2025 동계 세미나

RGB to IR transformation via text-guided feature representation

Pusan National University
Dept. of Electronic Engineering
Computer Vision & Signal Processing Lab

Presented By
정다운

CVSP Lab
Computer Vision & Signal Processing

Contents

- Paper review
  - Paper 1
    - Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[ICRA 2023]
  - Paper 2
    - DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[NeurIPS 2024]

Paper review 1

Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels

ICRA 2023

CVSP Lab
Computer Vision & Signal Processing

Paper 1

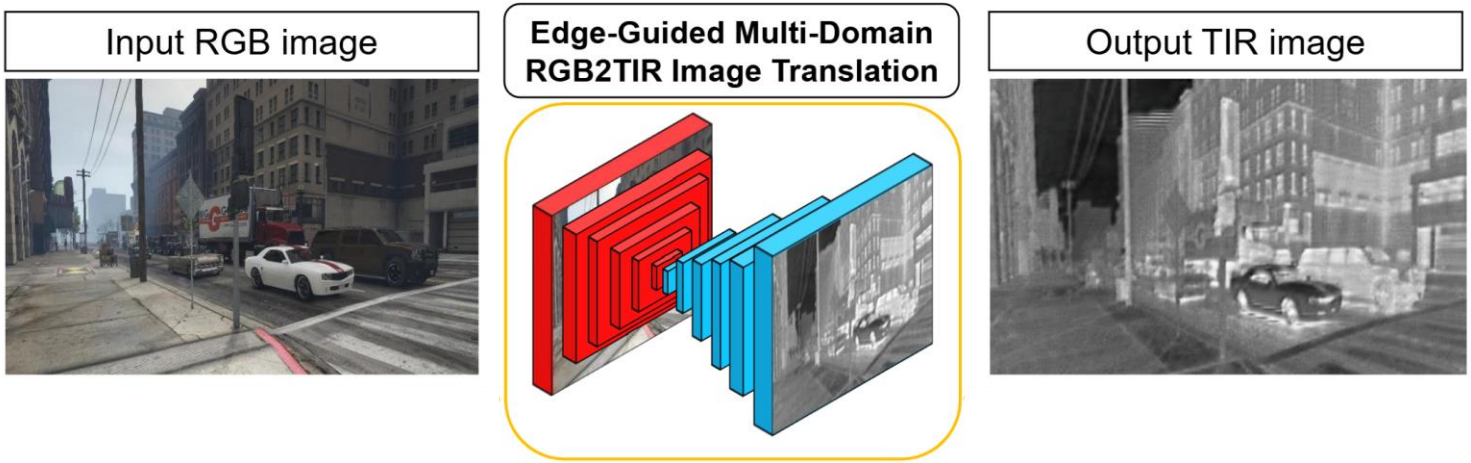
- Background
  - Thermal Infrared(TIR) Image
    - 정의
      - 가시광선 바깥의 적외선 스펙트럼 중 Thermal Infrared 대역에서 찍힌 이미지
        - ✓장파 적외선(LWIR, 8~14μm) 대역
    - 특징
      - 물체의 표면에서 방출하는 열 복사를 감지해서 생성
      - 열에 기반한 정보를 시각화하며 단일 채널(gray scale) 형태로 표현됨
      - 온도 차이에 기반한 밝기로 표현됨(0~255)
        - ✓빛이 없어 우리에게 보이지 않아도 감지가 가능함



<TIR image 예시, FLIR-ADAS dataset>

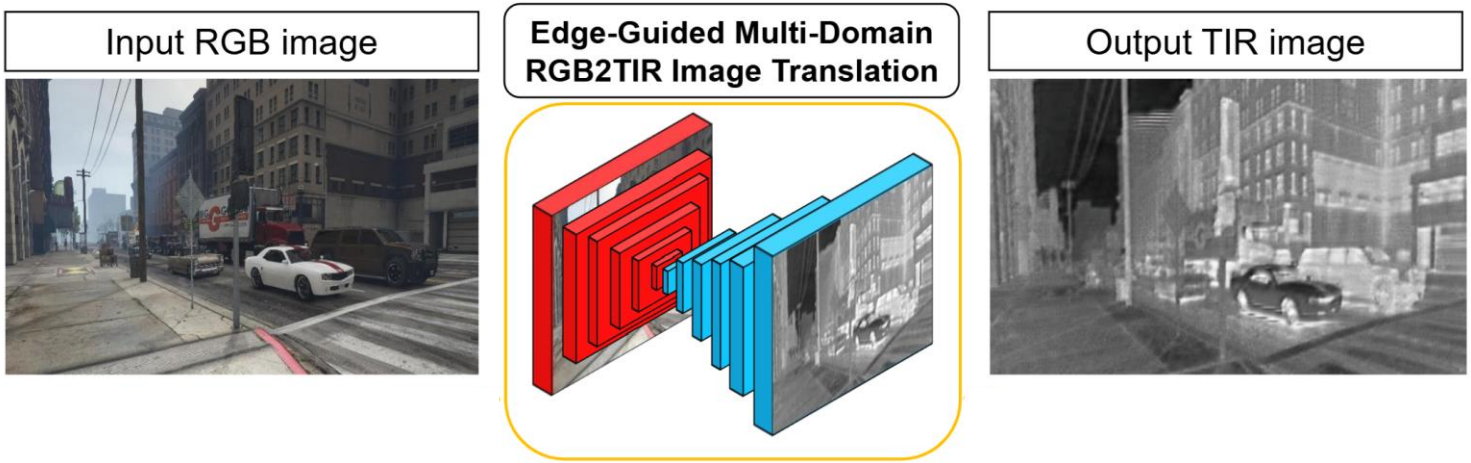
Paper 1

- Background
  - RGB to TIR image translation
    - 일반적인 가시광선(RGB) 이미지를 적외선(Thermal Infrared, TIR) 이미지로 변환
      - TIR 환경은 야간 등 악조건에서도 시각 정보를 활용할 수 있음
      - TIR 이미지 데이터셋 부족
        - ✓라벨링된 Optical flow, object detection TIR이미지 데이터셋 부족
      - RGB 이미지로 TIR 이미지 데이터를 생성하여 데이터 부족 문제 해결
        - ✓TIR 환경에서도 optical flow, object detection 등 downstream task들의 성능 향상



Paper 1

- Background
  - RGB to TIR image translation
    - 기존 방식
      - RGB환경(낮, 밤 등)에 따라 real TIR 이미지는 특성이 다름
        - ✓기존 RGB to TIR 모델은 변환 스타일이 하나의 스타일
        - ✓다양한 환경을 고려하지 않아 real TIR이미지와 거리가 멀었음



Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Introduction
    - Edge-guided와 style-control이 가능한 멀티도메인 기반 RGB to TIR 변환
      - Edge-guided를 이용한 RGB to IR
        - ✓RGB 이미지의 contents detail을 유지
      - Style-control 가능한 멀티도메인 기반 RGB to IR
        - ✓기존 방식과 달리 가장 적합한 스타일 벡터를 활용해서 현실적인 TIR이미지 생성
        - ✓라벨링이 어려운 optical flow estimation TIR 이미지 데이터셋을 생성

