

구조 집중 네트워크를 이용한 실시간 객체 추적

**Real-Time Visual Tracking
using Structure Attention Network**

**중앙대학교
첨단영상대학원 영상학과
2018126535 김영빈**

Table of Contents

Part 1

1. 논문의 성격

2. 논문의 개요

3. 기존 유사 연구

4. 제안 기술의 차별성

5. 예상 기여도

6. 연구 실적

7. 수강과목 성적

1. 논문의 성격



2. 논문의 개요

- 합성곱 신경망을 이용한 객체 추적에는 추적 객체에 대한 표현력 있는 특징과 적응적인 학습법이 필요하다.
- 실시간 객체 추적을 위해서는 얇은 층을 지니는 합성곱 신경망을 이용하기 때문에 추적 객체에 대한 표현력 있는 특징이 부족하다.
- 이러한 문제를 해결하기 위해 인코더 디코더 프레임워크를 이용해 추적 객체의 표현력 있는 특징을 생성한다.
- 또한 매프레임 변화하는 추적 객체에 대해 강건한 추적을 할 수 있도록 상관 필터와 합성곱 신경망을 함께 이용한다.
- 실시간성을 웃도는 속도와 함께 인코더 디코더를 통해 상관필터가 추적 객체의 표현력 있는 특징을 학습할 수 있게 되어 추적의 강건성과 정확도가 향상되었다.

3. 기존 유사 연구

- Wang, Q., Gao, J., Xing, J., Zhang, M., &Hu, W. (2017). Dcfnet: Discriminant correlation filters network for visual tracking. *arXiv preprint arXiv:1704.04057*.
- Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., &Torr, P. H. (2016, October). Fully-convolutional siamese networks for object tracking. In *European conference on computer vision* (pp. 850-865).
- Valmadre, J., Bertinetto, L., Henriques, J., Vedaldi, A., &Torr, P. H. (2017). End-to-end representation learning for correlation filter based tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2805-2813).
- Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2014). High-speed tracking with kernelized correlation filters. *IEEE transactions on pattern analysis and machine intelligence*, 37(3), 583-596.

4. 제안 기술의 차별성

1

적응적인 추적을 위해 상관필터와 합성곱 신경망을 함께 학습.

2

인코더 디코더 구조를 이용해 객체의 표현력 있는 특징을 생성 및 강조.

3

실시간성을 웃도는 성능과 함께 강건하고 정확한 객체 추적이 가능.

5. 예상 기여도

1

빠르고 정확한 객체 추적 기술로 다양한 감시 시스템에 적용 가능.

2

빠르고 가벼운 추적 알고리즘을 통해 다른 분야의 알고리즘들과 연계 가능.

3

제안하는 Structure Attention Network는 Correlation Filter 기반의 다른 추적 방법에도 적용 가능.

6. 연구 실적

국제 학술 대회

- [1] **Y. Kim**, H. Park, and J. Paik, "Deep Tracking using Convolutional Features and Adaptive Frame Update", IEEE ICCE-Berlin 2018, **(Accepted)**
- [2] V. Maik, **Y. Kim**, H. Park, S. Park, and J. Paik, "Improved Kernelized Correlation Filters in the Scale Space for Online Object Tracking", IEEE/IEIE ICCE-Asia 2018, **(Accepted)**
- [3] **Y. Kim**, J. Shin, H. Park, and J. Paik, "Visual Tracking via Deep Reinforcement Scene Learning", IEEE ICCV 2019, **(Submitted)**
- [4] J. Shin, H. Park, D. Kang, **Y. Kim**, and J. Paik, "CompGAN: Various-Haze Removal Method using a Complementary Generative Adversarial Network", IEEE ICCV 2019, **(Submitted)**

국내 학술 대회

- [1] **김영빈**, 이민정, 김형태, 백준기, "각주파수거리와 불변 모멘트를 이용한 잡음과 크기 변화에 강건한 표지판 검출", 대한전자공학회 학술대회, 2017
- [2] 박세인, **김영빈**, 박하실, 임재승, 백준기, "적응적 색상 특징을 적용한 식별모델 기반 상관 필터", 대한전자공학회 학술대회, 2018
- [3] 신중철, **김영빈**, 박하실, 김태경, 백준기, "경차 분류를 위한 상관적인 다중 해상도 합성곱신경망", 대한전자공학회 학술대회, 2018 (**LG전자 우수논문상 수상**)

6. 연구 실적

국제 학술지

[1] **Y. Kim**, H. Park, and J. Paik, "Robust Kernelized Correlation Filter using Adaptive Feature Weight", IEIE Trans. Smart Processing and Computing, vol. 7, no. 6, pp.433-439, Dec. 2018. **(Scopus, Published)**

국내 학술지

[1] 박상우, 박하실, **김영빈**, 김동근, "국방 경계 감시를 위한 의심객체 검출 및 추적 기술", 정보과학회지, 2018. **(Published)**

7. 수강 과목

첨단영상세미나

A

컴퓨터비전응용

A+

실시간디지털아트제작I

A

감성공학

A+

디지털미디어심리학I

A

최적영상처리론II

A

최적영상처리론I

A+

Table of Contents

Part 2

1. 연구 목표 및 중요성

2. 연구 배경

3. 제안된 방법

4. 실험 결과

5. 결론 및 추후 연구

1. 연구 목표 및 중요성

연구 중요성

- 객체 추적 기술의 특성상 빠르고 강건한 추적 기술이 필수적
- 가벼운 추적 알고리즘을 통해 다양한 감시시스템에서 복합적인 알고리즘들이 효율적으로 운용될 수 있다

연구 목표

- Siamese Network와 Discriminative Correlation Filter를 통한 빠르고 적응적인 추적 가능
- Encoder-Decoder Network를 통해 실시간으로 강건한 구조적인 특징을 집중 학습 가능

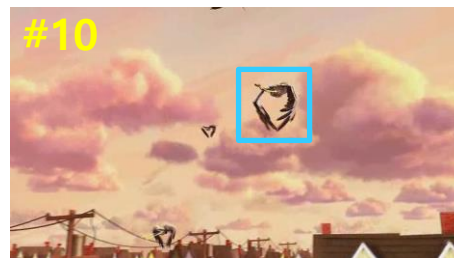
2. 연구 배경

시각적 객체 추적

- 단일 객체 추적, Target free한 특성 (첫 프레임에서만 객체에 대한 정보가 주어짐)



Initial bbox for target



...



