별 환경 데이터셋을 구축하였으며, 제안 데이터셋을 활용하여 기존 YOLOv4 네트워크 대비 향상된 표적 탐지 및 식별 성능을 보임을 확인하였다. 본 연구 논문은 2023년 한국군사과학기술학회로부터 박창규 학술상을 수상하였다.

II. 최근 VICLAB 수상 현황

1. 수상 실적

년도	수상명
2024	한국방송-미디어공학회 하계학술대회 논문 대상 (김태환 박사과정생), 우수논문상 수상 (권병준 박사과정생)
2023	한국방송-미디어공학회 하계학술대회 논문 대상 (Juan Gonzalez 박사과정생), 우수논문상 수상 (박종민 박사과정생) 한국군사과학기술학회 종합학술대회 박창규 학술상 수상 (박종민 박사과정생)
2021	KAIST 연구대상 (김문철 교수) Outstanding Ph. D. Research Award in Signal Division of School of EE for Year 2021 (Juan Gonzalez 박사과정생) Nomination Award of 2021 Microsoft Research Asia Fellowship (Juan Gonzalez 박사과정생) Google Ph.D. Fellow 선정 (김수예 박사과정생)
2020	ECCV AIM (Video Temporal Super-Resolution) Challenge 2위 수상 (심현준 박사과정생)
2019	제54회 발명의 날 대통령 표창 수상 (김문철 교수) KAIST 연구혁신상 수상 (김문철 교수) HEVC 표준 기술 공동개발 공로 - KBS 사장 감사패 수상 CVPRW "Deblurring Challenge" 3위 입상 (심현준 박사과정생)

2. 주요 논문 실적

년도	발행처	논문 수
2024	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	3
	IEEE Access	2
	IEEE Transactions on Geoscience and Remote Sensing (TGRS)	1
	IEEE Transactions on Image Processing (TIP)	1
	IEEE Transactions on Multimedia	1
	European Conference on Computer Vision (ECCV)	1
2023	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	1
	IEEE/CVF International Conference on Computer Vision (ICCV)	1
	IEEE Transactions on Geoscience and Remote Sensing (TGRS)	2
2022	Neural Information Processing Systems (NeurIPS)	1
	The European Conference on Computer Vision (ECCV)	1
	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	1
	IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)	1
	IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	1
2021	IEEE/CVF International Conference on Computer Vision (ICCV)	1
	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	3
	IEEE Access	3
	IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	3
2020	Neural Information Processing Systems (NeurIPS)	1
	The European Conference on Computer Vision (ECCV)	1
	International Conference on Learning Representations (ICLR)	1
	AAAI Conference on Artificial Intelligence (AAAI)	2
	IEEE Access	4
	IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	1



<그림 13> VICLAB 단체 사진

저 자 소 개



김문철

- 2001년 2월 ~ 2009년 2월 : 한국정보통신대학교 공학부 조교수/부교수
- 2009년 3월 ~ 2015년 8월 : 한국과학기술원 전기및전자공학부 부교수
- 2015년 9월 ~ 현재 : 한국과학기술원 전기및전자공학부 정교수
- 2021년 3월 ~ 2024년 2월 : 한국과학기술원 ICT 석좌교수

- 2001년 : MPEG 국제 표준화 회의 한국대표단 단장 (Head of Delegates)
- 2002년 7월 ~ 2018년 11월 : MPEG뉴미디어포럼 사무총장
- 2018년 4월 ~ 현재 : ETRI Journal Section Editor
- 2020년 11월 ~ 2023년 10월 : 대전광역시 컨벤션 대사