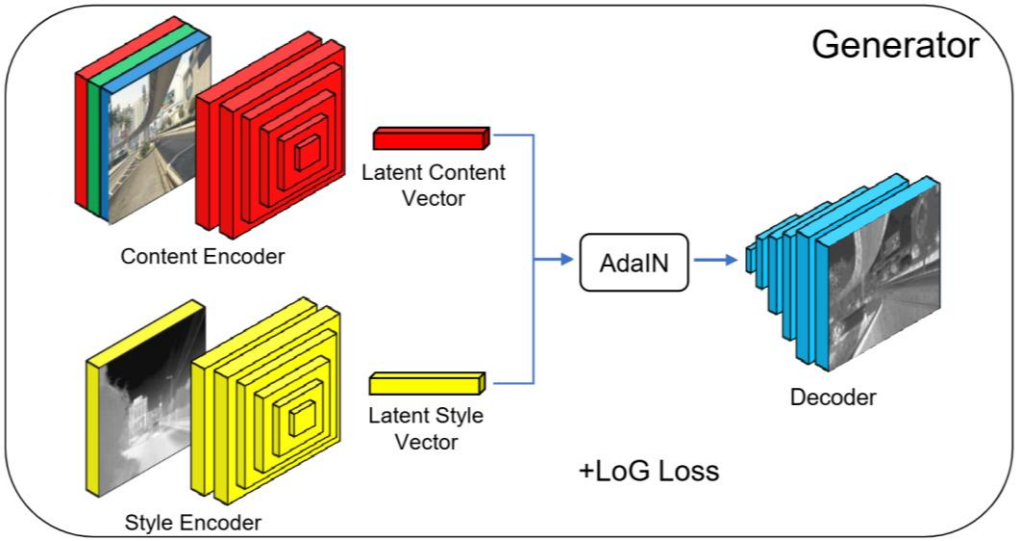


<RGB to IR image translation 결과 예시>

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - 네트워크 구조
      - GAN 기반의 멀티 도메인 unpaired image to image translation architecture
        - ✓Generator: 두 개의 encoders(content, style), 하나의 Decoder로 구성
          - RGB image를 Content Encoder, TIR image을 Style encoder에 입력
          - RGB image의 contents를 반영하는 TIR image style의 이미지 생성

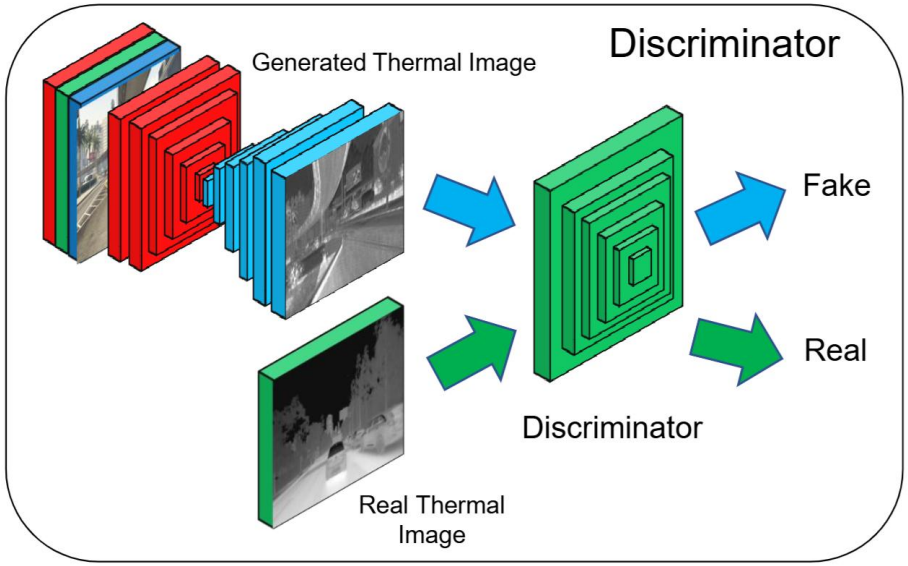


CVSP Lab
Computer Vision & Signal Processing

Paper 1

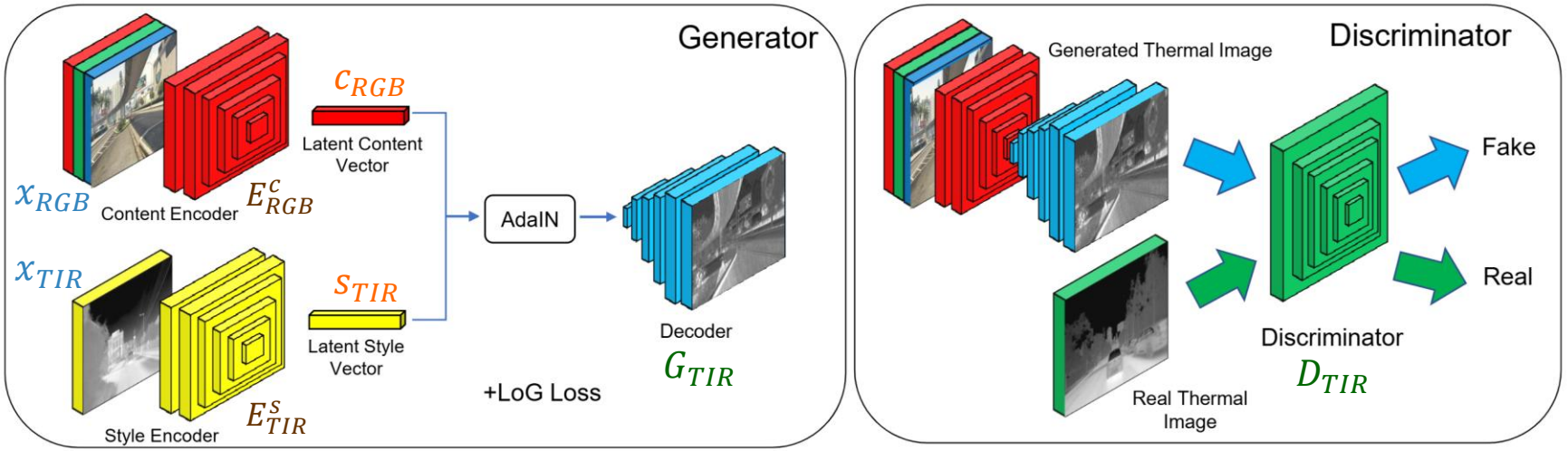
1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - 네트워크 구조
      - GAN 기반의 멀티 도메인 unpaired image to image translation architecture
        - ✓Discriminator: real TIR이미지와 생성된 TIR 이미지를 real인지 fake인지 판별
        - ✓Unpaired dataset으로 RGB to IR translation이 가능함



Paper 1

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Notations in paper
      - x_{RGB}, x_{TIR} : Input RGB and TIR images
      - E^c_{RGB}, E^s_{RGB} *and* E^c_{TIR}, E^s_{TIR}: Content and style encoder for RGB and TIR images
      - c_{RGB}, s_{RGB} *and* c_{TIR}, s_{TIR}: Latent content and style vector for RGB and TIR images
      - $G_{TIR}, G_{RGB}, D_{TIR}$: TIR decoder, RGB decoder, TIR discriminator



CVSP Lab
Computer Vision & Signal Processing

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

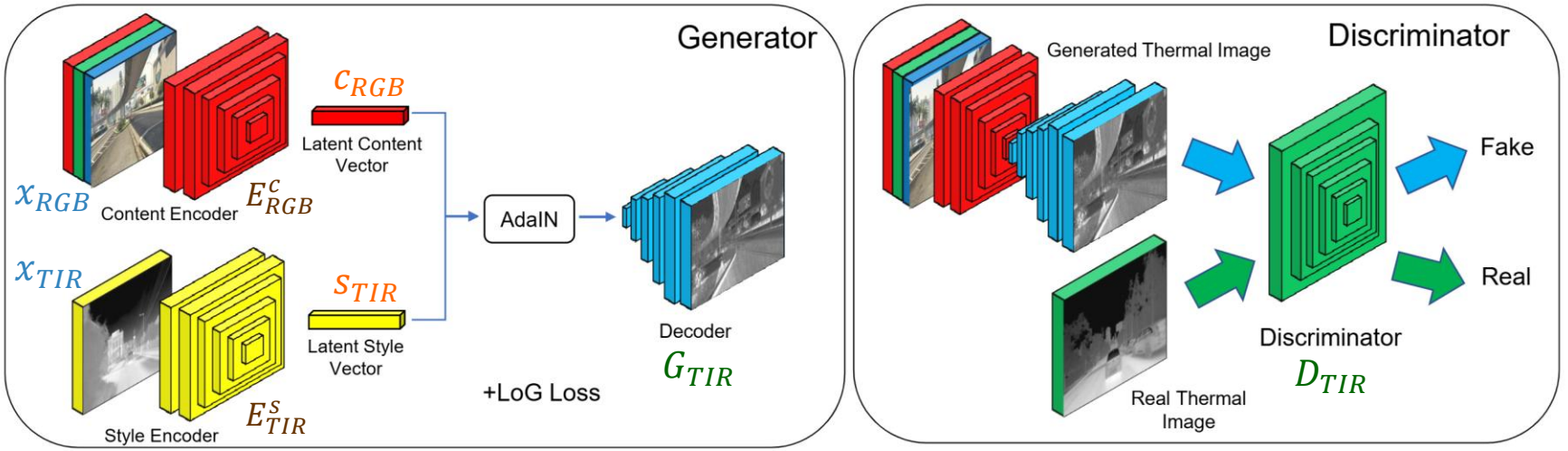
- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Loss
      - Adversarial loss for GAN training
      - Cyclic loss
      - Enforce edge-guidance multi-domain translation loss
        - ✓Image reconstruction Loss
        - ✓Content and Style Reconstruction Loss
        - ✓Laplace of Gaussian (LoG) Loss
      - Overall training objective
        - ✓$\mathcal{L}_{\mathcal{G}} = \mathcal{L}_{\mathcal{GAN}} + \lambda_{x_{recon}}\mathcal{L}_{recon}^{x\ TIR} + \lambda_{c_{RGB}}\mathcal{L}_{recon}^{c\ RGB} + \lambda_{S_{TIR}}\mathcal{L}_{recon}^{S\ TIR} + \lambda_{Lap}\mathcal{L}_{Lap} + \lambda_{cyc}\mathcal{L}_{cyc}$