- 다양한 외형변화에 강건하기 위해 표현력 있는 풍부한 특징 요구



Background clutters

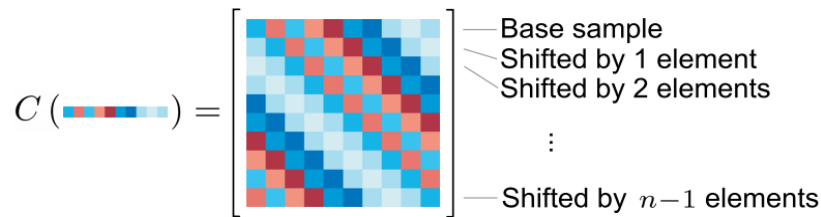


Deformations

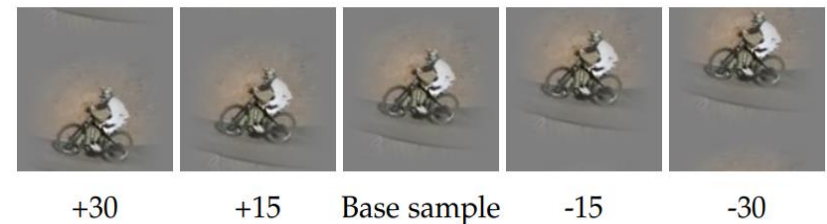
2. 연구 배경

기존 연구 1: Discriminative Correlation Filter

- [1] Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2014). High-speed tracking with kernelized correlation filters. *IEEE transactions on pattern analysis and machine intelligence*, 37(3), 583-596.



순환 행렬 예시



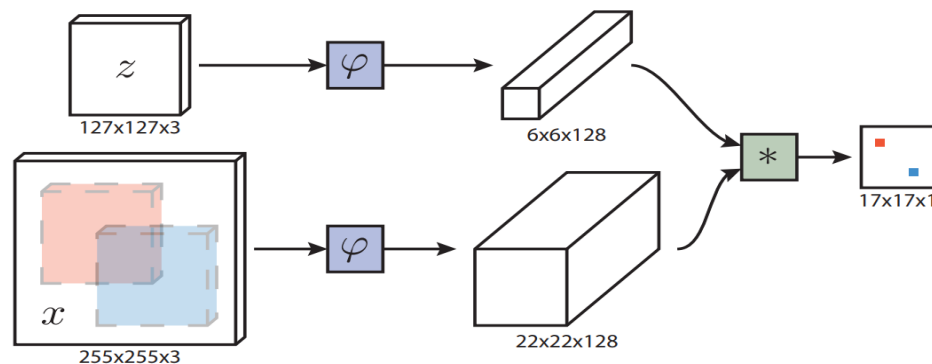
Shift된 학습 샘플

- 특징
 - Ridge regression과 Circulant한 구조를 이용해 학습데이터가 부족한 문제를 해결하고 빠르고 효율적인 계산 가능
 - 실시간으로 학습할 수 있기 때문에 변화하는 객체에 적응적인 추적이 가능
- 문제점
 - Shift된 학습데이터로 인해 Boundary effect 문제를 야기

2. 연구 배경

기존 연구 2: SiameseFC

- [2] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. (2016, October). Fully-convolutional siamese networks for object tracking. In *European conference on computer vision* (pp. 850-865).

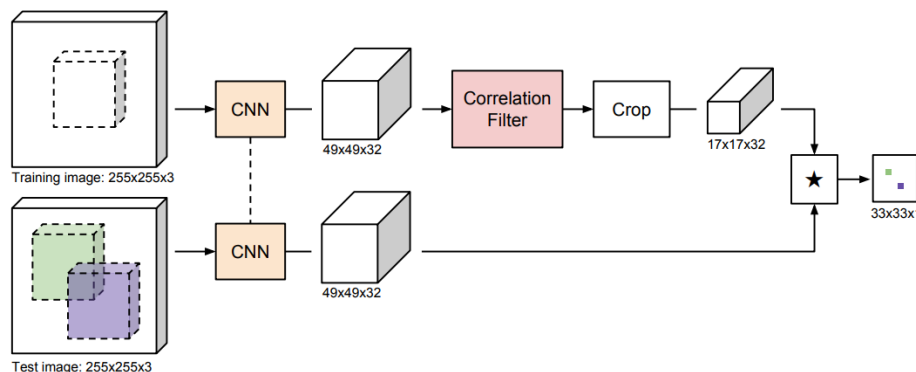


- 특징
 - Siamese network와 Correlation layer를 이용해 객체의 유사도를 비교하고 추적
 - End-to-end 방식으로 실시간 학습이 필요 없으며, 실시간 추적이 가능
- 문제점
 - 실시간 학습을 하지 않기 때문에 다양한 외형 변화에 적응적이지 못함

2. 연구 배경

기존 연구 3: Correlation Filter Network

- [3] Valmadre, J., Bertinetto, L., Henriques, J., Vedaldi, A., & Torr, P. H. (2017). End-to-end representation learning for correlation filter based tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2805-2813).

