IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

F2LM 생성자와 파괴자를 이용한 비디오 이상 탐지 방법

Seungkyun Hong*, Sunghyun Ahn*, Youngwan Jo, Sanghyun Park†

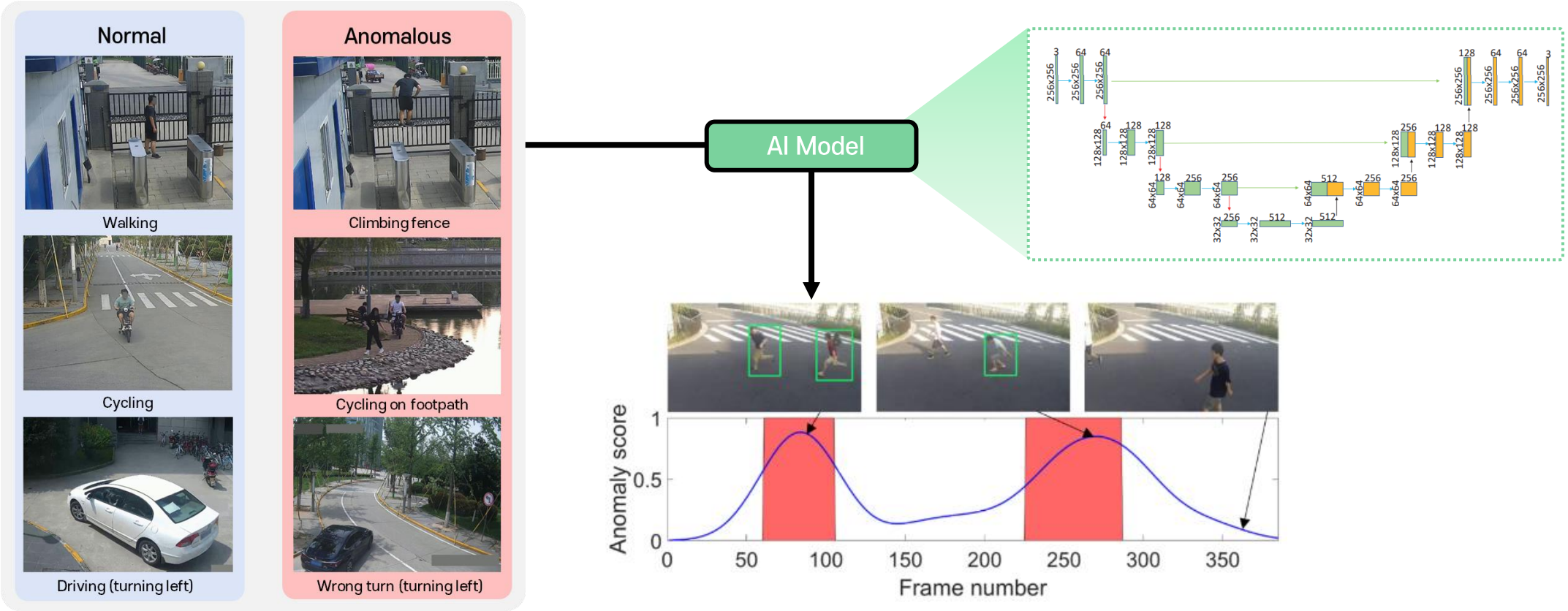
Department of Computer Science, Yonsei University
Seoul, Republic of Korea
{highsk, skd, jyy1551, sanghyun}@yonsei.ac.kr

* Equal contribution
† Corresponding author

Introduction

Video Anomaly Detection

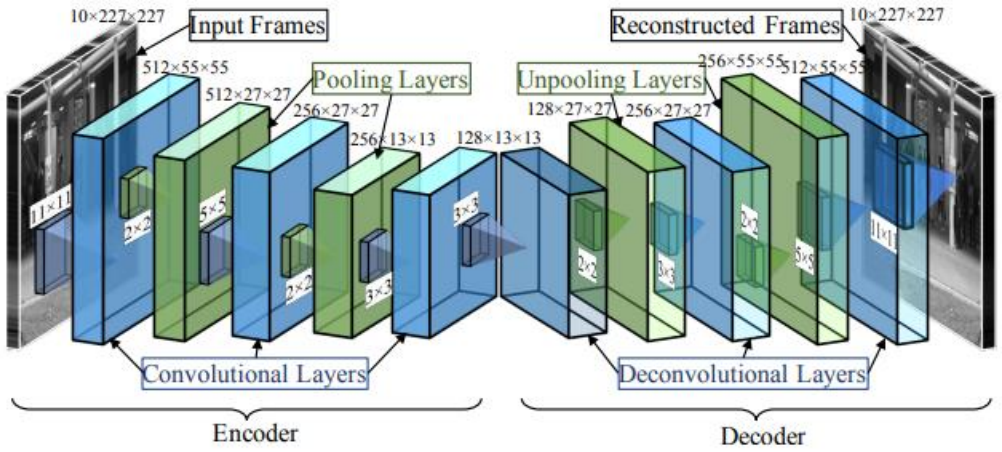
- 비디오에서 비정상적인 상황을 감지하여 피해를 예방하는 기술
- 비정상 상황은 배경에 적합하지 않은 객체의 출현이나 행동을 포함할 수 있음
- 비정상 데이터의 수가 부족하므로, 제한된 비정상의 수에 적응 가능한 이상 탐지 AI 모델 연구가 필요함



Introduction

Unsupervised Video Anomaly Detection

- 정상 데이터로만 학습하고, 정상 데이터의 패턴과 유사하지 않으면 모두 비정상으로 판단하는 이상 탐지 방법
- 정상 데이터의 수가 비정상의 수보다 훨씬 많기 때문(클래스 불균형), Real world의 비정상은 어떤 형태일지 예상할 수 없기 때문(과적합)
- 프레임 재구축 혹은 미래 프레임 예측 방식으로 학습하고, 정답 영상과의 유사도(ex. PSNR)를 패턴으로 이용하여 이상 탐지를 시도함



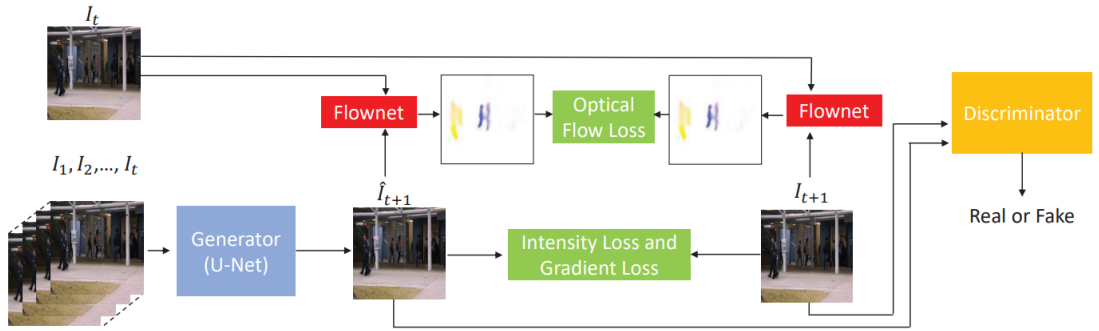


Figure 2. The pipeline of our video frame prediction network. Here we adopt U-Net as generator to predict next frame. To generate high quality image, we adopt the constraints in terms of appearance (intensity loss and gradient loss) and motion (optical flow loss). Here Flownet is a pretrained network used to calculate optical flow. We also leverage the adversarial training to discriminate whether the prediction is real or fake.

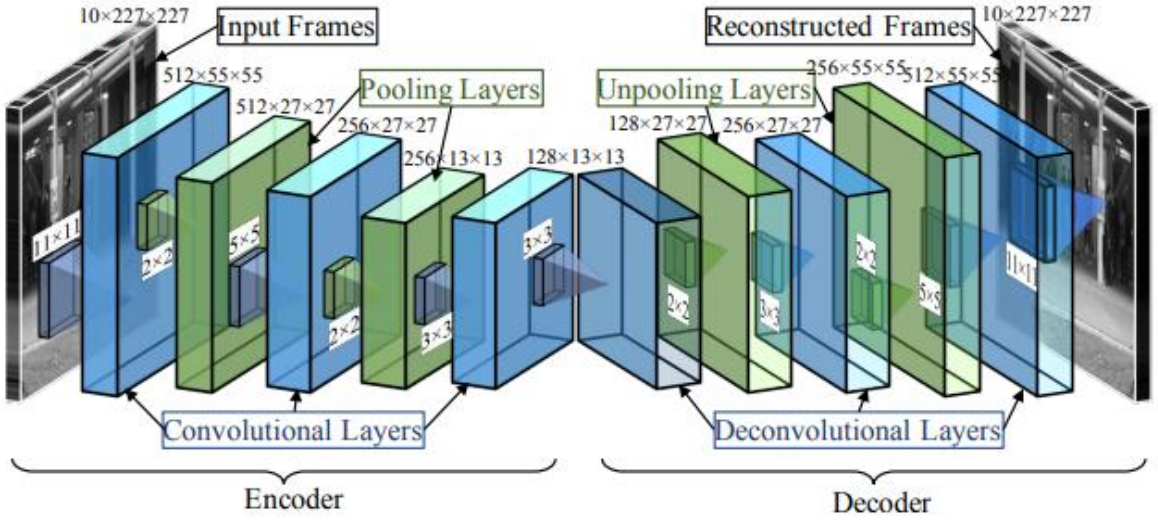
Hasan, Mahmudul, et al. "Learning temporal regularity in video sequences." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Liu, Wen, et al. "Future frame prediction for anomaly detection–a new baseline." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Introduction