Paper 1

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Loss
      - Image reconstruction Loss

✓Generator가 원본 RGB이미지를 잘 복원하도록 하는 loss

- $L_{recon}^{x\,RGB} = E[\|G_{RGB}(E_{RGB}^{c}(x_{RGB}), E_{RGB}^{s}(x_{RGB})) - x_{RGB}\|_1]$



CVSP Lab
Computer Vision & Signal Processing

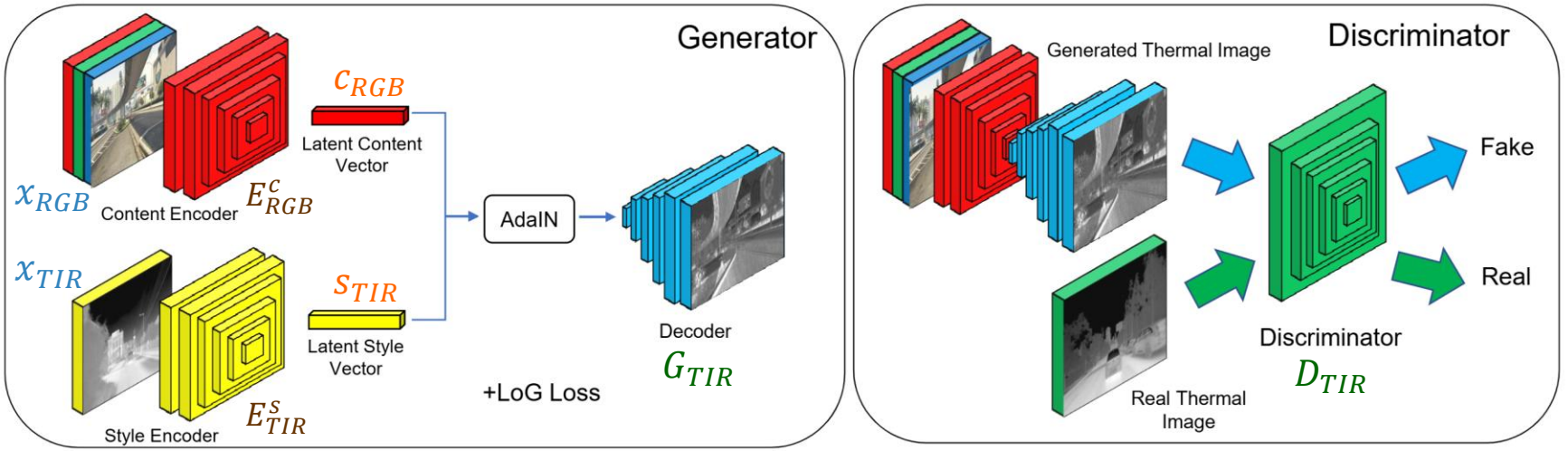
Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Loss
      - Content and Style Reconstruction Loss

✓RGB 이미지의 contents와 IR이미지의 스타일을 보존하기 위한 loss

- $L_{recon}^{c\ RGB} = E[\|E_{TIR}^{c}(G_{TIR}(c_{RGB}, s_{TIR})) - c_{RGB}\|_1]$
- $L_{recon}^{s\ TIR} = E[\|E_{TIR}^{s}(G_{TIR}(c_{RGB}, s_{TIR})) - s_{TIR}\|_1]$



Paper 1

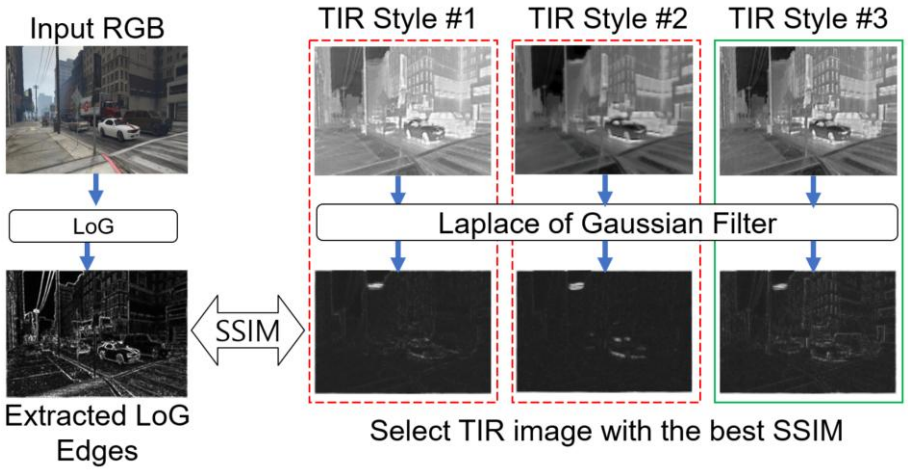
- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Loss
      - Laplace of Gaussian (LoG) Loss
        - ✓이미지의 edge를 강조하는 Laplace filter와 가우시안 필터를 결합한 필터 loss
          - $L_{Lap} = E[\|L(x_{TIR}) - L(x_{TIR,recon})\|_1]$
            - TIR이미지의 edge를 잘 보존하도록 학습
          - $L_{Lap} = E[\|L(x_{RGB}) - L(x_{RGB,recon})\|_1]$
            - RGB 이미지의 edge를 잘 보존하도록 학습

CVSP Lab
Computer Vision & Signal Processing

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Method
    - Translation style selection
      - 다양한 style코드 중 가장 적합한 style을 선택하여 현실적인 TIR의 특징을 묘사
      - 방법
        - ✓Laplacian of Gaussian(LoG)필터를 이용하여 RGB이미지, style 이미지의 edge 추출
        - ✓RGB 이미지의 LoG edge와 SSIM*이 가장 높은 스타일 코드 선택 후 최적화



*SSIM (Structural Similarity Index Measure): 두 이미지 간의 **구조적 유사성**을 정량적으로 평가하기 위한 척도.
픽셀 간 차이만을 측정하는 것이 아니라 사람의 인지방식과 유사하게 이미지를 비교함.