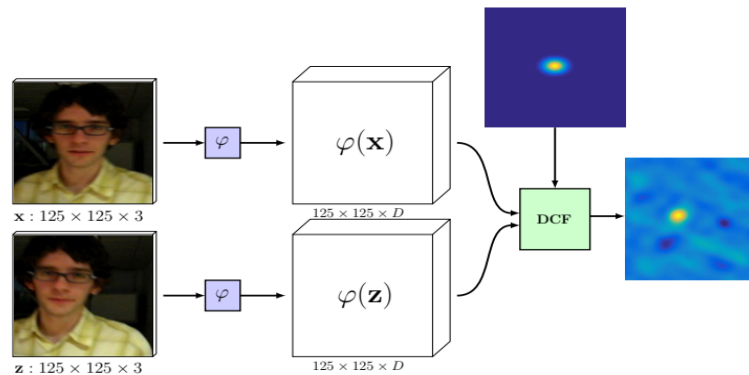


- 특징
 - 기존의 SiamseFC에 Correlation filter layer를 추가하였고, feature crop을 통해 성능을 올림
- 문제점
 - 다음 프레임의 위치를 추정시에 여전히 correlation layer를 이용하기 때문에 객체의 변화에 민감함

2. 연구 배경

기존 연구 3: SiameseDCF

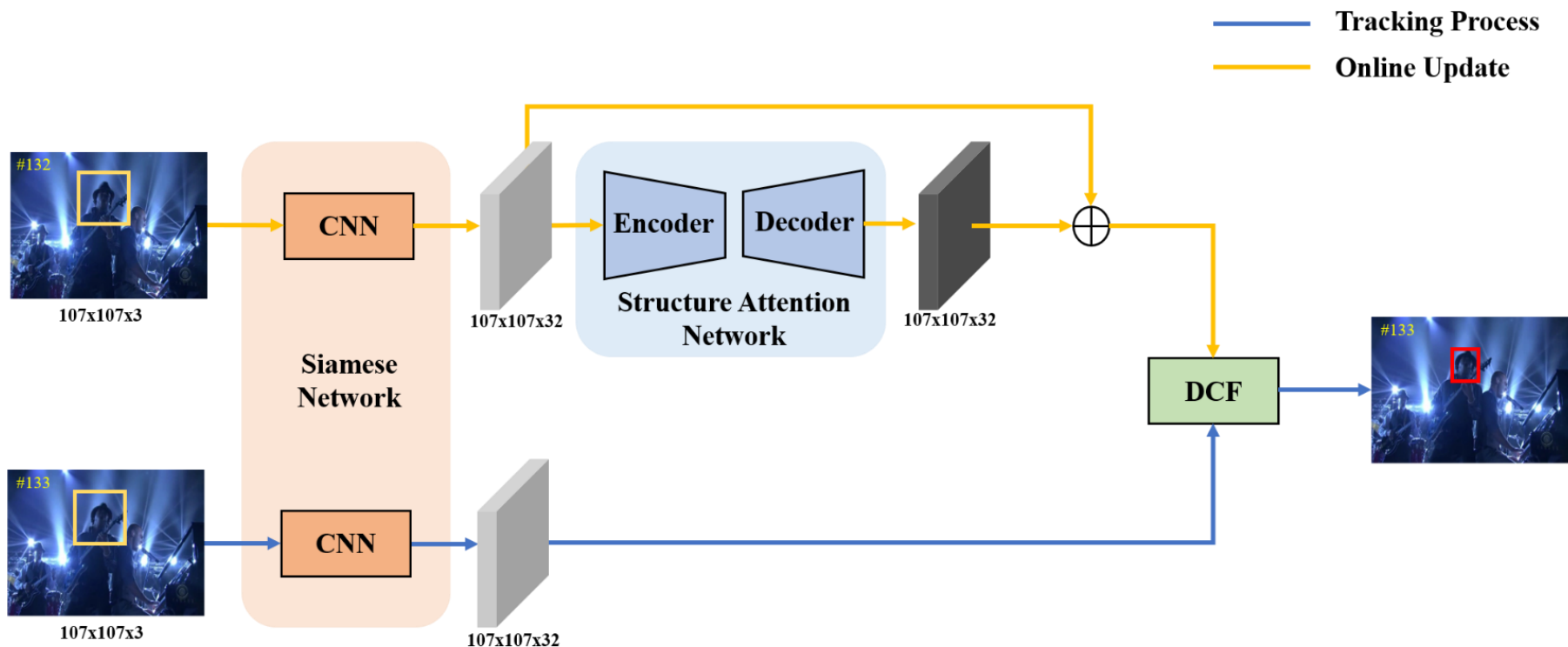
- [4] Wang, Q., Gao, J., Xing, J., Zhang, M., & Hu, W. (2017). Dcfnet: Discriminant correlation filters network for visual tracking. *arXiv preprint arXiv:1704.04057*.