Frame Reconstruction

- 정상 프레임만 재구축하도록 학습하면 비정상 프레임은 재구축하지 못 한다는 가정
- 정상 프레임의 PSNR과 비정상 프레임의 PSNR 차이를 통해 이상 탐지를 시도
- 재구축 모델의 높은 일반화 능력으로 인해 비정상 프레임도 재구축한다는 문제점이 발생함



Hasan, Mahmudul, et al. "Learning temporal regularity in video sequences." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Introduction

Future Frame Prediction

- 정상 미래 프레임만 예측하도록 학습하면 비정상 미래 프레임은 못 예측한다는 가정
- 정상 미래 프레임의 PSNR과 비정상 미래 프레임의 PSNR 차이를 통해 이상 탐지를 시도
- 정상과 비정상 간의 PSNR 차이가 기대만큼 크지 않아 성능(AUC) 차이도 크지 않음

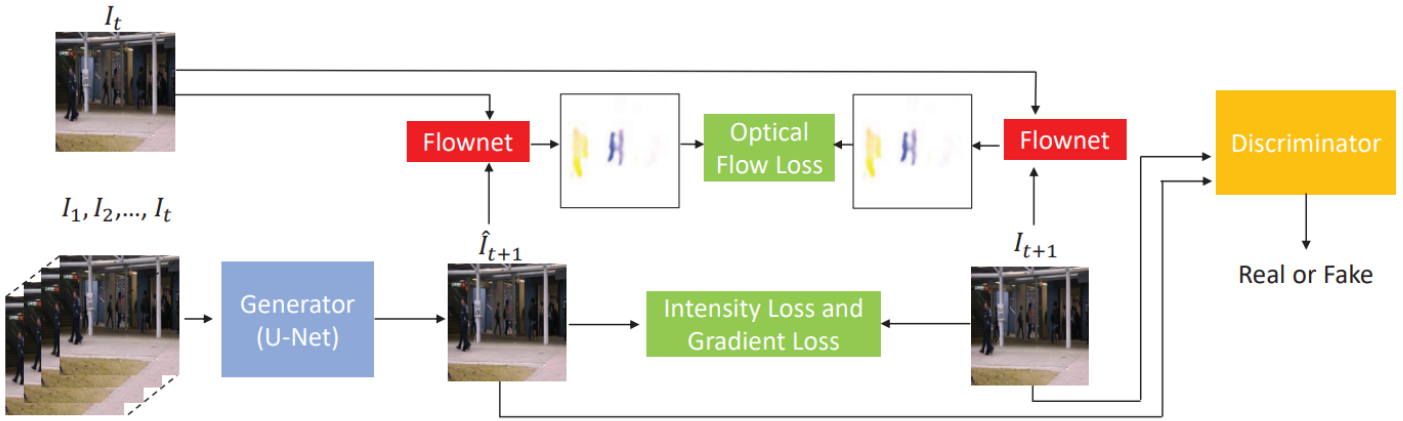
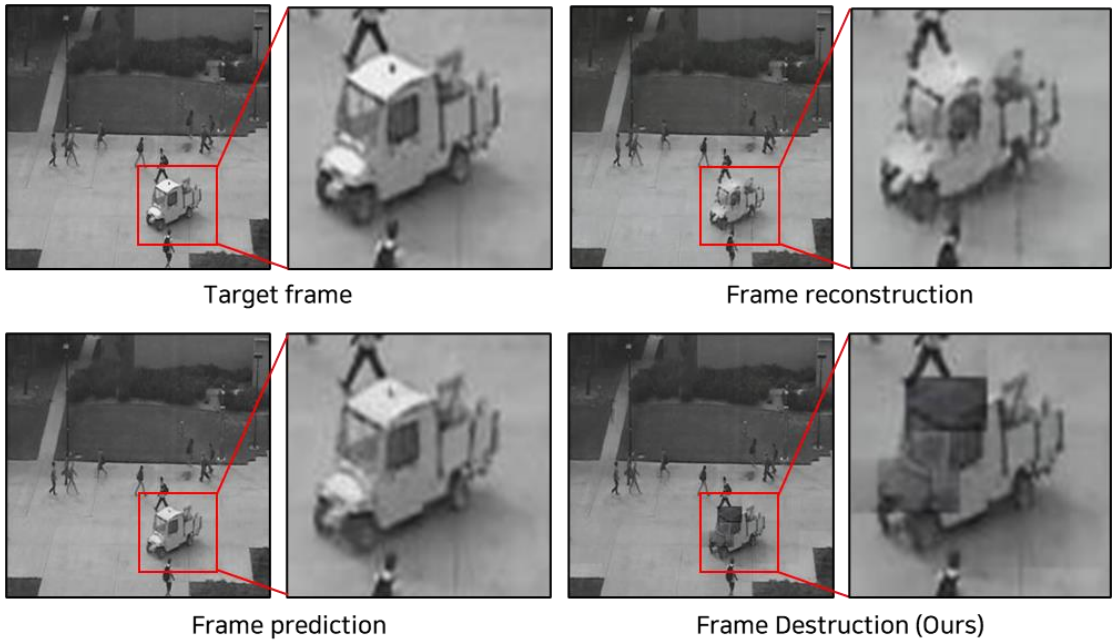


Figure 2. The pipeline of our video frame prediction network. Here we adopt U-Net as generator to predict next frame. To generate high quality image, we adopt the constraints in terms of appearance (intensity loss and gradient loss) and motion (optical flow loss). Here Flownet is a pretrained network used to calculate optical flow. We also leverage the adversarial training to discriminate whether the prediction is real or fake.

Liu, Wen, et al. "Future frame prediction for anomaly detection–a new baseline." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Making Anomalies More Anomalous

- 프레임에 위치한 비정상 영역(ex. 트럭)을 정답과 유사하지 않도록 생성하는 것을 목표로 함
- 비정상 영역을 관련이 없는 검은 영상으로 변환하여 더 비정상처럼 만드는 Frame Destruction 방법을 고안함

