

특집논문 (Special Paper)

방송공학회논문지 제28권 제4호, 2023년 7월 (JBE Vol.28, No.4, July 2023)

<https://doi.org/10.5909/JBE.2023.28.4.374>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 마스크 기반 특징 맵 교환을 통한 객체 검출기 지식 증류

백 건 우<sup>a)</sup>, 조 현<sup>a)</sup>, 황 원 준<sup>a)‡</sup>

### Distilling Dense Object Detector via Masked Feature Exchange

Geonwoo Baek<sup>a)</sup>, Hyeon Cho<sup>a)</sup>, and Wonjun Hwang<sup>a)‡</sup>

#### 요 약

객체 탐지를 위한 지식 증류는 다양한 응용 분야에 적용할 수 있기 때문에 큰 주목을 받고 있다. 그러나 대부분의 연구는 문맥 정보를 고려하지 않고 **intermediate feature**를 모방하는 것을 목표로 한다. 본 논문에서는 교사의 **feature**를 효과적으로 전달하기 위한 마스크 기반 지식 증류 프레임워크를 제안한다. **Feature**에 전경-배경 사이 문맥 정보를 부여하기 위해 교사의 관점에서 전경을 추출하는 신뢰 기반 마스크를 사용한다. 이 마스크는 전경-배경 사이 **gradient gap**을 고려하기 위해 교사와 학생 특징에 각각 적용되어 교환을 수행한다. 마지막으로, 양방향 **KL divergence** 손실을 사용하여 교사의 전경 특징과 학생의 배경 특징 간의 문맥적 격차를 줄인다. 다양한 실험을 통해 우리의 방법이 **state-of-the-art** 방법들과 경쟁력 있는 결과를 달성하는 것을 보여준다.

#### Abstract

Knowledge distillation for object detection has attracted attention to apply in various applications. However, most studies aim to mimic intermediate feature without considering contextual information. In this paper, we propose context-based knowledge distillation framework to transfer the feature of the teacher. In order to contain contextual information in the feature, we employ confidence-based mask that extracts foreground feature from the teacher's perspective. The mask is applied to the teacher and the student features, exchanging them to account for the foreground-background gradient gap. Finally, we use bidirectional KL divergence loss to bridge the contextual gap between the teacher's foreground features and the student's background features. Through various experiments, we demonstrate that our method achieves comparable result with state-of-the-art methods.

Keyword : Object Detection, Knowledge Distillation

---

a) 아주대학교 소프트웨어학과(Department of Software at Ajou University)

‡ Corresponding Author : 황원준(Wonjun Hwang)

E-mail: [wjhwang@ajou.ac.kr](mailto:wjhwang@ajou.ac.kr)

Tel: +82-31-219-2632

ORCID: <https://orcid.org/0000-0001-8895-0411>

※ This work was supported by BK21 FOUR program of NRF (NRF5199991014091), Korea NRF grant (NRF-2022R1A2C1091402), and ITRC support program (IITP-2023-2018-0-01431).

· Manuscript May 31, 2023; Revised July 18, 2023; Accepted July 18, 2023.

## 1. 서론

최근 딥러닝 분야는 상당한 발전이 이루어져 분류, 객체 검출, 세그먼테이션 등 다양한 컴퓨터 비전 task에 대한 활발한 연구가 이루어지고 있다. 객체 검출은 많은 분야에서 중요한 과제로 남아 있으며, 지속적으로 연구가 이루어지고 있다. 객체 검출기는 지역 제안 네트워크가 있는 2단계 검출기와 지역 제안 네트워크가 없는 1단계 검출기(Dense Object Detector) 두 가지 유형으로 나눌 수 있다<sup>[1]</sup>. 2단계 물체 검출기의 높은 정확도에도 불구하고, 1단계 물체 검출 방법은 더 빠르고 효율적이기 때문에 실제 물체 검출에 일반적으로 사용된다. 1단계 물체 검출기는 병렬적으로 위치 파악 및 분류를 동시에 수행하므로 추론 속도가 빠른 장점이 있다. 객체 검출기가 갖는 대용량 특징 인코더의 메모리 및 계산 비용을 줄이기 위해 심층 네트워크(예: 가지 치기, 지식 증류, 신경 아키텍처 검색)의 효율성에 대한 다양한 연구가 진행되고 있다. 특히 모델 복잡도와 정확도 사이의 trade-off를 최소화하기 위해 지식 증류(Knowledge Distillation, KD) 연구가 활발히 진행되고 있다.