CVSP Lab
Computer Vision & Signal Processing

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Experiments
    - A. Synthetic RGB to TIR image translation
      - 생성된 TIR 이미지가 원본 RGB 이미지의 edge를 잘 유지하는지 실험
      - 지표: APCE(Average Precision of Canny Edges)
        - ✓Canny edge를 기준으로 생성된 IR이미지와 원본 RGB 이미지의 edge 간 유사성 평가
    - B. Deep supervised TIR optical flow estimation
      - Labeling된 RGB optical flow 데이터셋을 TIR 이미지로 변환 후 모델을 학습
      - IR환경에서 optical flow 예측 평가
      - 지표: EPE (Endpoint Error)
        - ✓실제 optical flow와 예측된 optical flow 간 차이 측정
    - C. Extension to object detection
      - Labeling된 object detection 데이터셋을 TIR 이미지로 변환 후 object detection 모델 학습
      - TIR환경에서 object detection 성능 평가

CVSP Lab
Computer Vision & Signal Processing

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Experiments
    - A. Synthetic RGB to TIR image translation
      - 생성된 TIR 이미지가 원본 RGB 이미지의 edge를 잘 유지하는지 실험
      - 지표: APCE(Average Precision of Canny Edges)
        - ✓Canny edge를 기준으로 생성된 IR이미지와 원본 RGB 이미지의 edge 간 유사성 평가

Models	Proposed	CycleGAN	MUNIT	UNIT
Synthetic Images				
GTA-Day	**0.0793**	0.0495	0.0175	0.0322
GTA-Night	**0.1264**	0.0351	0.0262	0.0304
Synthia	**0.0967**	0.0753	0.0289	0.0404
Virtual Kitti	0.1221	**0.1516**	0.0165	0.0138
Sintel-1	**0.1301**	0.0606	0.0226	0.0067
Sintel-2	**0.1246**	0.0398	0.0118	0.0584
Average	**0.11**	0.07	0.02	0.03
Real Images				
Valley-Morning	**0.0936**	0.045	0.0232	0.0207
SNU Afternoon	**0.1502**	0.0354	0.0248	0.0514
KAIST Day	**0.1683**	0.0658	0.032	0.0389
KAIST Night	**0.1432**	0.0456	0.0399	0.046
FLIR Day-1	**0.0544**	0.0395	0.0204	0.0281
FLIR Day-2	**0.0943**	0.0486	0.0261	0.0451
Average	**0.12**	0.05	0.03	0.04

Paper 1

1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Experiments
    - B. Deep supervised TIR optical flow estimation
      - Labeling된 RGB optical flow 데이터셋을 TIR 이미지로 변환 후 모델을 학습
      - TIR환경에서 optical flow 예측 평가
      - 지표: EPE (Endpoint Error)
        - ✓실제 optical flow와 예측된 optical flow 간 차이를 측정
        - ✓두 flow의 유클리디안 거리를 픽셀별로 구해 평균

Model	EPE (RGB)	Training EPE (Thermal)	EPE (Thermal)
RAFT	43.762	3.069	16.18
GMFlow	44.106	8.536	24.568
GMA-RAFT	41.740	**2.917**	**15.733**

CVSP Lab
Computer Vision & Signal Processing

Paper 1

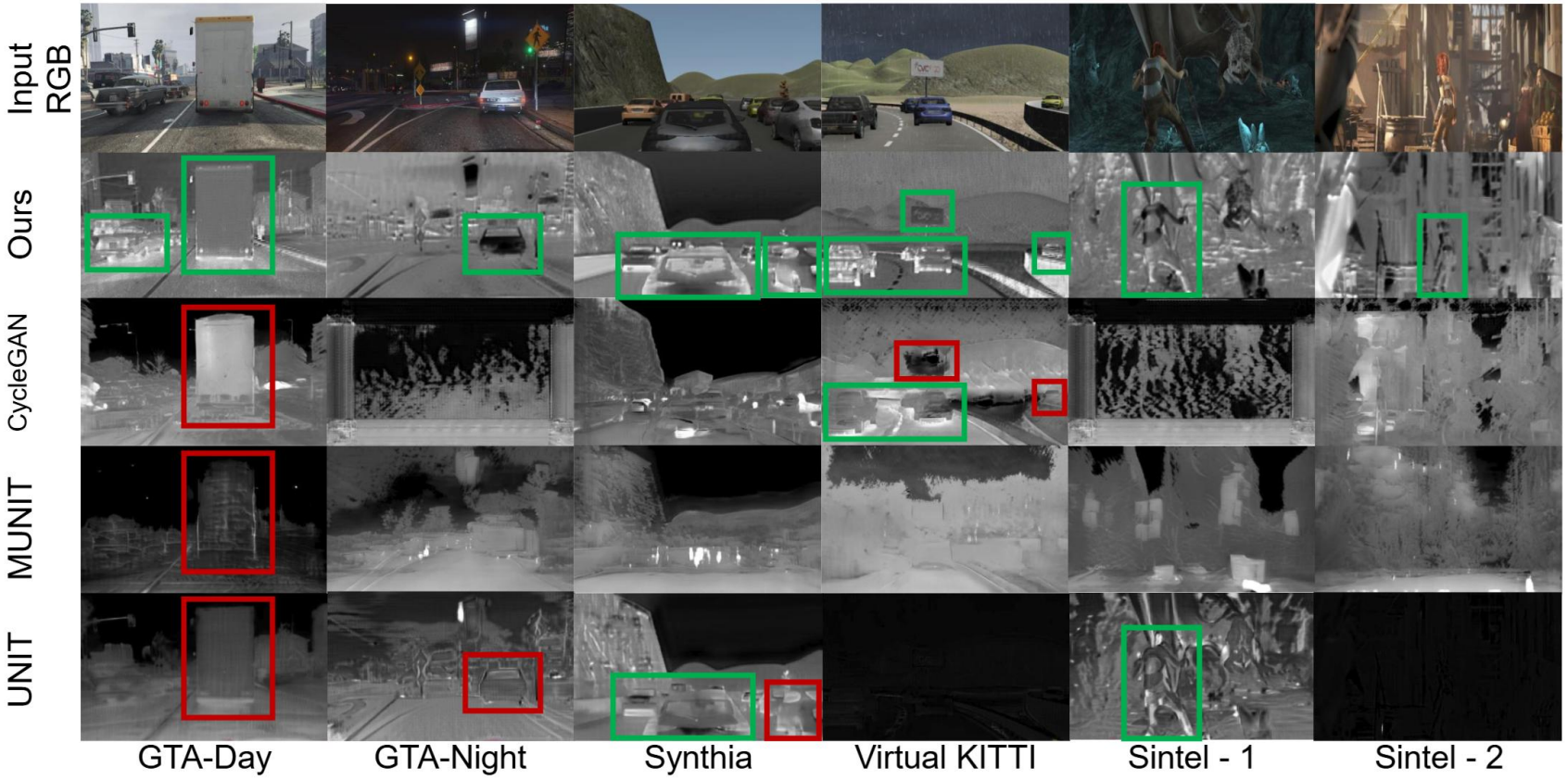
1) Lee, Dong–Guw, et al. "Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.

- Edge-guided Multi-domain RGB-to-TIR image Translation for Training Vision Tasks with Challenging Labels[1)]
  - Experiments
    - C. Extension to object detection
      - Labeling된 RGB dataset을 TIR로 변환하여 Detection model 학습
        - ✓다른 변환방법들에 비해 가장 높은 mAP인 0.239 달성

Methods	mAP	mAP_50	mAP_75	mAP_S	mAP_M	mAP_L
Proposed	**0.239**	**0.408**	**0.237**	0.061	**0.441**	**0.692**
CycleGAN	0.007	0.01	0.01	0	0.005	0.009
MUNIT	0.069	0.137	0.06	0.007	0.117	0.413
UNIT	0.225	0.384	0.222	**0.062**	0.414	0.671