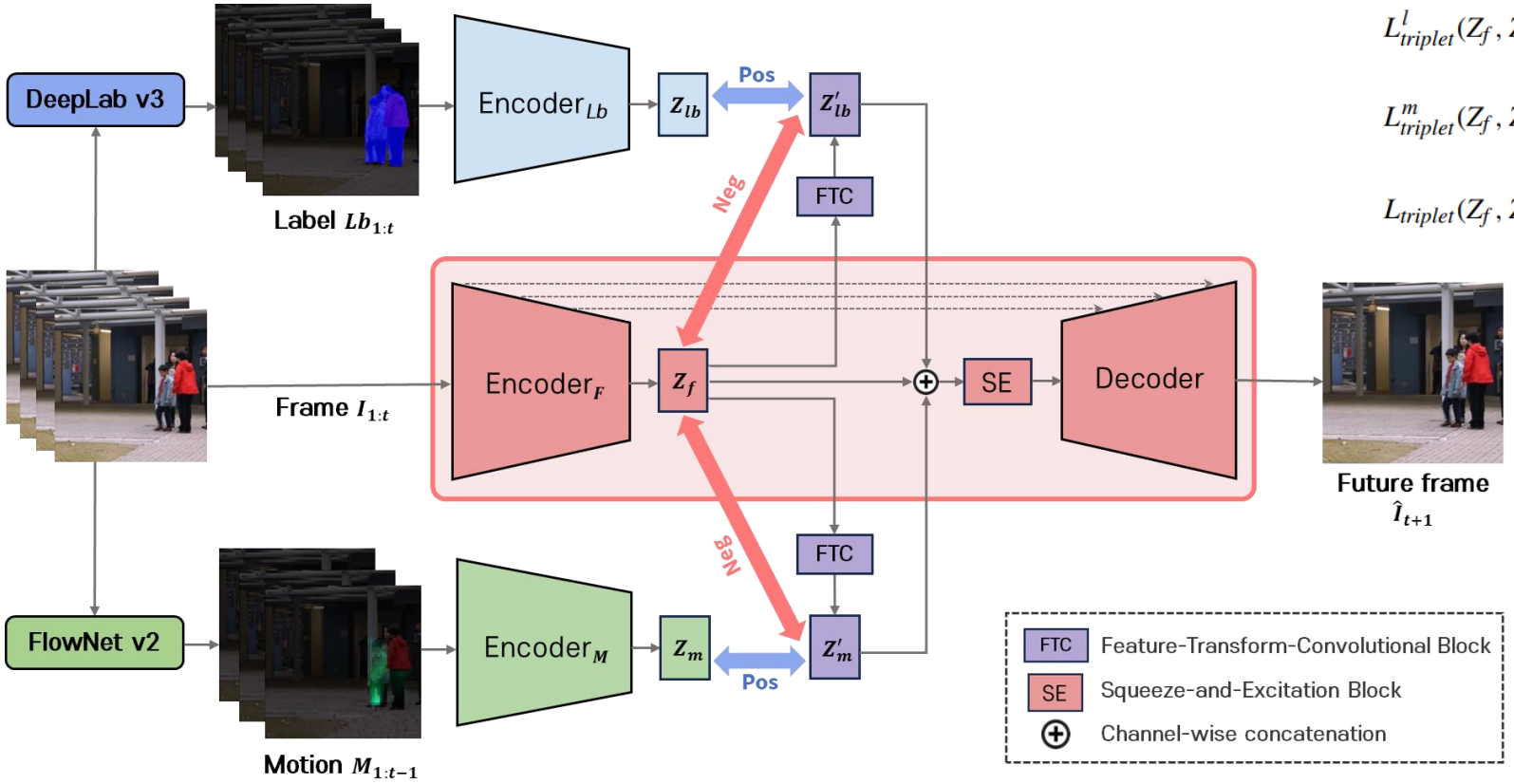


- 정상 미래 프레임은 잘 예측하고 비정상 미래 프레임은 못 예측하는 새로운 Generator, F2LM Generator(F2LM 생성자) 제안
- 잘 예측된 미래 프레임은 유지하고 못 예측된 미래 프레임은 파괴하는 Destroyer(파괴자) 제안 (파괴: 관련없는 영상으로 변환시킴)

Method

F2LM Generator

- 라벨, 프레임, 모션 정보를 입력받아 feature level에서 다른 타입으로 변환을 한 다음, 융합하여 미래 프레임을 예측하는 모델
- 학습 관점: 디코더는 의미있는 라벨, 프레임, 모션 정보를 입력으로 받아 미래 프레임을 잘 예측함 (어떤 객체가 어떻게 움직이는지의 정보)
- 테스트 관점: 비정상이 입력되면 처음 보는 라벨, 모션 특징으로 변환이 어렵고, 노이즈 융합 특징이 디코더에 입력되어 올바른 생성이 힘듦



$$L^l_{triplet}(Z_f, Z_l, Z'_l) = \max\{d(Z'_l, Z_l) - d(Z'_l, Z_f) + \alpha, 0\},$$

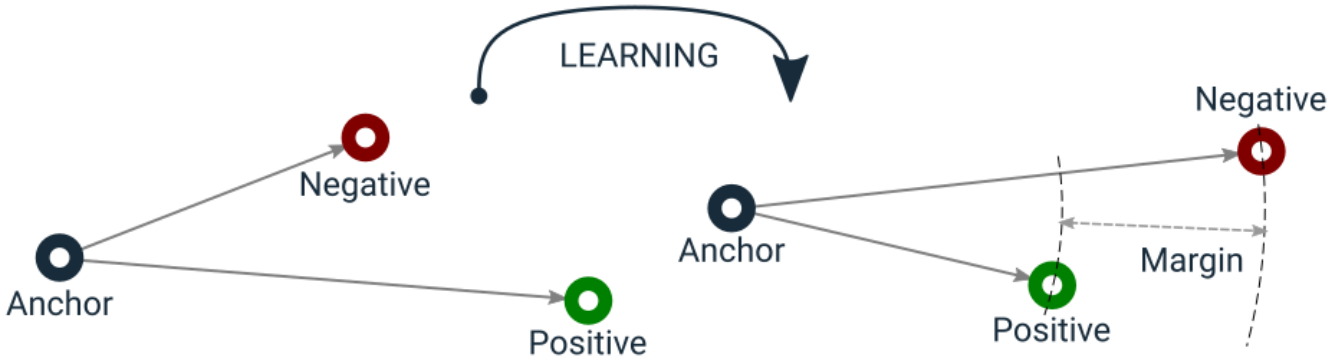
$$L^m_{triplet}(Z_f, Z_m, Z'_m) = \max\{d(Z'_m, Z_m) - d(Z'_m, Z_f) + \alpha, 0\},$$

$$L_{triplet}(Z_f, Z_l, Z_m, Z'_l, Z'_m) = L^l_{triplet}(Z_f, Z_l, Z'_l) + L^m_{triplet}(Z_f, Z_m, Z'_m),$$

Method

Triplet Loss

- Anchor와 Positive 간의 거리에 Margin을 더한만큼 Negative와 멀어지도록 유도하는 목적함수
- 입력(a, p, n)에 L2 정규화를 적용해서 margin을 선택하는 것을 용이하게 하였음



$$L(a, p, n) = \max\{d(a_i, p_i) - d(a_i, n_i) + \text{margin}, 0\}$$

$$d(x_i, y_i) = \|\mathbf{x}_i - \mathbf{y}_i\|_p$$

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

Method

U-Net Structure

- FFP(Future Frame Prediction)에서 사용한 U-Net Encoder와 동일함, Triplet loss를 계산하기 위해 L2 Norm을 적용함
- FFP에서 사용한 U-Net Decoder와 동일함, 첫 입력($32 \times 32 \times 512$)으로 SEBlock을 통과한 융합 특징을 받음

TABLE 1. Detailed network architecture of the encoder in the F2LM generator. Abbreviations: *k*: kernel, *p*: padding, *s*: stride, *H*: height, *W*: width, *C*: channel.

Layer name	Layer	Input size ($H \times W \times C$)	Output size ($H \times W \times C$)
inconv	[Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$256 \times 256 \times C$	$256 \times 256 \times 64$
downconv$_1$	MaxPool2d(k=2, s=2) [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$256 \times 256 \times 64$	$128 \times 128 \times 128$
downconv$_2$	MaxPool2d(k=2, s=2) [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$128 \times 128 \times 128$	$64 \times 64 \times 256$
downconv$_3$	MaxPool2d(k=2, s=2) [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$64 \times 64 \times 256$	$32 \times 32 \times 512$
LN	L_2 normalization	$32 \times 32 \times 512$	$32 \times 32 \times 512$

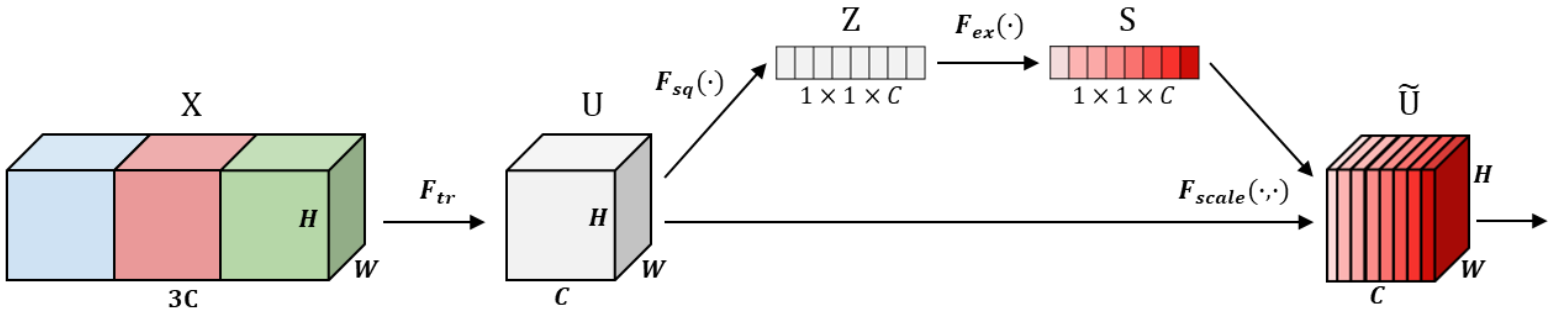
TABLE 2. Detailed network architecture of the decoder in the F2LM generator. Abbreviations: *k*: kernel, *p*: padding, *s*: stride, *H*: height, *W*: width, *C*: channel.