객체 검출 KD에서는 의미 정보와 위치 정보를 모두 고려해야 하기 때문에 특징 맵에서 어떤 영역을 어떤 가중치로 증류할지 선택하는 것이 중요하다. 최근의 연구에서는 전

경, 배경 영역의 supervision loss에 대한 gradient를 측정했을 때, 통계적으로 전경의 gradient가 배경보다 큰 것을 식별하였다. 이에 따라 이러한 gradient gap 문제를 해결하기 위해 ground-truth bounding box를 기반으로 영역을 분리한 후, 서로 다른 가중치를 부여하여 증류하는 방법인 [5,16,17]이 큰 성과를 거두었다.”로 변경하였다. 그러나 그림 1 (a)와 같은 region-to-region 증류 방법은 여전히 학생 네트워크가 교사 특징 맵의 전경과 배경 사이의 문맥 정보를 충분히 활용하지 못한다는 문제점이 있다. 최근 [18]에서는 인스턴스와 배경 사이의 문맥 정보가 탐지 성능에 결정적인 영향을 미친다고 말하며 신중한 고려가 필요하다고 강조했다.

본 논문에서는 교사 모델로부터 학생에게 효과적으로 문맥 정보를 전달하기 위한 특징 전달 기반 지식 증류 방법을 제안한다. 제안하는 방법은 (1) 마스크 기반 특징 교환(masked feature exchange)과 (2) 특징 문맥 브리징 두 부분으로 구성된다. 마스크 기반 특징 교환은 특징 맵을 전경과 배경으로 나눈 후, 교사 모델의 전경 특징을 학생 모델의 전경 특징과 교환하여 두 영역 간 gradient gap을 고려할 수 있도록 한다. 특징 문맥 브리징은 컨볼루션 모듈을 사용하여 상대 모델의 전경 특징과 자신의 배경 특징을 혼합하여 gradient gap을 줄이면서 효과적으로 특징 지식을 증류한다.

## II. 관련 연구

지식 증류는 반응 기반<sup>[2,6,8]</sup>, 특징 기반<sup>[5,12,15]</sup>, 관계 기반<sup>[9,10,14]</sup> 세 종류의 방법으로 나뉜다. 객체 검출에서, 특징 기반 방법은 지역화와 분류에 관한 정보를 모두 고려해야 하기 때문에 주로 연구되어왔다. 특히 영상 내부의 인스턴스 위치에 따라 어떤 영역을 증류해야 하는지에 대한 연구가 많이 이루어졌다. 방법 [8]은 지역 제안 네트워크에서 제안된 지역의 특징 맵을 증류함으로써 교사의 지식을 모방하려고 한다. 방법 [15]는 물체가 위치한 지역 뿐만 아니라 그 근처 영역에도 의미 있는 정보가 있다고 주장했다. 방법 [4]는 교사와 학생 예측 사이의 차가 큰 일반적인 영역을 증류하는 연구를 시도했다. 방법 [20]은 교사의 어두운 지식(dark knowledge)을 전달하기 위해 교사의 분류 점수가