- 특징
 - Discriminative correlation filter를 Siamese network와 함께 사전 학습함으로써 기존의 correlation layer를 통한 추적보다 정확한 위치 추정이 가능
 - 실시간 학습이 가능하기 때문에 객체의 변화에 적응적임
- 문제점
 - 얇은 layer를 이용하기 때문에 correlation filter가 객체의 표현력있는 특징을 학습하는데 한계가 있음

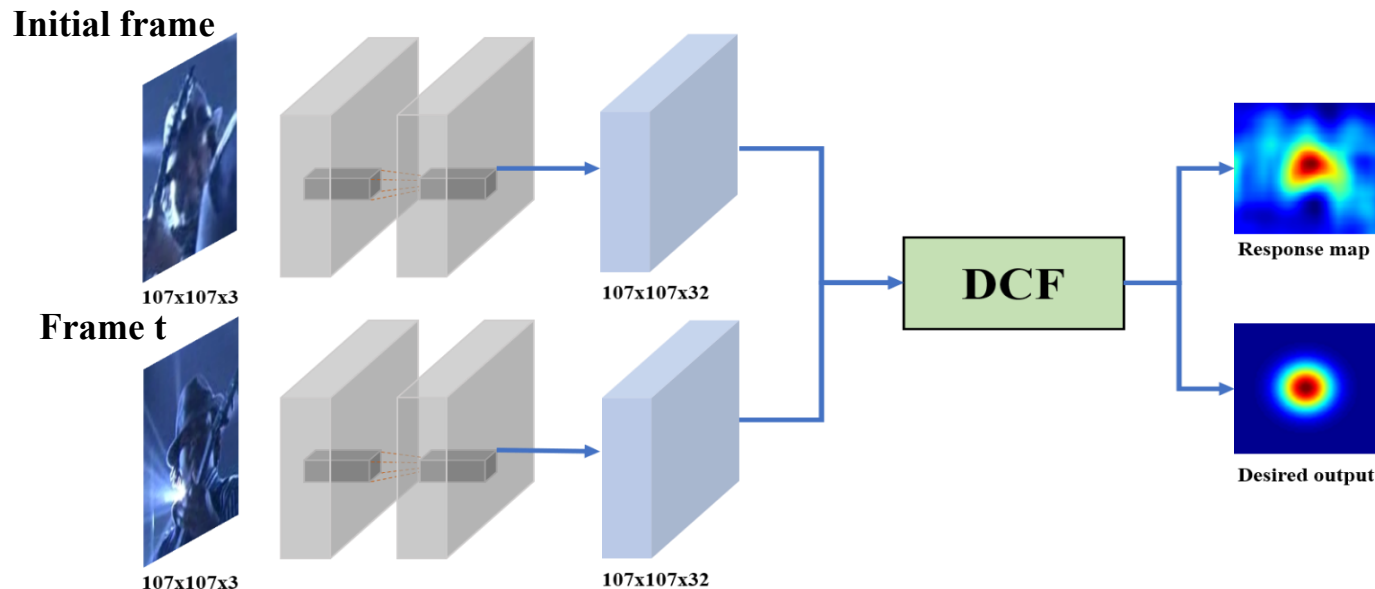
3. 제안된 방법

제안하는 방법의 전체 구조



3. 제안된 방법

Siamese Network



- 첫 프레임에서 주어진 객체의 외형(Template)과 다른 프레임의 객체의 외형(Search)간의 유사도를 Discriminative Correlation Filter를 이용해 비교
- 변형된 추적 객체의 Response map이 Gaussian shape의 형태를 띠도록 네트워크를 학습 (Discriminative Correlation Filter에 특화된 특징을 생성하도록 학습)

3. 제안된 방법

Siamese Network

- Discriminative Correlation Filter 정의

$$\min_w \left\| \sum_i w_i \cdot x_i - y \right\|^2 + \lambda \sum_i \|w_i\|^2 \quad (1) \quad \Rightarrow \quad \hat{w}_i = \frac{\hat{x}_i \odot \hat{y}^*}{\hat{x}_i \odot \hat{x}_i^* + \lambda} \quad (2)$$

- Fast Detection

$$g = F^{-1} \left(\sum_i \hat{w}_i^* \odot \hat{z}_i \right) \quad (3)$$

- Online Training

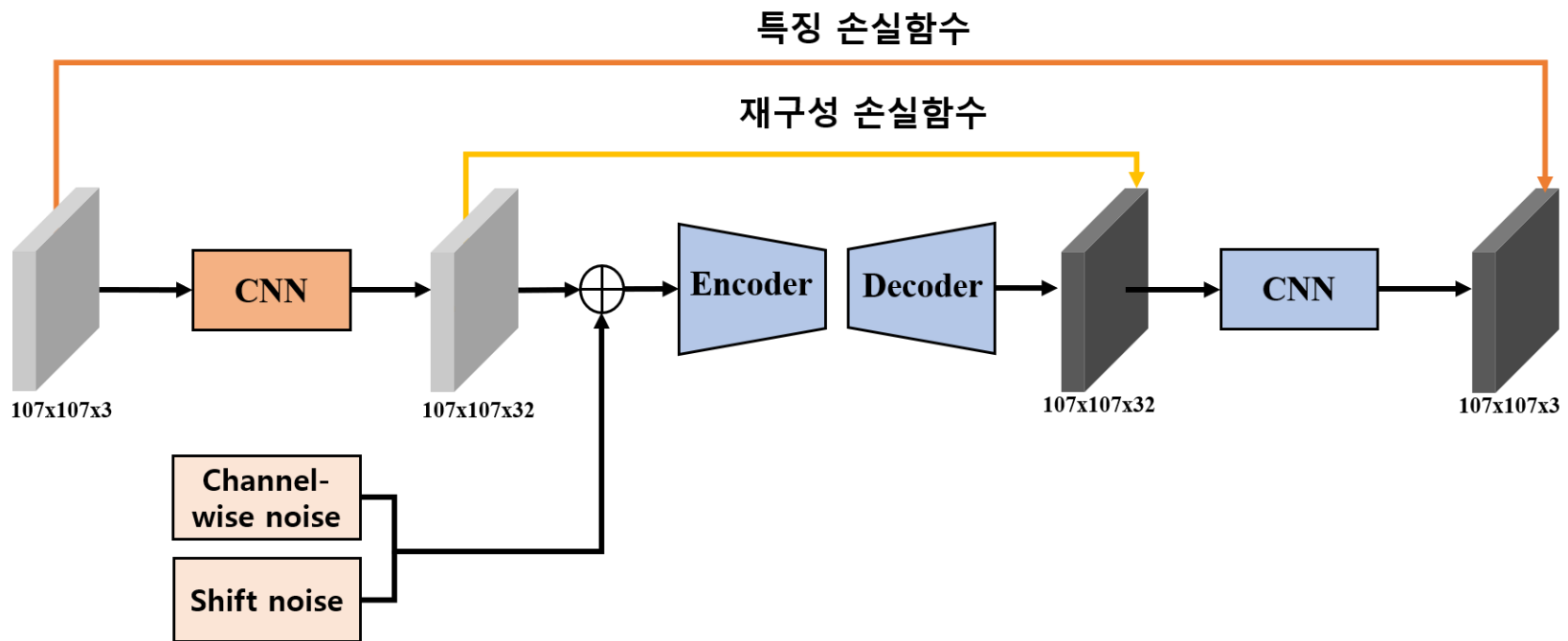
$$\begin{aligned} A_i^t &= (1 - \eta) A_i^{t-1} + \eta \hat{x}_i^t \odot \hat{y}^* \\ B_i^t &= (1 - \eta) B_i^{t-1} + \eta \hat{x}_i^t \odot \hat{x}_i^{t*} \end{aligned} \quad (4) \quad \hat{w}_i^t = \frac{A_i^t}{B_i^t + \lambda} \quad (5)$$