Layer name	Layer	Input size ($H \times W \times C$)	Output size ($H \times W \times C$)
upconv$_1$	ConvTranspose2d(k=2,s=2) channel-wise concat [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$32 \times 32 \times 512$ $64 \times 64 \times 256$	$64 \times 64 \times 256$
upconv$_2$	ConvTranspose2d(k=2,s=2) channel-wise concat [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$64 \times 64 \times 256$ $128 \times 128 \times 128$	$128 \times 128 \times 128$
upconv$_3$	ConvTranspose2d(k=2,s=2) channel-wise concat [Conv2D (k=3, p=1,s=1) BatchNorm2d ReLU] $\times$ 2	$128 \times 128 \times 128$ $256 \times 256 \times 64$	$256 \times 256 \times 64$
outconv	Conv2D (k=3, p=1,s=1) tanh	$256 \times 256 \times 64$	$256 \times 256 \times 3$

Method

SE Block Structure

- 모션, 프레임, 라벨 정보를 채널 방향으로 합치고, 채널 어텐션을 통해 중요한 채널 특징을 선별하였음
- 이러한 구조를 통해 미래 프레임을 더 잘 예측할 수 있음



$$X = [Z_l'; Z_f; Z_m']$$

$$u_c = \delta(f_c^{3\times3}(X))$$

$$z_c = F_{sq}(u_c) = AvgPool(u_c)$$

$$S = F_{ex}(Z) = \sigma\big(MLP(Z)\big) = \sigma(W_2\delta(W_1 Z))$$

$$\tilde{u}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c$$

$$U = [u_1; u_2; \ldots; u_c]$$

$$Z = [z_1; z_2; \ldots; z_c]$$

$$S = [s_1; s_2; \ldots; s_c]$$

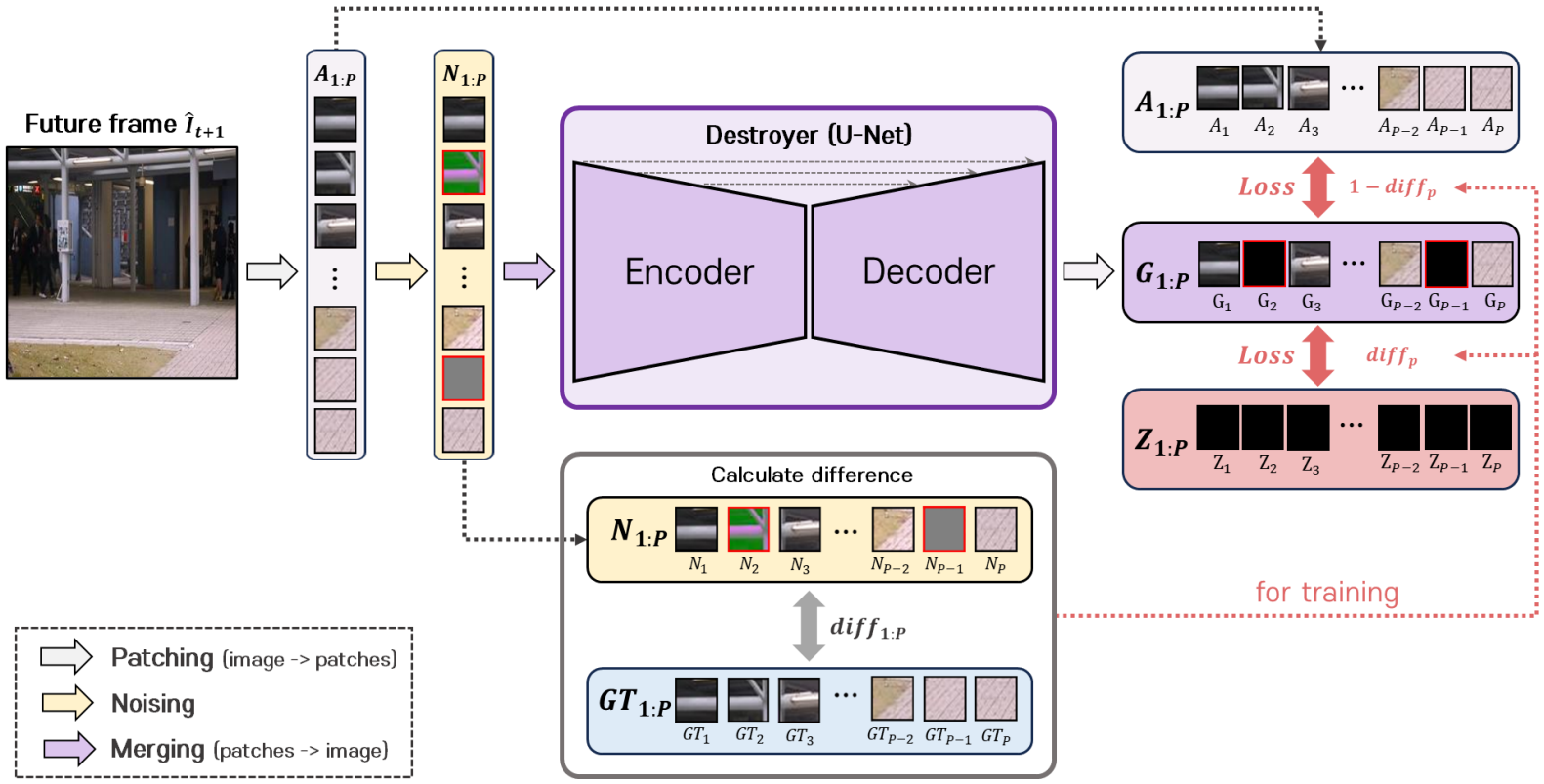
$$\tilde{U} = [\tilde{u}_1; \tilde{u}_2; \ldots; \tilde{u}_c]$$

Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

Method

Destroyer

- 미래 프레임의 패치 중 잘 예측된 패치는 유지하고 못 예측된 패치는 파괴하는 모델
- 미래 프레임은 정상 데이터만으로 학습한 F2LM 생성자의 출력이므로 대부분 잘 예측된 패치임 (노이즈를 주입하여 비정상 패치를 생성하여 학습)

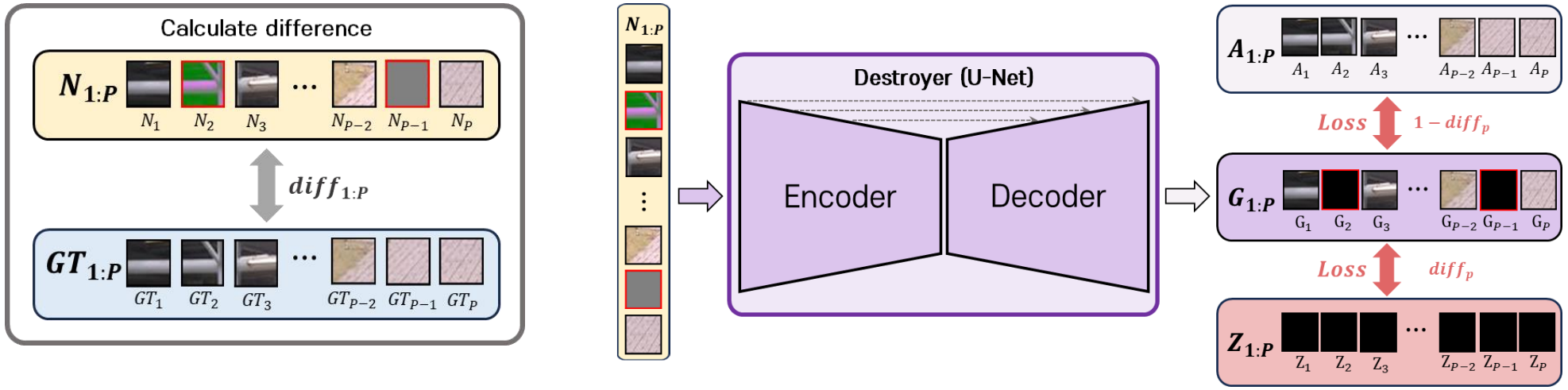


$$L_{Destroyer}(diff, R, Z, A) = \sum_{p=1}^{P} (diff_p \cdot \|Z_p - R_p\|_2^2 + (1 - diff_p) \cdot \|A_p - R_p\|_2^2).$$

Method

Destroyer

- Destoyer Loss: diff가 높은 패치는 Z(제로 패치)로 변환하고, 낮은 패치는 A(원본 패치)로 변환하는 목적함수
- diff: 이미지의 품질을 평가하는 SSIM을 사용함, Destroyer loss 각 항의 가중치를 0~1로 조정하기 위해 MIN 함수를 사용함
- lambda: 비정상 패치의 diff가 기대보다 작은 경우, Z로 학습이 잘 되지 않으므로, 하이퍼 파라미터 lambda를 통해 Z로 학습되는 강도를 높임