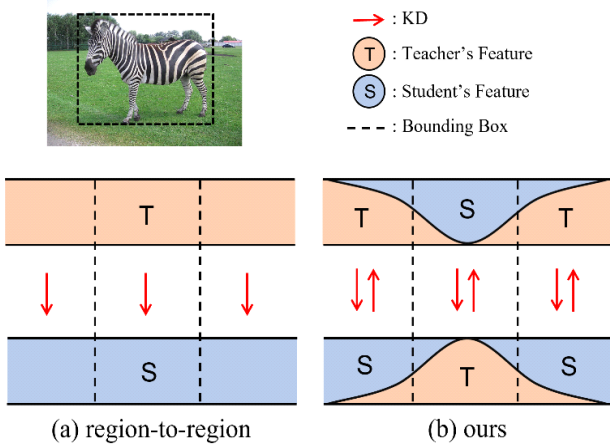


그림 1. (a) region-to-region 방법과 (b) 제안하는 방법의 개념 그림  
 Fig. 1. Conceptual illustration of (a) region to region method and (b) ours

높은 지역의 특징 맵을 증류했다. 특히, [5] 방법은 실제 물체의 위치인 **ground-truth bounding box**의 내부와 외부 사이의 **gradient gap**이 존재함을 실험적으로 증명하고, 이 문제를 해결하기 위해 분리된 특징 증류를 제안한다. 후속 연구인 [16, 17] 방법들은 **ground-truth bounding box**의 내부를 전경 및 외부 배경으로 분할하는 것 외에도 관심 및 품질 점수로 전경 영역에 증류 가중치를 할당하여 증류를 수행한다.

그러나 이러한 영역 간 특징 추출 방법은 객체 검출 성능에 큰 영향을 미치는 객체와 배경 간의 상황 정보를 충분히 활용하지 못한다. 이러한 문제점을 보완하기 위해 일부 연구에서는 배경을 포함하여 인스턴스와 픽셀 간의 관계를 증류하여 성능 향상을 시도했다<sup>[4,17]</sup>. 그러나 이러한 접근법은 전경과 노이즈가 많은 배경을 동등하게 취급하기 때문에 전경-배경 사이의 **gradient gap** 문제를 여전히 가지고 있다. 따라서 교사의 문맥을 고려한 특징을 편향 없이 전달하기 위해서는 지역적 차이를 고려한 증류 방법이 필요하다.

### III. 제안하는 방법

이번 장에서는 교사-학생 특징 맵의 전경 교환을 통해 복잡한 교사의 문맥 정보를 효과적으로 전달하기 위한 새로

운 특징 증류 프레임워크를 설명한다. 그림 2는 해당 방법의 전반적인 구조를 보여준다. 먼저, 신뢰도 점수에 따라 생성된 전경 마스크를 이용해 마스크 기반 특징 교환을 수행한다. 다음으로, 특징 문맥 브리징을 통해 교사의 문맥 정보를 효과적으로 증류한다.

#### 1. 마스크 기반 특징 교환

이 방법의 목표는 문맥, **gradient gap**이 존재하는 전경, 배경 영역의 정보를 효과적으로 전이하는 것이다. 기존 **ground-truth bounding box** 기반으로 생성한 전경은 객체가 실제로 존재하지 않는 픽셀을 가지고, 객체가 존재하는 픽셀이더라도 일부는 예측에 큰 영향을 미치지 않는다. 따라서, 정답 레이블 기반 영역 분리는 픽셀 단위의 정보를 처리할 수 없기 때문에 차선택이 될 수 있다. 따라서 교사 예측 값에 대한 **gradient**를 반영할 수 있는 픽셀 단위의 전경을 제안한다. 분류와 회귀에 대한 중요도를 모두 고려하기 위해 분류 점수와 회귀 점수의 곱으로 표현된 신뢰 점수를 사용한다. 이 신뢰 기반 전경 마스크는 다음과 같은 과정을 따라 생성된다.