- Siamese Network 손실 함수

$$L = \frac{1}{n} \sum_j \|y - g_j\|^2 \quad (6)$$

3. 제안된 방법

Structure Attention Network

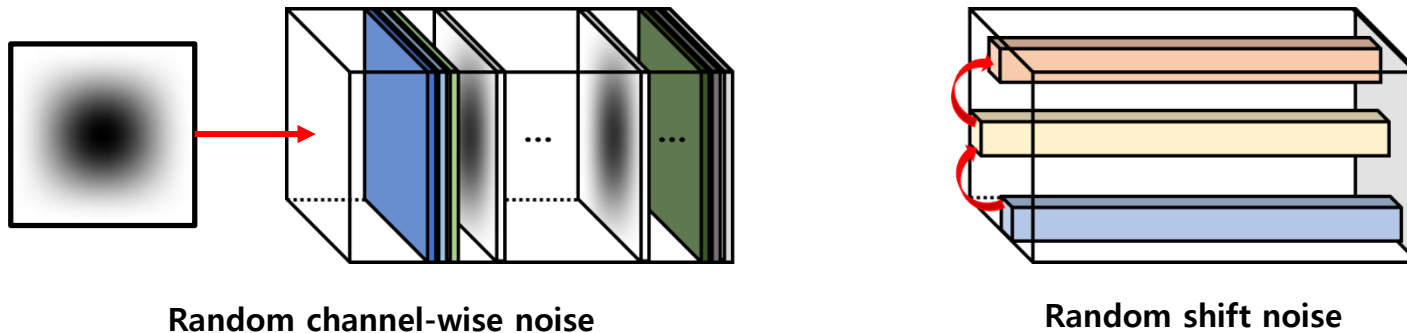


- Discriminative Correlation Filter가 객체의 표현력 있고 구조적인 특징들에 집중해서 학습 할 수 있도록, 제안하는 손실함수와 두개의 노이즈를 통해 Encoder-Decoder Network를 학습

3. 제안된 방법

Structure Attention Network

- 매 프레임 학습을 하는 Discriminative Correlation Filter가 추적 객체의 강건한 구조적 특징을 학습할 수 있도록, 두 가지의 노이즈를 통해 네트워크를 학습한다.



- 강건한 특징 생성을 위해 특징 손실함수와 재구성 손실함수를 통해 네트워크를 학습한다.

$$L = \underbrace{\frac{1}{m} \sum_i \|X_i - \text{VAE}(\hat{X}_i)\|^2}_{\text{재구성 손실함수}} + \underbrace{\frac{1}{m} \sum_l \|H_l - D(\text{VAE}(\hat{X}_j))\|^2}_{\text{특징 손실함수}} - \underbrace{\frac{1}{2} \sum_j (1 + \log(\sigma_j^2) - \mu_j^2 + \sigma_j^2)}_{\text{KLD}}$$