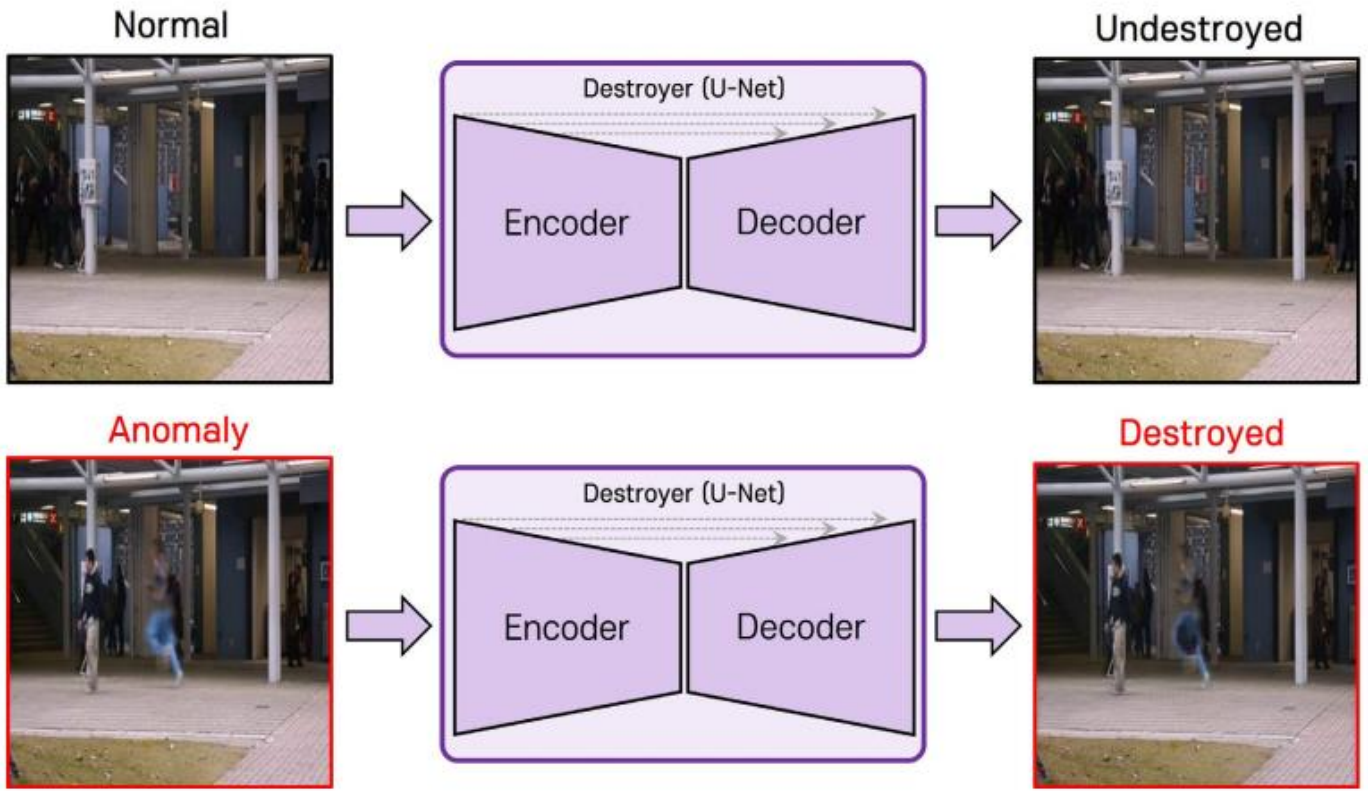


$$diff_p = \mathrm{MIN}(\lambda(1 - \mathrm{SSIM}(N_p, GT_p)), 1), \quad L_{Destroyer}(diff, R, Z, A) = \sum_{p=1}^{P} (diff_p \cdot \|Z_p - R_p\|_2^2 + (1 - diff_p) \cdot \|A_p - R_p\|_2^2).$$

Method

Destroyer

- 자기-지도학습을 통해 품질이 낮은 패치를 파괴하였으므로, 테스트 과정에서 품질이 낮은 비정상 패치가 파괴됨



Method

Loss Function

- Intensity loss: 픽셀 간 유사도를 높이기 위한 목적함수
- Gradient loss: 생성된 영상을 더 부드럽게(자연스럽게) 하기 위한 목적함수

$$L_{int}(\hat{I}, I) = \|\hat{I} - I\|_2^2,$$

$$L_{gd}(\hat{I}, I) = \sum_{i,j} \| |\hat{I}_{i,j} - \hat{I}_{i-1,j}| - |I_{i,j} - I_{i-1,j}| \|_1 + \| |\hat{I}_{i,j} - \hat{I}_{i,j-1}| - |I_{i,j} - I_{i,j-1}| \|_1,$$

Method

Loss Function

- Discriminator adversarial loss: 정답 미래 프레임과 생성된 미래 프레임을 잘 구분하기 위한 목적함수
- Generator adversarial loss: Discriminator를 속일 정도로 미래 프레임을 잘 생성하기 위한 목적함수
- Triplet loss: frame에서 label과 motion feature로 변환하기 위한 목적함수

$$L_{adv}^{D}(\hat{I}, I) = \sum_{i,j} \frac{1}{2} L_{MSE}(D(I)_{i,j}, 1) + \sum_{i,j} \frac{1}{2} L_{MSE}(D(\hat{I})_{i,j}, 0),$$

$$L_{adv}^{G}(\hat{I}) = \sum_{i,j} \frac{1}{2} L_{MSE}(D(\hat{I})_{i,j}, 1),$$

$$L_{triplet}^{l}(Z_f, Z_l, Z_l') = \max\{d(Z_l', Z_l) - d(Z_l', Z_f) + \alpha, 0\},$$

$$L_{triplet}^{m}(Z_f, Z_m, Z_m') = \max\{d(Z_m', Z_m) - d(Z_m', Z_f) + \alpha, 0\},$$

$$L_{triplet}(Z_f, Z_l, Z_m, Z_l', Z_m') = L_{triplet}^{l}(Z_f, Z_l, Z_l') + L_{triplet}^{m}(Z_f, Z_m, Z_m'),$$

Method

Loss Function

- Generator Loss: F2LM 생성자를 훈련하기 위한 최종 목적함수
- Destroyer Loss: 파괴자를 훈련하기 위한 목적함수

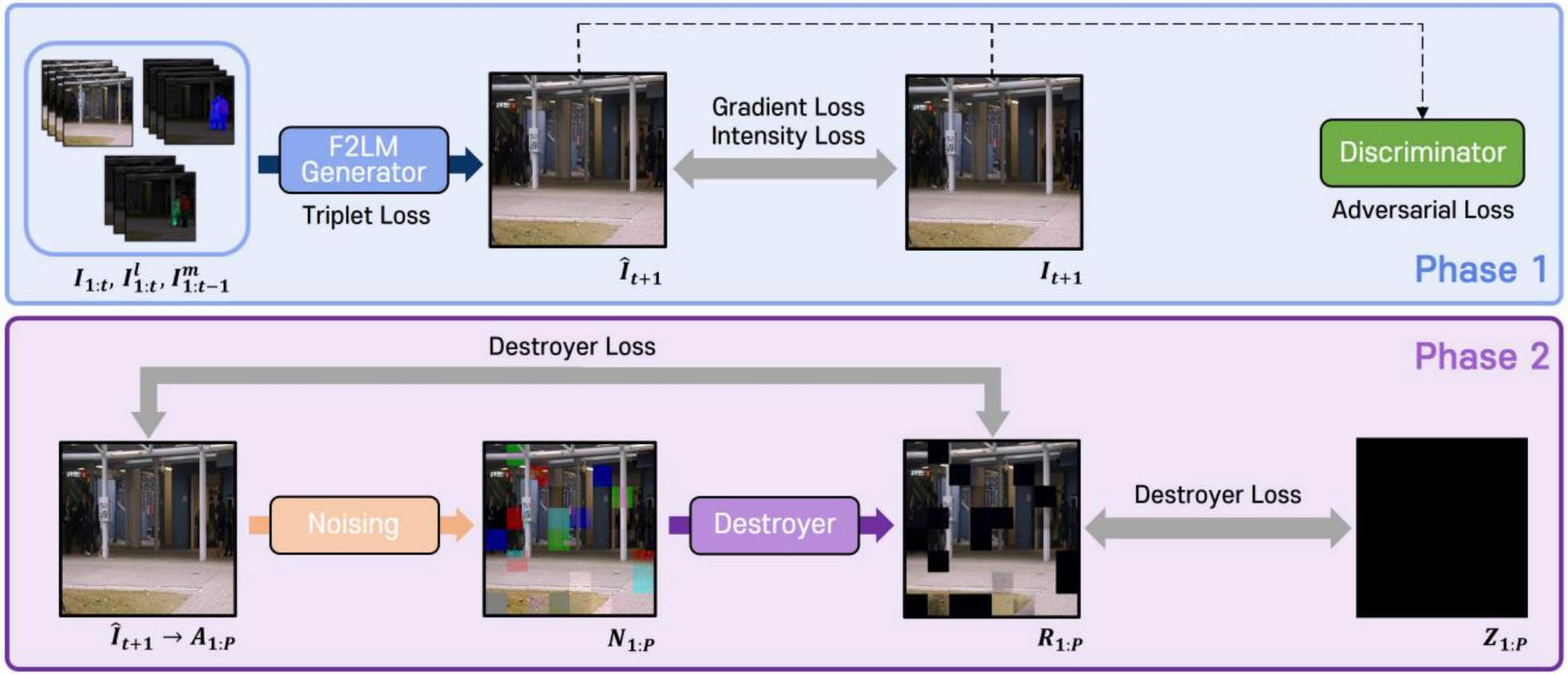
$$L_{Generator} = \delta_{int} L_{int}(\hat{I}, I) + \delta_{gd} L_{gd}(\hat{I}, I) + \delta_{adv} L_{adv}^{G}(\hat{I}) + \delta_{tri} L_{triplet}(Z_f, Z_l, Z_m, Z_l', Z_m'),$$

$$L_{Destroyer}(diff, R, Z, A) = \sum_{p=1}^{P} (diff_p \cdot \|Z_p - R_p\|_2^2 + (1 - diff_p) \cdot \|A_p - R_p\|_2^2).$$