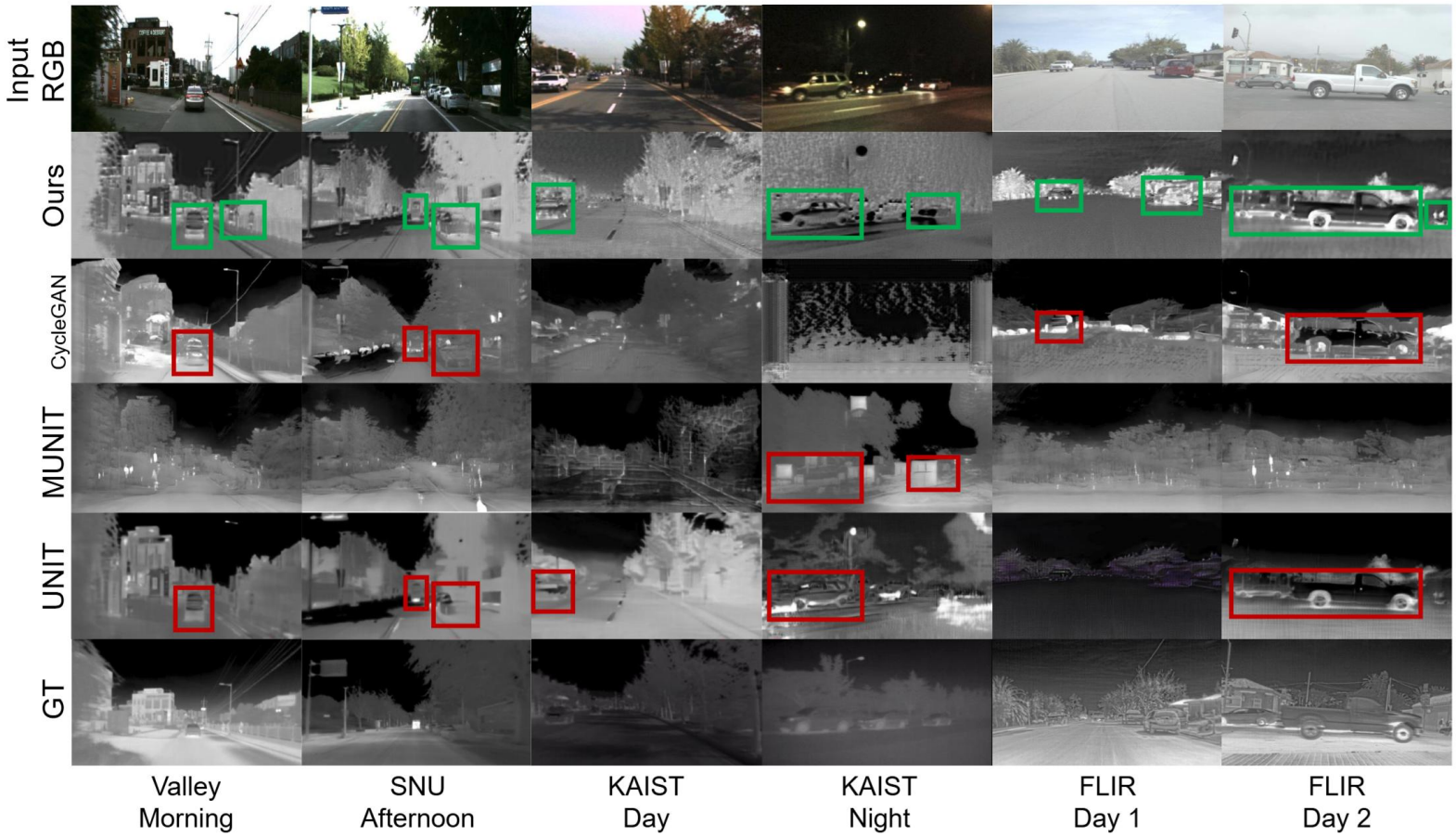
CVSP Lab
Computer Vision & Signal Processing

- Qualitative results
  - Synthetic dataset – object detection



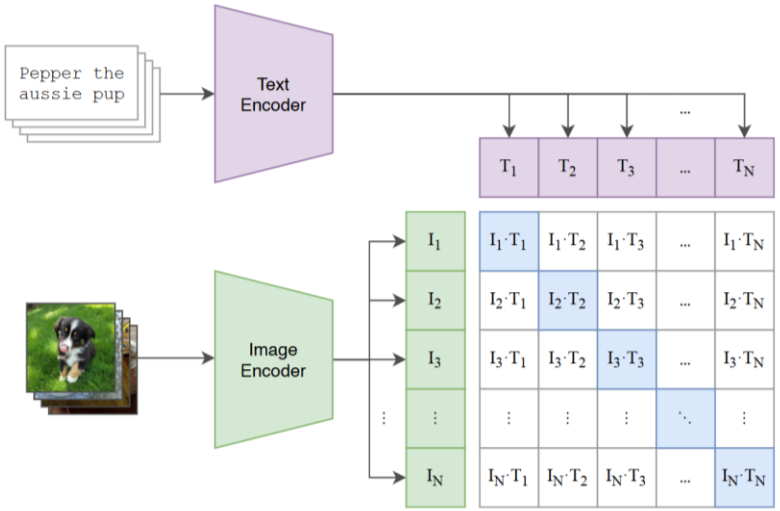
CVSP Lab
Computer Vision & Signal Processing

- Qualitative results
  - Real RGB dataset – object detection



CVSP Lab
Computer Vision & Signal Processing

- Background
  - Contrastive Language-Image Pretraining(CLIP)[2)]
    - 이미지와 텍스트 pair간의 관계를 학습하는 멀티모달 모델
      - 학습 방법
        - ✓이미지와 텍스트 페어를 각각 임베딩 벡터로 변환
        - ✓정답 pair는 벡터 간 거리를 가깝게 오답 pair는 거리가 멀어지도록 학습
      - 다양한 downstream task: 텍스트 입력으로 이미지 검색, 이미지 입력으로 텍스트 생성 등

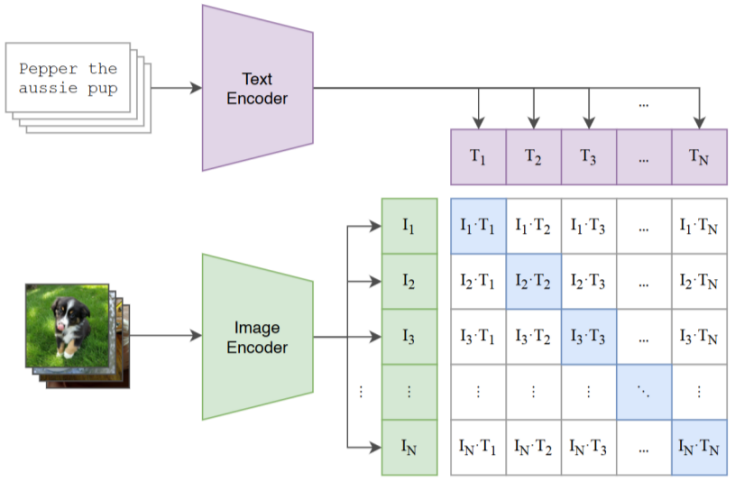


<CLIP 모델의 pre-training>

- Background
  - Contrastive Language-Image Pretraining(CLIP)[2)]
    - Loss
      - Contrastive learning을 기반으로 한 InfoNCE loss

✓ $\mathcal{L} = \mathcal{L}_{image-text} + \mathcal{L}_{text-image}$

- $\mathcal{L}_{image-text} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{(sim(z_i,z_t)/\tau)}{\sum_{j=1}^{N}\exp(sim(z_i,z_j)/\tau)}$
- $\mathcal{L}_{text-image} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{(sim(z_t,z_i)/\tau)}{\sum_{j=1}^{N}\exp(sim(z_t,z_j)/\tau)}$



<CLIP 모델의 pre-training>

CVSP Lab
Computer Vision & Signal Processing

Paper review 2

DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor

NeurIPS 2024

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

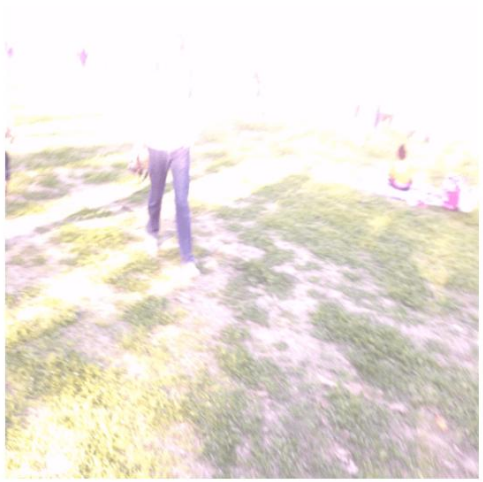
- Background
  - Image Degradation이란
    - 이미지 품질이 망가지거나 왜곡되는 과정
      - 유형: Noise, Blur, Exposure, Low Resolution, Compression Artifacts