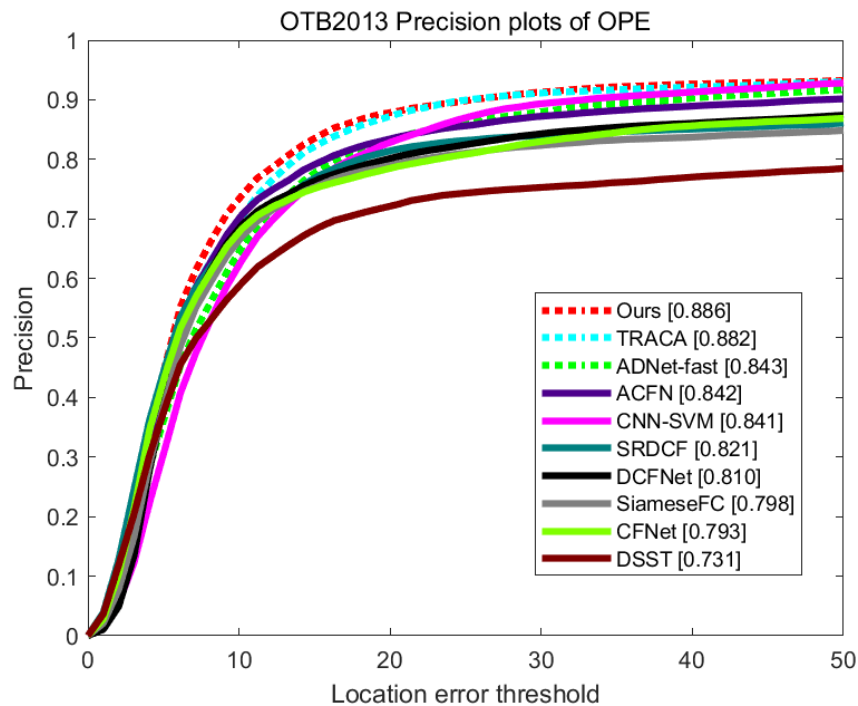
4. 실험 결과

동영상 시연

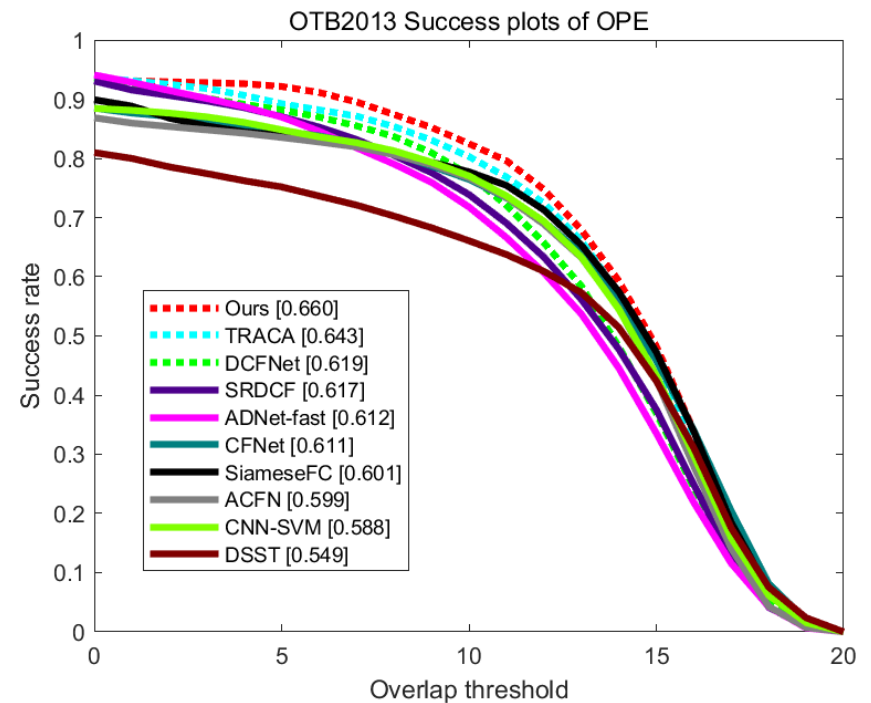


4. 실험 결과

OTB2013 Dataset 실험 결과



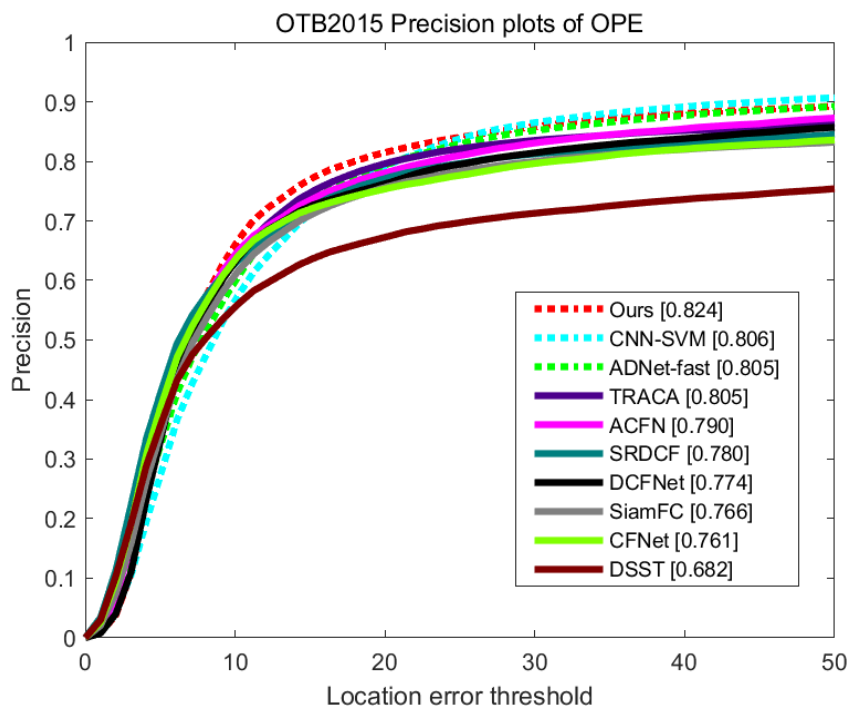
Distance Precision



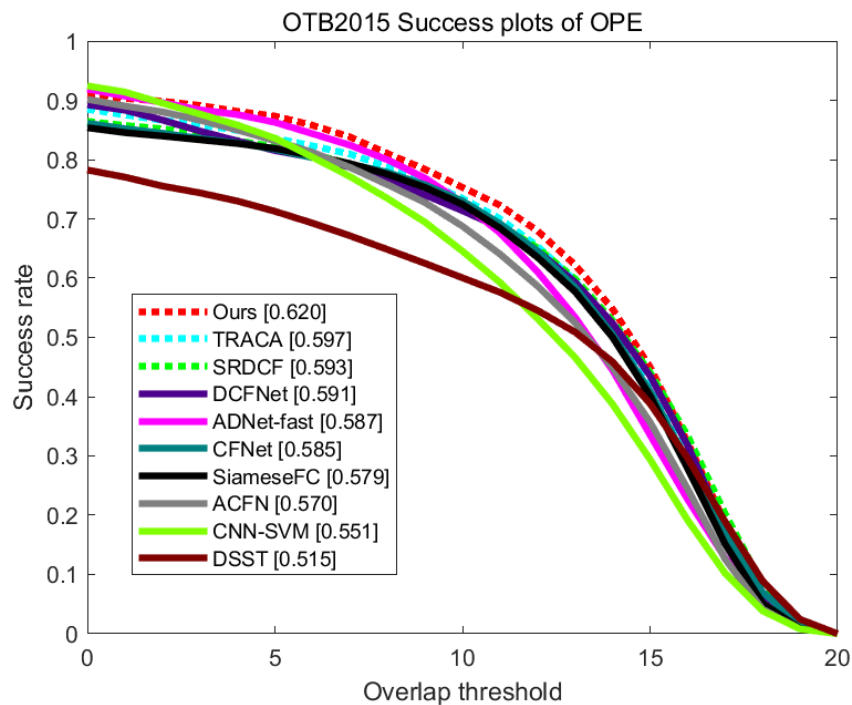
Overlap Precision

4. 실험 결과

OTB2015 Dataset 실험 결과



Distance Precision



Overlap Precision

4. 실험 결과

OTB Dataset 실험 결과

	Ours		CFNet		SRDCF		SiamFC		DCFNet		ADNet		DSST		TRACA		CNN-SVM		ACFN	
	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC	Dp	AUC
OTB 2013	0.886	0.660	0.793	0.611	0.821	0.617	0.798	0.601	0.810	0.619	0.843	0.612	0.731	0.549	0.882	0.643	0.841	0.588	0.842	0.599
OTB 2015	0.824	0.620	0.761	0.585	0.780	0.593	0.766	0.579	0.774	0.591	0.805	0.587	0.682	0.515	0.805	0.597	0.806	0.551	0.790	0.570

OTB2013, 2015 dataset 실험 결과

	Ours		CFNet		SRDCF		SiamFC		DCFNet		ADNet		DSST		TRACA		CNN-SVM		ACFN	
	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015	OTB 2013	OTB 2015
OP 0.5	0.852	0.784	0.794	0.757	0.787	0.745	0.789	0.753	0.793	0.740	0.809	0.768	0.683	0.625	0.830	0.762	0.759	0.694	0.776	0.727
OP 0.75	0.593	0.548	0.545	0.505	0.560	0.528	0.546	0.500	0.574	0.524	0.481	0.445	0.514	0.458	0.570	0.516	0.445	0.387	0.477	0.449

강건도 측정 결과

4. 실험 결과

세부 평가