Method

Training process

- Phase 1에는 F2LM 생성자를 학습하고, Phase2에는 파괴자를 학습함으로써 두 모델을 각각 안정적으로 최적화함



Method

Anomaly Scoring

- Feature level에서 얼마나 변환이 잘 이루어졌는지 확인함(1,2), Frame level에서 얼마나 정답과 유사한지 확인함(3,4)
- 모든 점수를 조합하고, 0과 1 사이의 점수로 정규화를 진행함

$$
\begin{aligned}
Scaled\ Score = {} & \gamma_1 \cdot SL^{l}_{triplet} + \gamma_2 \cdot SL^{m}_{triplet} \\
& + \gamma_3 \cdot SL^{Generator}_{MSE} + \gamma_4 \cdot SL^{Destroyer}_{MSE},
\end{aligned}
$$

$$
SL_{triplet} = \frac{L_{triplet} - \mu(L_{triplet})}{\sigma(L_{triplet})},
$$

$$
SL_{MSE} = \frac{L_{MSE} - \mu(L_{MSE})}{\sigma(L_{MSE})},
$$

$$
\begin{aligned}
Anomaly\ Score & \\
& = \frac{Scaled\ Score - \mathrm{MIN}(Scaled\ Score)}{\mathrm{MAX}(Scaled\ Score) - \mathrm{MIN}(Scaled\ Score)}.
\end{aligned}
$$

Experiments

Comparison with baseline

- F2LM 생성자를 통해 비정상 미래 프레임을 잘 생성하지 못하여서 AUC가 (2.1%, 3.1%, 1.5%) 증가함
- Destroyer를 통해 비정상 영역이 파괴되어 AUC가 (2.8%, 6.1%, 3.7%) 증가함

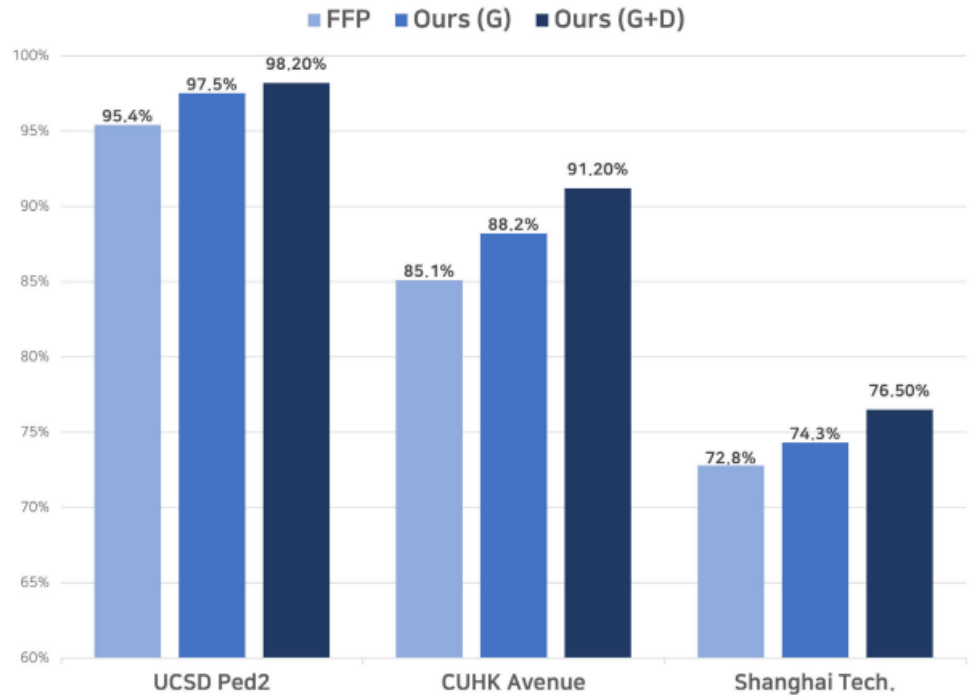


FIGURE 8. AUC comparison with baseline [10]. Abbreviations: FFP: future frame prediction, G: F2LM generator, D: Destroyer.

Experiments

Network design

- 하이퍼 파라미터를 변경하는 실험을 여러 차례 진행하여 최적의 모델을 설계함
- 하이퍼 파라미터 변경에 따른 성능 변화는 오른쪽과 같음

TABLE 10. Hyperparameters for the entire network.

Hyper-P	Value	Description
α	0.2	Margin of triplet loss
noise	CI	Noising method for the Destroyer training
patch size	32	Size of patch for patching
λ	4	Number for adjusting $diff_p$
Z_p	zero vector	Destroying method
δ_{int}	1	Weight of intensity loss
δ_{gd}	1	Weight of gradient loss
δ_{adv}	0.05	Weight of adversarial loss
δ_{tri}	1	Weight of triplet loss
γ_1	(0.02, 0.94, 0.68)	Weight of label triplet loss for testing
γ_2	(0.50, 0.02, 0.06)	Weight of motion triplet loss for testing
γ_3	(0.48, 0.04, 0.26)	Weight of F2LM generator MSE loss for testing
γ_4	(1.00, 1.00, 0.25)	Weight of Destroyer MSE loss for testing

TABLE 4. AUC comparison based on the feature fusion method. Best results are bolded.

Fusion method	UCSD Ped2	CUHK Avenue	Shanghai Tech.
Add	97.3%	87.5%	72.6%
Concat & SE block	**97.5%**	**88.2%**	**74.3%**

TABLE 5. AUC comparison based on the hyperparameter α of the triplet loss. Best results are bolded.

Triplet α	UCSD Ped2	CUHK Avenue	Shanghai Tech.
0.2	**97.5%**	**88.2%**	**74.3%**
0.5	97.0%	87.5%	72.4%
0.8	95.9%	87.0%	72.4%
1.0	95.6%	86.7%	72.3%

TABLE 6. AUC comparison based on the dropout noise method. Abbreviations: CD: channel dependent, CI: channel independent. Best results are bolded.

Noise method	UCSD Ped2	CUHK Avenue	Shanghai Tech.
None	97.6%	90.0%	75.1%
Dropout(CD)	97.8%	90.6%	75.7%
Dropout(CI)	**98.2%**	**91.2%**	**76.5%**

TABLE 7. AUC comparison according to patch size. Best results are bolded.

Patch size	UCSD Ped2	CUHK Avenue	Shanghai Tech.
16×16	97.8%	90.4%	75.8%
32×32	**98.2%**	**91.2%**	**76.5%**
64×64	97.7%	89.8%	76.1%
128×128	97.6%	89.7%	75.9%

TABLE 8. AUC comparison based on the hyperparameter λ of the Destroyer. Best results are bolded.