Noise



Blur



Exposure

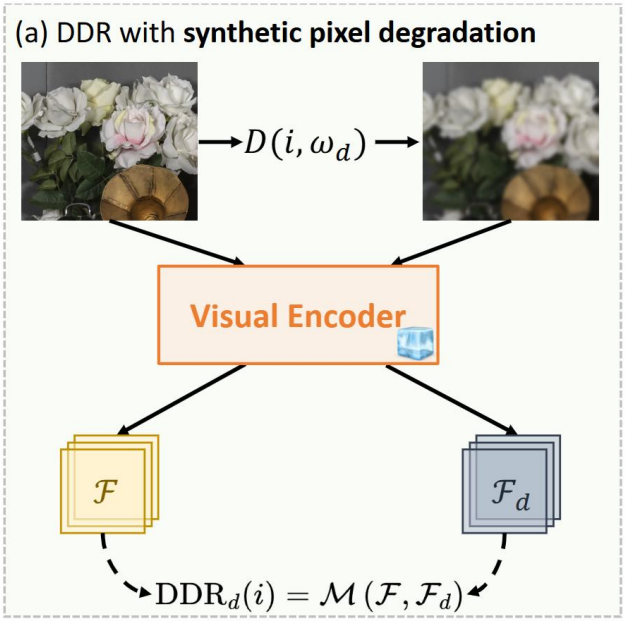
<Degraded Images 예시>

CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Introduction
    - Deep Degradation Response(DDR)
      - 이미지의 feature가 다양한 degradation 유형에서 변화하는 양을 정량화
        - ✓사람이 직접 라벨링한 IQA 데이터셋 없이도 IQA metric으로 사용 가능
        - ✓DDR을 이미지 Restoration model의 학습 목표로 추가하여 성능 향상

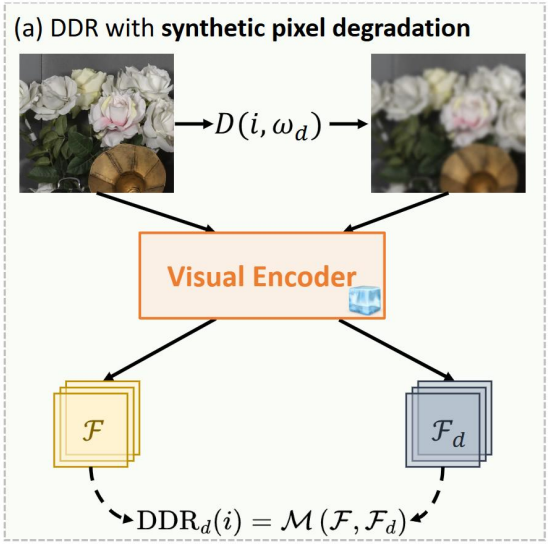


CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

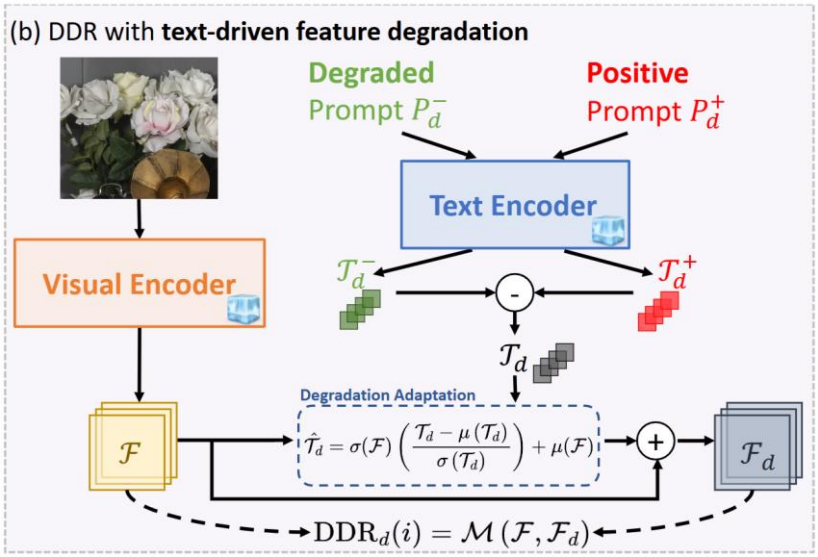
- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Deep Degradation Response(DDR)
      - 특정 유형의 degradation이 적용됐을 때 이미지 feature에서 변화를 측정
        - ✓$DDR_d(i) = M(F, F_d)$
          - d: $degradation$ 유형
          - $M(\cdot,\cdot)$: 차이 측정(L_n 또는 코사인 거리 등)
          - $F = \Phi_v(i)$: 이미지 feature
          - $F_d = degradation$ 된 feature



CVSP Lab
Computer Vision & Signal Processing

Paper 2

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Deep Degradation Response(DDR)
      - CLIP model의 이미지 인코더/텍스트 인코더를 사용해서 text기반으로 feature단에서 degradation을 표현
        - ✓Positive prompt와 degraded prompt를 모델링해서 얻은 feature 간 차이를 이미지 feature에 융합
        - ✓Text prompt로 degradation된 이미지 feature를 얻을 수 있음

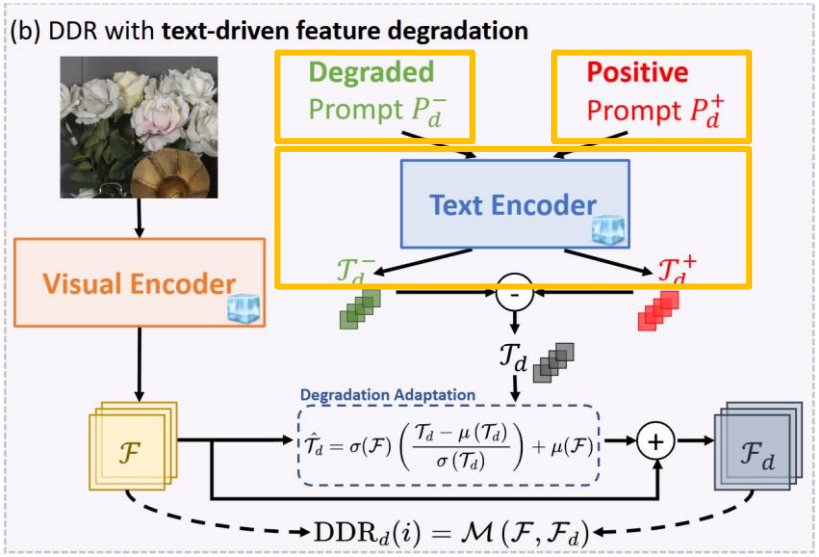


CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Fusing text-driven degradation
      - Feature domain에서 text-driven prompts로 degradation을 합성
        - ✓Text encoder $\Phi(\cdot)$를 통해 P_d^-와 P_d^+의 feature T_d^- 와 T_d^+ 생성
          - prompt P_d^- : degraded 이미지를 표현하는 prompt
            - ex) "A noise degraded photo with low-quality."
          - Prompt P_d^+ : 동일한 이미지를 degradation 없이 표현하는 prompt
            - ex) "A clean photo with high-quality."



CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Fusing text-driven degradation
      - Feature domain에서 text-driven prompts로 degradation을 합성
        - ✓Text encoder $\Phi(\cdot)$를 통해 P_d^-와 P_d^+의 feature T_d^- 와 T_d^+ 생성
          - Degraded prompt P_d^- 와 clean prompt P_d^+

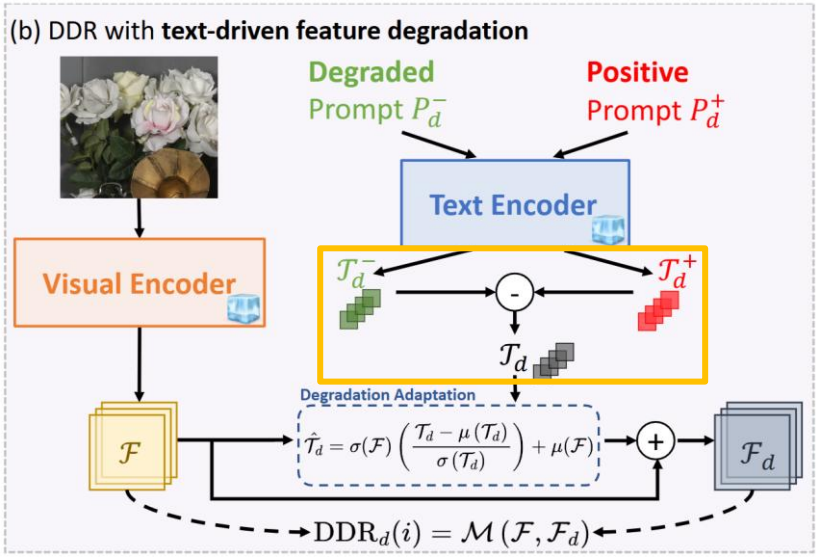
```
self.clean_prompt['color'] = "A real color photo with high-quality."
self.degraded_prompt['color'] = "A unnatural color photo with low-quality."
self.clean_prompt['noise'] = "A clean photo with high-quality."
self.degraded_prompt['noise'] = "A noise degraded photo with low-quality."
self.clean_prompt['blur'] = "A sharp photo with high-quality."
self.degraded_prompt['blur'] = "A blurry photo with low-quality."
self.clean_prompt['content'] = "A clear content photo with high-quality."
self.degraded_prompt['content'] = "A bad content photo with low-quality."
self.clean_prompt['exposure'] = "A natural exposure photo with high-quality."
self.degraded_prompt['exposure'] = "A unnatural exposure photo with low-quality."
```