- 총 439개의 비디오 데이터셋, 11개의 세부 항목에 대한 평가

IV(38): Illumination Variation dataset 38개

FM(39): Fast Motion dataset 39개

SV(64): Scale Variation dataset 64개

IPR(51): In-plane Rotation dataset 51개

OCC(49): Occlusion dataset 49개

OPR(63): Out-of-Plane Rotation dataset 63개

DEF(44): Deformation dataset 44개

OV(14): Out-of-View dataset 14개

MB(29): Motion Blur dataset 29개

BC(31): Background Clutters 31개

LR(9): Low Resolution 9개

4. 실험 결과

세부 평가

	Ours	CFNet	SRDCF	SiamFC	DCFNet	ADNet	DSST	TRACA	CNN-SVM	ACFN
	DP	DP	DP	DP	DP	DP	DP	DP	DP	DP
IV	0.767	0.680	0.765	0.713	0.711	0.834	0.703	0.813	0.761	0.761
SV	0.769	0.719	0.737	0.729	0.744	0.766	0.640	0.758	0.761	0.751
OCC	0.783	0.681	0.710	0.701	0.762	0.693	0.585	0.750	0.704	0.724
DEF	0.708	0.672	0.707	0.670	0.714	0.796	0.531	0.744	0.766	0.746
MB	0.766	0.602	0.731	0.677	0.721	0.719	0.543	0.720	0.720	0.686
FM	0.792	0.682	0.730	0.713	0.729	0.692	0.556	0.710	0.712	0.715
IPR	0.815	0.749	0.721	0.724	0.757	0.767	0.680	0.788	0.791	0.756
OPR	0.784	0.723	0.719	0.739	0.744	0.794	0.639	0.802	0.776	0.754
OV	0.752	0.528	0.586	0.661	0.732	0.597	0.467	0.685	0.626	0.656
BC	0.782	0.728	0.772	0.685	0.733	0.794	0.697	0.795	0.766	0.755
LR	0.843	0.805	0.757	0.899	0.798	0.913	0.673	0.850	0.914	0.810