λ	UCSD Ped2	CUHK Avenue	Shanghai Tech.
1	97.5%	90.2%	75.6%
2	97.6%	90.4%	76.1%
3	97.8%	91.0%	76.3%
4	**98.2%**	**91.2%**	**76.5%**
5	98.0%	90.7%	76.2%
6	98.0%	90.6%	76.3%

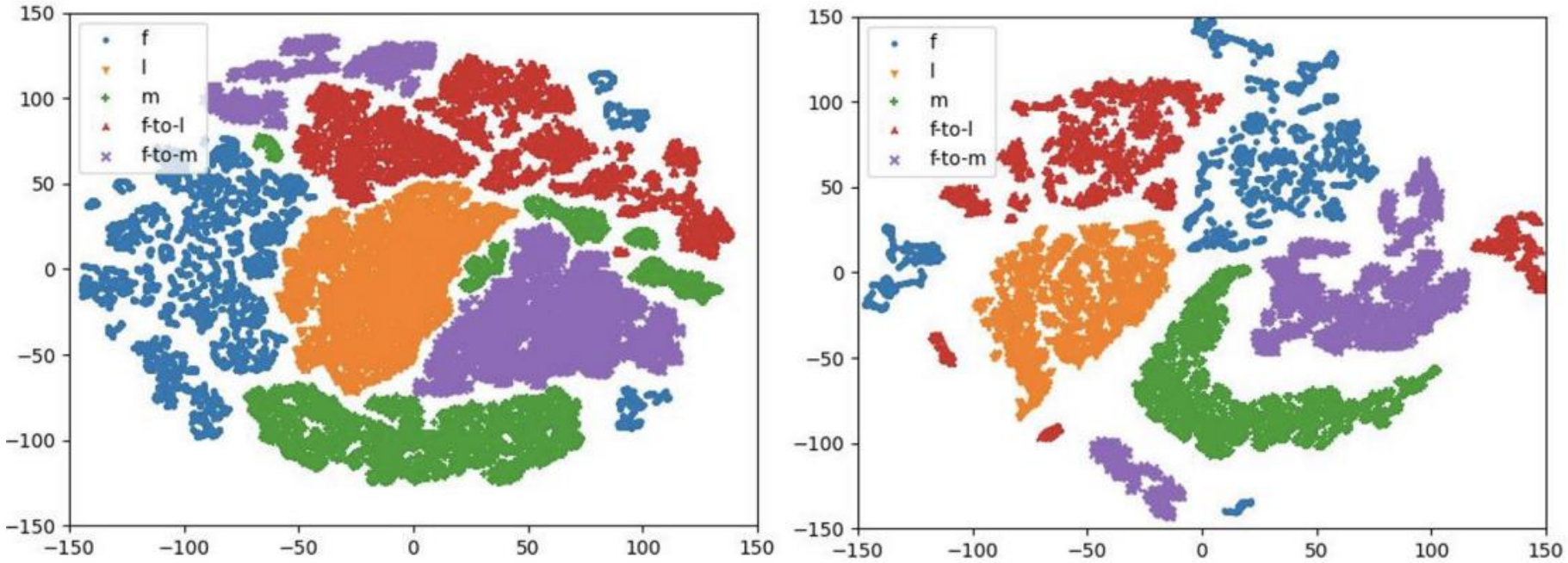
TABLE 9. AUC comparison based on the selection of Z_p. Best results are bolded.

Z_p	UCSD Ped2	CUHK Avenue	Shanghai Tech.
None	97.5%	88.2%	74.3%
Background vector	97.6%	91.0%	-
Zero vector	**98.2%**	**91.2%**	**76.5%**

Experiments

Qualitative analysis

- 정상 데이터는 Negative인 frame 간의 거리는 멀고, Positive인 motion 간의 거리가 더 가까움 (변환을 잘 함)
- 비정상 데이터는 Negative인 frame간의 거리, Positive인 motion의 거리가 비슷함 (변환을 잘 못 함)
- label을 기준으로 거리를 계산해도 비슷한 경향을 보임



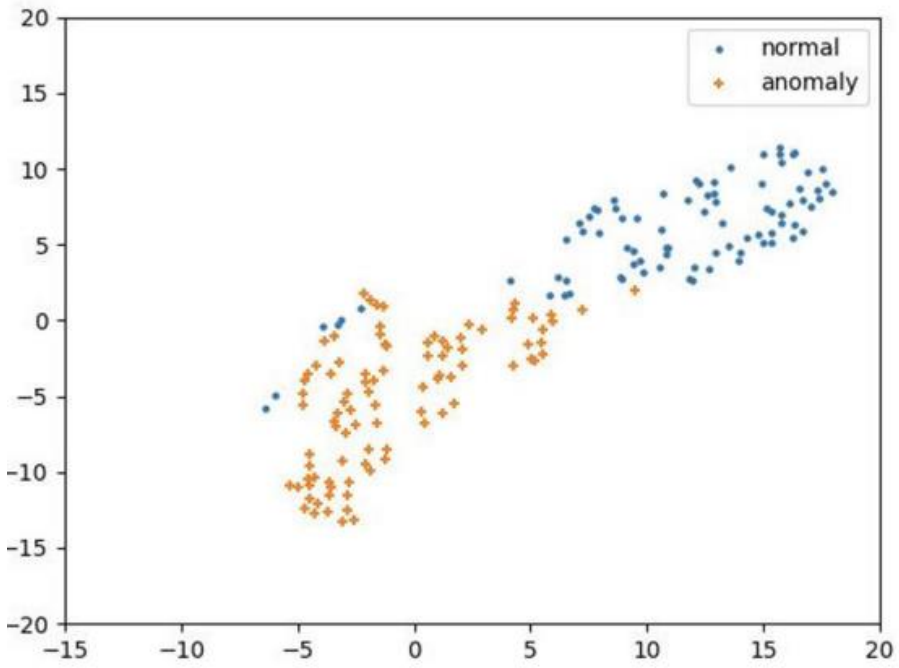
(a) Normal features

(b) Anomaly features

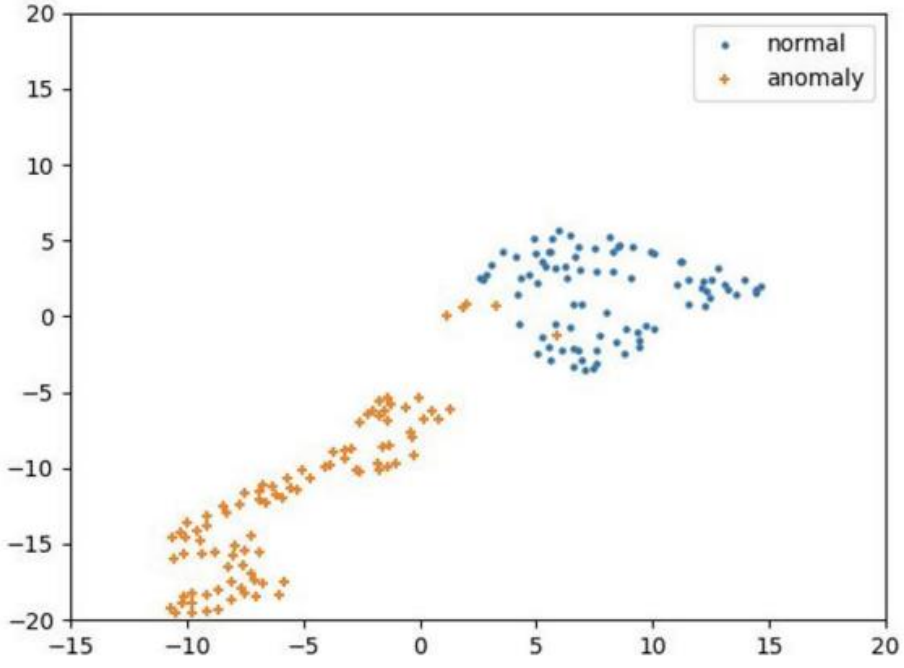
Experiments

Qualitative analysis

- F2LM 생성자를 이용하면 정상과 비정상을 어느정도 구분할 수 있음
- Destroyer까지 이용하면 정상과 비정상을 더욱 효과적으로 구분할 수 있음



(a) The F2LM generator



(b) The Destroyer

Experiments

Ablation study

- 세 개의 인코더와 FTC를 이용하는 방법이 가장 효과적임을 증명함
- 파괴자가 다른 생성자들에 적용해도 효과적임을 증명함

E_f	E_l	E_m	*FTC*	UCSD Ped2	CUHK Avenue	Shanghai Tech.
√				95.4%	85.1%	72.8%
√	√		√	95.8%	85.7%	73.0%
√		√	√	96.8%	86.2%	73.4%
√	√	√	√	**97.5%**	**88.2%**	**74.3%**
√	√	√		95.2%	83.2%	71.0%

Method	Generator	Destroyer	UCSD Ped2	CUHK Avenue	Shanghai Tech.
Frame-Prediction [10]	√		95.4%	84.9%	72.5%
Frame-Prediction (w/ Destroyer)	√	√	96.3%(+0.9%)	87.7%(+2.4%)	72.7%(+0.2%)
TransAnomaly [12]	√		96.2%	85.6%	N/A
TransAnomaly (w/ Destroyer)	√	√	96.8%(+0.6%)	88.2%(+2.6%)	N/A
Ours (w/o Destroyer)	√		97.5%	88.2%	74.3%
Ours	√	√	98.2%(+0.7%)	91.2%(+3.0%)	76.5%(+2.2%)