<논문에서 쓰인 text prompt P_d^-, P_d^+ >

CVSP Lab
Computer Vision & Signal Processing

Paper 2

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Fusing text-driven degradation
      - Feature domain에서 text-driven prompts로 degradation을 합성
        - ✓Feature domain 상에서 두 feature 간 차이를 구함
          - $T_d^- - T_d^+$: Text feature 간의 차

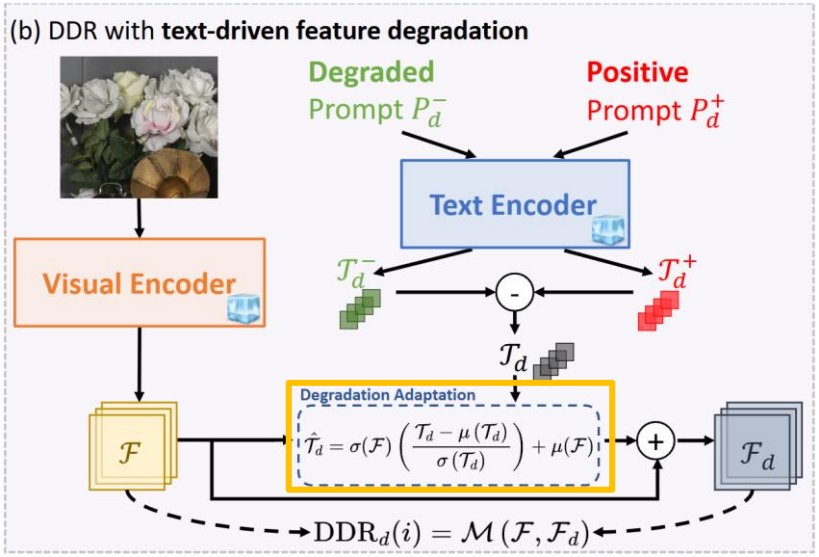


CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

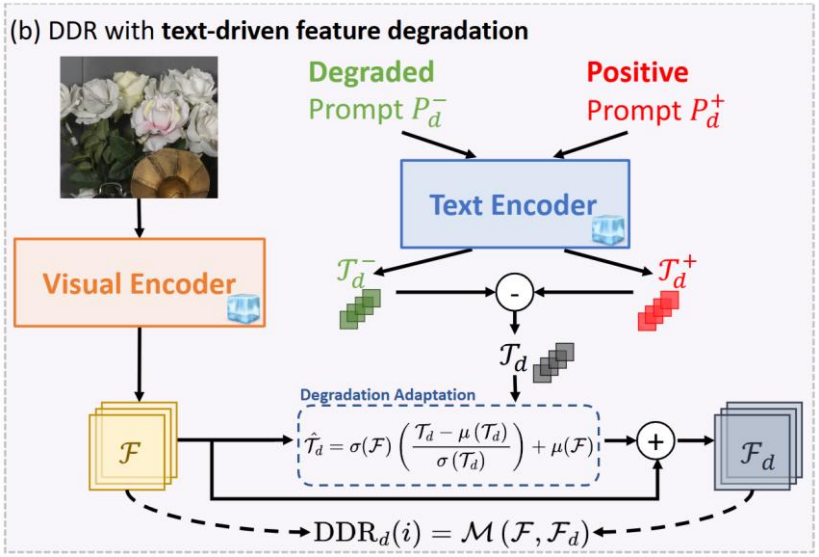
- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - Fusing text-driven degradation
      - Feature domain에서 text-driven prompts로 degradation을 합성
        - ✓Feature domain 상에서 두 feature 간 차이를 구함
          - $T_d^- - T_d^+$: Text feature 간의 차
        - ✓Image와 text의 도메인이 다르기 때문에 AdaIN의 컨셉을 따라 normalize
          - Text feature를 image feature domain에 맞춰 주고 더함
          - $\hat{T}_d = \sigma(F)\frac{T_d - \mu(T_d)}{\sigma(T_d)} + \mu(F)$, clean image feature F + $\hat{T}_d = F_d$



CVSP Lab
Computer Vision & Signal Processing

Paper 2

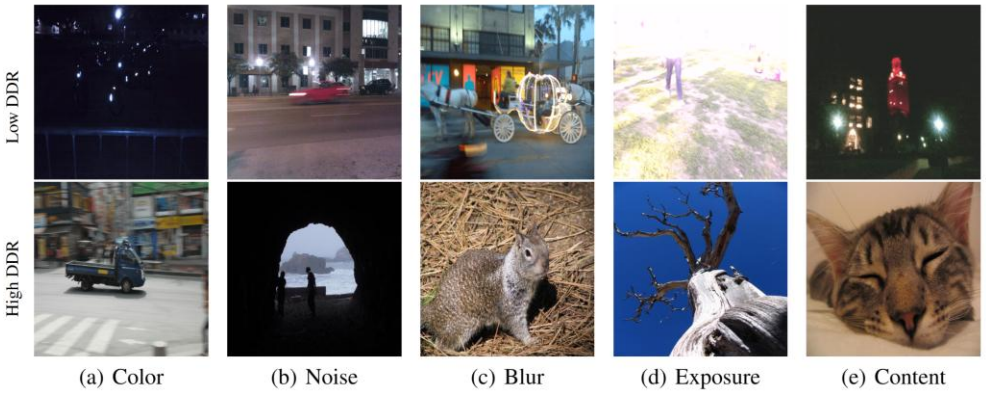
- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - DDR as a Flexible Image Descriptor
      - DDR과 image의 degradation정도는 상관관계가 있음
        - ✓Degradation유형별로 프롬프트 설정 후 DDR 측정
          - Blur 유형의 경우 'blurry'와 'sharp'로 설정
        - ✓DDR과 degradation 수준 간 음의 상관관계
          - DDR 값 ↑ degradation 정도 ↓ : high quality image
          - DDR 값 ↓ degradation 정도 ↑ : low quality image



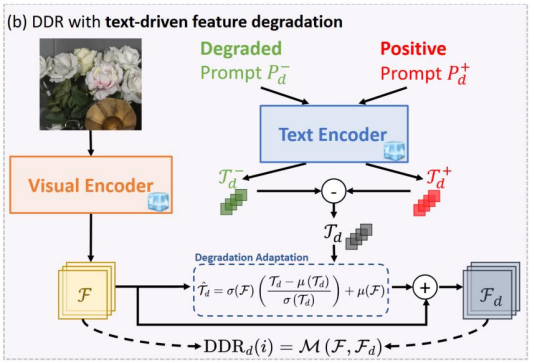
CVSP Lab
Computer Vision & Signal Processing

Paper 2

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - DDR as a Blind Image Quality Assessment Metric
      - DDR과 이미지의 품질 점수 간 양의 상관관계가 있음
        - ✓특정 degradation에 대해 높은 DDR 값을 가진 이미지는 더 높은 quality를 가질 가능성이 있음



<DDR 값과 degradation 유형에 따른 이미지 예시>



<DDR 정의>

CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - DDR as a Blind Image Quality Assessment Metric
      - DDR과 이미지의 품질 점수 간 양의 상관관계가 있음
        - ✓특정 degradation에 대해 높은 DDR 값을 가진 이미지는 더 높은 quality를 가질 가능성이 있음
        - ✓Ex) degradation type이 color일 때:
          - colorfulness와 sharpness, quality측면에서의 IQA점수와 양의 상관관계가 있음
          - DDR이 IQA metric으로 가치가 있음 → 사람의 품질평가 데이터셋 없이도 IQA가 가능함을 보여줌