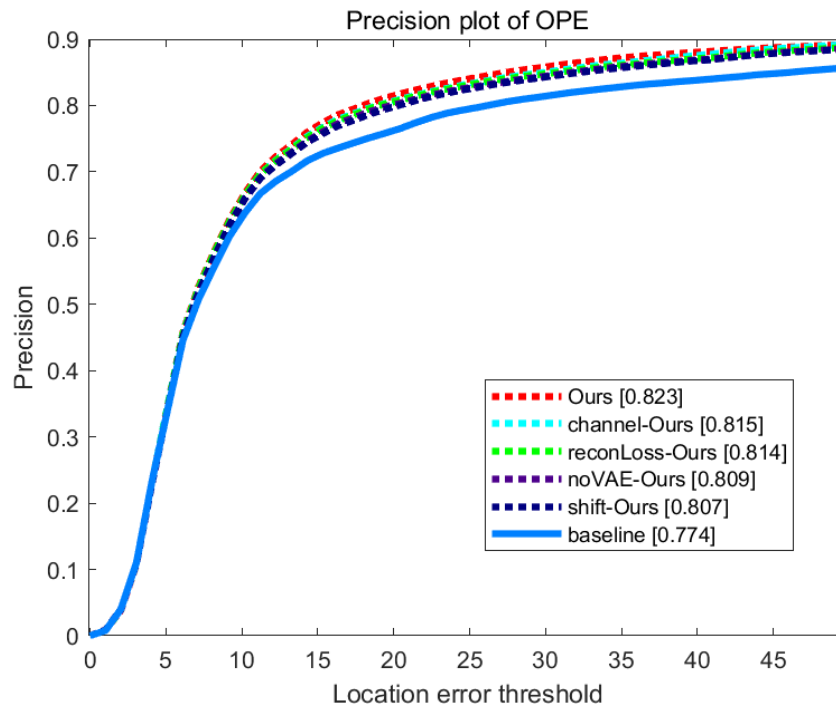
4. 실험 결과

세부 평가

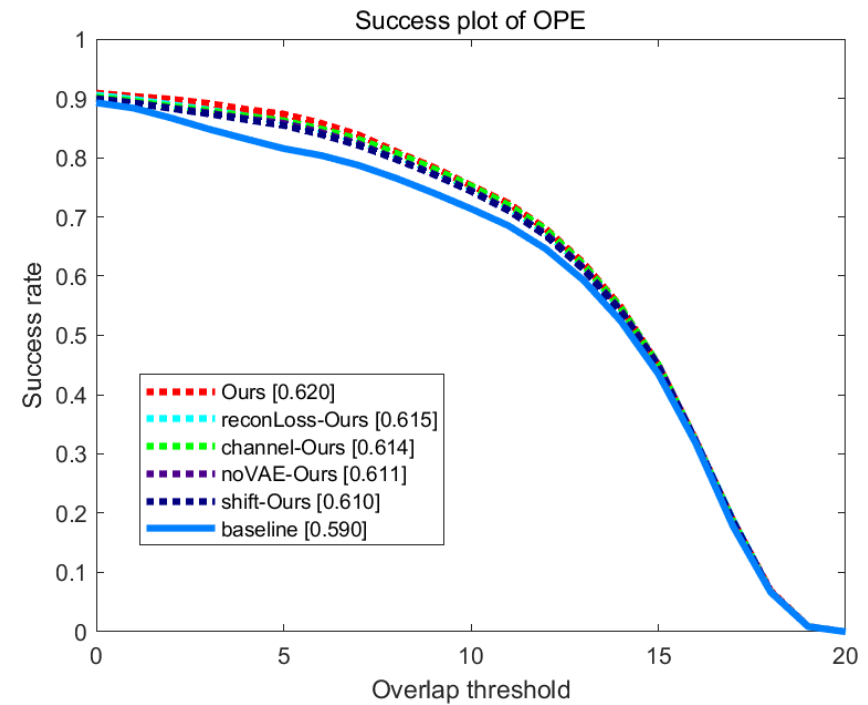
	Ours	CFNet	SRDCF	SiamFC	DCFNet	ADNet	DSST	TRACA	CNN-SVM	ACFN
	OP	OP	OP	OP	OP	OP	OP	OP	OP	OP
IV	0.594	0.541	0.599	0.560	0.562	0.612	0.550	0.608	0.529	0.558
SV	0.585	0.550	0.561	0.552	0.568	0.563	0.475	0.554	0.490	0.547
OCC	0.599	0.533	0.549	0.536	0.585	0.518	0.454	0.561	0.507	0.531
DEF	0.536	0.500	0.533	0.498	0.537	0.555	0.420	0.550	0.538	0.527
MB	0.631	0.503	0.577	0.539	0.596	0.565	0.460	0.573	0.565	0.550
FM	0.632	0.546	0.581	0.556	0.593	0.550	0.460	0.561	0.534	0.551
IPR	0.604	0.564	0.534	0.550	0.568	0.559	0.500	0.571	0.540	0.536
OPR	0.593	0.541	0.542	0.552	0.571	0.571	0.472	0.586	0.542	0.538
OV	0.579	0.423	0.460	0.507	0.565	0.479	0.385	0.547	0.488	0.493
BC	0.597	0.565	0.583	0.520	0.563	0.588	0.535	0.591	0.551	0.539
LR	0.543	0.588	0.513	0.621	0.523	0.573	0.381	0.501	0.378	0.514

4. 실험 결과

내부 평가



Distance Precision



Overlap Precision

4. 실험 결과

내부 평가

	Ours	Channel -Ours	ReconLoss -Ours	noVAE -Ours	Shift -Ours	Baseline
Dp	0.823	0.815	0.814	0.809	0.807	0.774
AUC	0.620	0.614	0.615	0.611	0.610	0.590
Fps	89	89	89	106	89	101

내부 평가 실험 결과

5. 결론 및 추후 연구

결론

- Siamese Network와 Discriminative Correlation Filter를 함께 학습하여 빠르고 정확한 추적 가능
- 인코더 디코더 구조를 이용해 Correlation Filter가 실시간으로 객체의 강건한 구조적인 특징을 학습 가능
- 실시간성이 요구되는 다양한 분야에 적용 가능

추후 연구

- 더욱 풍부하고 강건한 특징을 얻기 위해 깊은 네트워크가 필요
- 다중 객체 추적을 위한 병렬적인 알고리즘 처리 연구 예정



경청해 주셔서 감사합니다.