Experiments

Comparison with SOTA

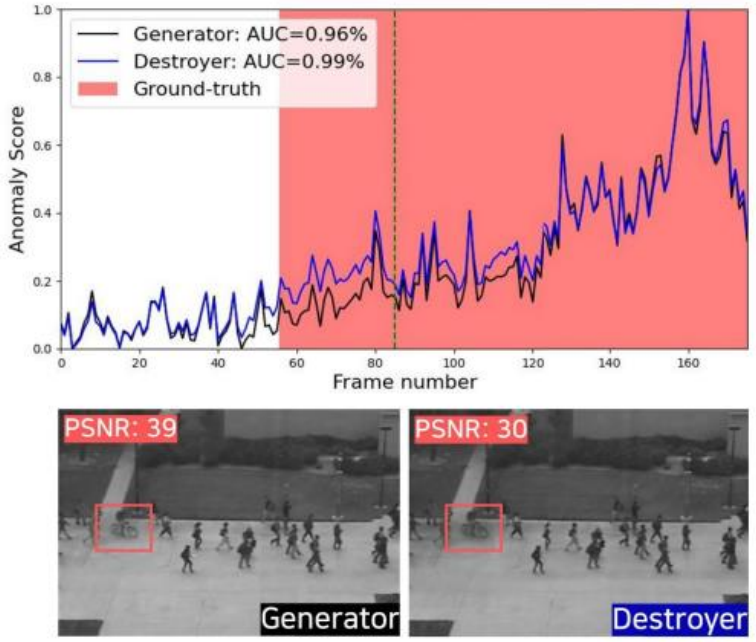
- 최첨단 모델들과 비교하였을 때, Ped2 데이터셋는 Best-second, 나머지 데이터셋은 Best AUC 성능을 달성함

Year	Method	UCSD Ped2		CUHK Avenue		Shanghai Tech.		Publisher
		AUC	EER	AUC	EER	AUC	EER	
2018	FFP [10]	95.4%	11.7%	85.1%	21.4%	72.8%	33.1%	CVPR
	Wang et al. [38]	96.4%	8.9%	85.3%	23.9%	-	-	MM
2019	MemAE [6]	94.1%	-	83.3%	-	71.2%	-	ICCV
	AMC [8]	96.2%	-	86.9%	-	-	-	ICCV
	AnoPCN [11]	96.8%	-	86.2%	-	73.6%	-	MM
	AnomalyNet [39]	94.9%	10.3%	86.1%	22.0%	-	-	IEEE Transactions
2020	DSTN [40]	95.5%	9.4%	87.9%	20.2%	-	-	IEEE Access
	Siamese [41]	94.0%	14,1%	-	-	-	-	WACV
	GMM-FCN [42]	92.2%	12.6%	83.4%	22.7%	-	-	CVIU
	Tang et al. [43]	96.3%	10.0%	85.1%	-	73.0%	-	Pattern Recognition Letters
	FFP+MS_SSIM+FCN [44]	95.9%	11.1%	85.9%	20.4%	73.5%	32.5%	ICCC
	Dual D-b GAN [45]	95.6%	-	84.9%	-	73.7%	32.2%	IEEE Access
	MNAD-P [5]	97.0%	-	88.5%	-	70.5	-	CVPR
2021	TransAnomaly [12]	96.4%	-	87.0%	-	-	-	IEEE Access
	BR-GAN [21]	97.6%	7.6%	88.6%	19.0%	74.5%	31.6%	IEEE Access
	Multi-scale U-Net [46]	95.7%	12.0%	86.9%	20.2%	73.0%	32.3%	IEEE Access
	HMCF [47]	93.7%	18.8%	83.2%	20.0%	-	-	MobileHCI
	HF2-VAD [17]	**99.3%**	-	91.1%	-	76.2%	-	ICCV
2022	Zhong et al. [9]	97.7%	-	88.9%	-	70.7%	-	Pattern Recognition
	DLAN-AC [16]	97.6%	-	89.9%	-	74.7%	-	ECCV
	DR-STN [48]	97.6%	6.9%	90.8%	**11.0%**	-	-	Pattern Recognition Letters
2023	Scene-Aware [49]	-	-	89.6%	21.1%	74.7%	**28.6%**	MM
	MsMp-net [50]	97.6%	6.6%	89.0%	18.1%	-	-	IEEE Access
	Bi-READ [51]	97.7%	7.9%	86.7%	19.5%	-	-	VCIR
	USTN-DSC [52]	98.1%	-	89.9%	-	73.8%	-	CVPR
	SwinAnomaly [53]	98.2%	-	84.8%	-	76.3%	-	IEEE Access
Proposed	**Ours (w/o Destroyer)**	97.5%	7.0%	88.2%	19.1%	74.3%	31.4%	
	Ours	98.2%	**5.9%**	**91.2%**	15.5%	**76.5%**	30.0%	

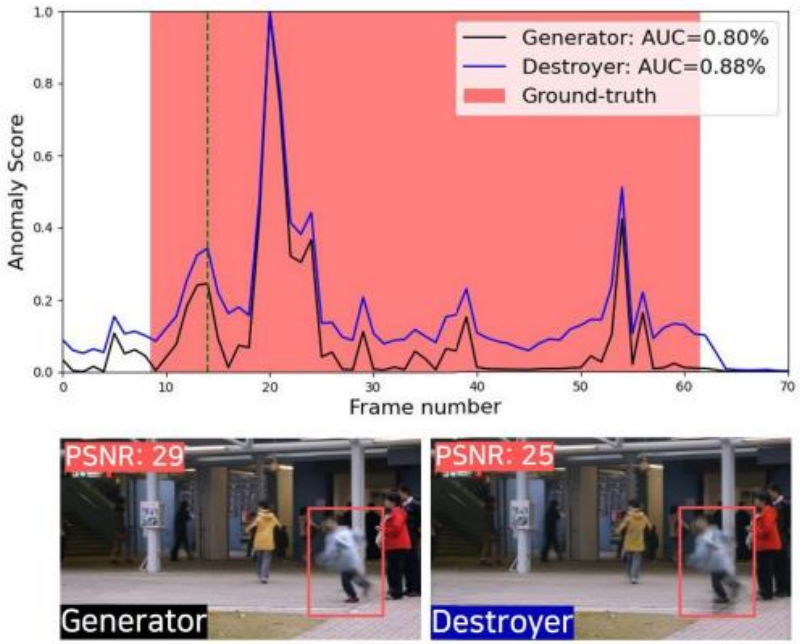
Experiments

Visualization

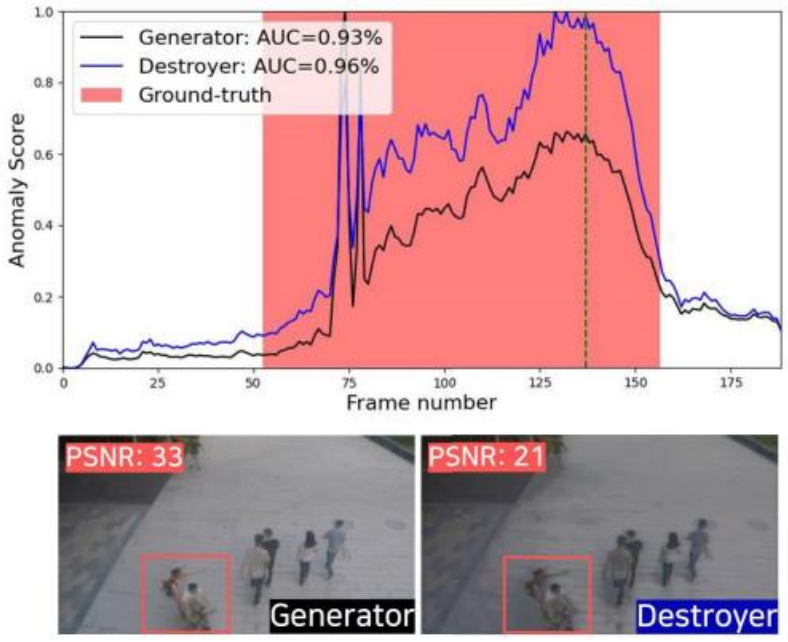
- Destroyer를 이용하면 비정상 영역이 파괴되며, 정상과 비정상 사이의 Anomaly Score 차이가 더 커짐
- Anomaly Score 차이가 더 커지면, 정상과 비정상을 더 잘 분류하므로 성능(AUC)도 증가함



(a) UCSD Ped2, 85^{th} frame of 1^{st} video



(b) CUHK Avenue, 14^{th} frame of 21^{st} video



(c) Shanghai Tech., 137^{th} frame 32^{nd} video

Conclusion

MAMA

- 비정상 프레임을 파괴하여 비디오 이상 탐지를 수행하는 F2LM 생성자와 파괴자를 제안함
- 기존의 미래 프레임 예측 방법과 비교하여 모든 데이터셋에서 성능 향상을 이루었으며, 정성 분석을 통해 두 모델이 효과적임을 증명함
- 비정상 영역을 파괴하는 다양한 방법들이 개발되기를 기대함

Thank you