Degradation Type	Complexity [25]	Colorfulness [23]	Sharpness [44]	Quality
color	-0.223	0.757	0.715	0.790
noise	-0.444	0.673	0.600	0.694
blur	-0.206	0.612	0.732	0.756
exposure	-0.435	0.684	0.699	0.770
content	-0.357	0.612	0.561	0.642

<degradation 유형별 DDR과 IQA 점수의 SRCC>

* Spearman 순위 상관계수(SRCC): 각 데이터 분포들을 단순히 순위로 보고 상관계수를 구함

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Method
    - DDR as an Unsupervised Learning Objective
      - Image restoration task에서 모델을 학습하는 데 사용할 수 있음
        - ✓Image restoration: degradation image에서 clean image를 복구하도록 예측
          - 정답 이미지와 예측한 이미지의 픽셀 간 reconstruction loss를 사용해 모델 학습
        - ✓DDR as learning objective
          - Reconstruction loss와 동시에 DDR 최대화를 학습목표로 해서 모델 학습

$$\min_{\theta}\left(\mathcal{L}_{rec}\left(R_{\theta}(i), i_{gt}\right) - \lambda_d \sum_{d\in\mathcal{D}}^{d} DDR_d(R_{\theta}(i))\right)$$

CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Experiments
    - (1) Opinion Unaware Blind Image Quality Assessment(OU-BIQA)
      - 인간의 주관적 평가 없이 이미지의 quality 평가
    - (2) Image restoration tasks
      - DDR 을 학습 loss로 추가하여 사용했을 때 성능 향상
        - ✓Motion deblurring
        - ✓Single Image Super Resolution

CVSP Lab
Computer Vision & Signal Processing

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Experiments
    - (1) Opinion Unaware Blind Image Quality Assessment(OU-BIQA)
      - IQA(Image Quality Assessment) 데이터셋의 정답들과 DDR의 품질평가 간 SRCC측정*
        - ✓IQA 데이터셋의 품질점수들과 DDR 값의 SRCC가 높음
          - 사람이 작성한 IQA dataset을 학습하지 않고도, clean pair dataset 없이도 Image Quality assessment metric으로 사용할 수 있음을 보임

Datasets	NIQE	QAC	PIQE	LPSI	ILNIQE	dipIQ	SNP-NIQE	NPQI	ContentSep	Ours
CSIQ	0.6191	0.4804	0.5120	0.5218	0.8045	0.5191	0.6090	0.6341	0.5871	**0.8289**
LIVE	0.9062	0.8683	0.8398	0.8181	0.8975	**0.9378**	0.9073	0.9108	0.7478	0.8793
TID2013	0.3106	0.3719	0.3636	0.3949	0.4938	0.4377	0.3329	0.2804	0.2530	**0.5844**
KADID	0.3779	0.2394	0.2372	0.1478	0.5406	0.2977	0.3719	0.3909	0.5060	**0.5968**
KonIQ	0.5300	0.3397	0.2452	0.2239	0.5057	0.2375	0.6284	0.6132	0.6401	**0.6455**
LIVEitw	0.4495	0.2258	0.2325	0.0832	0.4393	0.2089	0.4654	0.4752	0.5060	**0.6613**
CID2013	0.6589	0.0299	0.0448	0.3229	0.3062	0.3776	0.7159	0.7698	0.6116	**0.8009**
SPAQ	0.3105	0.4397	0.2317	0.0001	0.6959	0.2189	0.5402	0.5999	0.7084	**0.7249**

<DDR과 IQA 점수 간 순위상관계수>

* Spearman 순위 상관계수(SRCC): 각 데이터 분포들을 단순히 순위로 보고 상관계수를 구함

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Experiments
    - (2) Image restoration task
      - Motion deblurring
        - ✓PSNR loss만 최적화 했을 때보다 predict된 이미지의 DDR을 최대화 하면 deblurring이 더 효과적임
          - Deblurring dataset에서 PSNR과 SSIM기준 성능이 제일 높음

model	loss	GoPro [51]		RealBlur [52]	
		PSNR	SSIM	PSNR	SSIM
NAFNet	PSNR	33.1717	0.9482	30.6373	0.9038
	PSNR + LPIPS [3]	33.1660	0.9481	30.7245	0.9044
	PSNR + CTX [2]	32.7879	0.9436	30.4394	0.8985
	PSNR + PDL [7]	32.9417	0.9463	30.6270	0.9039
	PSNR + FDL [8]	32.8321	0.9420	30.1743	0.8864
	PSNR + DDR(ours)	**33.3427**	**0.9500**	**30.7982**	**0.9049**
Restormer	PSNR	33.3398	0.9494	31.9816	0.9098
	PSNR + LPIPS [3]	33.3717	0.9495	31.9639	0.9099
	PSNR + CTX [2]	33.2834	0.9483	31.9893	0.9101
	PSNR + PDL [7]	33.2905	0.9487	31.9900	0.9106
	PSNR + FDL [8]	33.3560	0.9489	31.7673	0.9034
	PSNR + DDR(ours)	**33.4946**	**0.9513**	**32.1759**	**0.9121**

<DDR을 학습목표로 추가했을 때 deblurring 성능>

Paper 2

3) Wu, Juncheng, et al. "DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor." arXiv preprint arXiv:2406.08377 (2024).

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Experiments
    - (2) Image restoration task
      - Motion deblurring
        - ✓PSNR loss만 최적화 했을 때보다 predict된 이미지의 DDR을 최대화 하면 deblurring이 더 효과적임
          - Deblurring dataset에서 PSNR과 SSIM기준 성능이 제일 높음

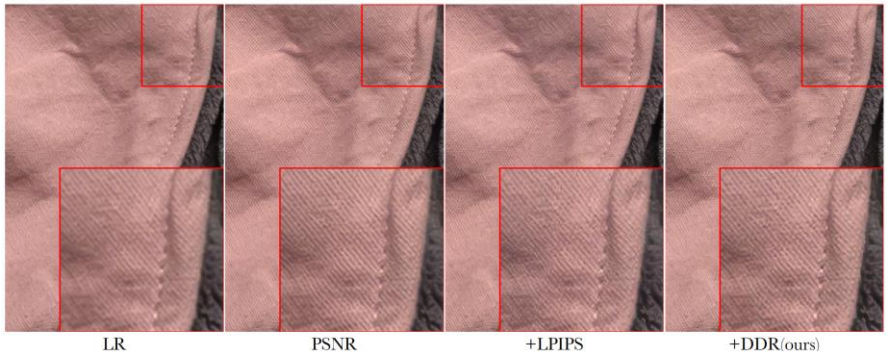


<DDR을 학습목표로 추가했을 때 deblurring 결과 시각화>

CVSP Lab
Computer Vision & Signal Processing

Paper 2

- DDR: Exploiting Deep Degradation Response as Flexible Image Descriptor[3)]
  - Experiments
    - (2) Single Image Super Resolution task
      - Predict된 이미지의 DDR을 최대화하는 loss를 추가
        - ✓SISR super resolution dataset을 통해 측정
        - ✓PSNR 단독, LPIPS를 추가하여 학습했을 때보다 DDR을 최대화 하는 loss를 추가
          - PSNR, SSIM기준



loss	NAFNet [62]		Restormer [63]	
	PSNR	SSIM	PSNR	SSIM
PSNR	27.0856	0.8917	28.1491	0.8986
PSNR + LPIPS [3]	27.2835	**0.8938**	28.1221	0.8985
PSNR + CTX [2]	27.0985	0.8867	28.0933	0.8964
PSNR + PDL [7]	26.9467	0.8907	28.1413	0.8985
PSNR + FDL [8]	27.0263	0.8809	28.0746	0.8966
PSNR + DDR(ours)	**27.3121**	0.8923	**28.1668**	**0.8990**

<DDR을 학습목표로 추가했을 때 Super Resolution 성능>

Thank you

CVSP Lab
Computer Vision & Signal Processing