$$M_{i,j} = (P_{i,j}^t)^\alpha \cdot (B_{i,j}^t)^{(1-\alpha)}, \quad (1)$$

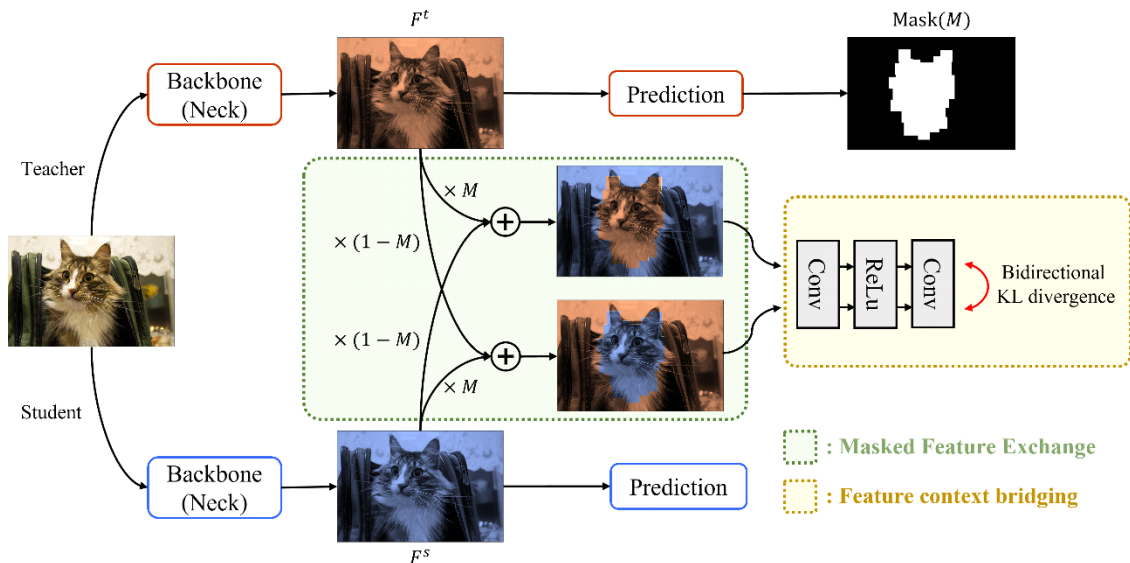


그림 2. 제안하는 방법의 전체적인 구조  
Fig. 2. Overall framework of our method

여기서  $P^t$ 는 교사 예측의 분류 점수 맵,  $B^t$ 는 ground-truth bounding box와 교사 예측 간의 Intersection Over Union(IOU) 점수이다.  $i, j$ 는 각각 수평 및 수직 방향에 해당하는 인덱스를 나타내고,  $\alpha$ 는 결합 확률의 균형 매개변수이다. 그런 다음, 마스크  $M$ 과 Feature Pyramid Network(FPN)의 출력 특징 맵 간의 Hadamard 곱을 통해 교사와 학생의 전경 특징 맵을 얻는다. 그 후, 다음과 같이 마스크 기반 영역의 특징을 교환한다:

$$\begin{aligned} F^{ts} &= F^t \cdot M + F^s \cdot (1 - M), \\ F^{st} &= F^t \cdot (1 - M) + F^s \cdot M. \end{aligned} \quad (2)$$

$F^t$ 와  $F^s$ 는 각각 교사와 학생의 FPN 출력 특징 맵을 나타낸다.  $F^{ts}$ 는 교사의 전경과 학생 배경 특징의 원소 별 합이며,  $F^{st}$ 는 교사의 배경과 학생 전경 특징의 원소 별 합이다. 두 수정된 특징 맵은 그림 1(b)와 같이 표현된다. 편의를 위해 수식에서 FPN 레벨은 생략하였다.

## 2. 특징 문맥 브리징

$F^t$ 와  $F^s$ 는 서로 다른 깊이를 갖는 네트워크로부터 비롯되었기 때문에 서로 다른 특징 임베딩 공간에 속한다. 따라서 특징의 일부분이 교환된  $F^{ts}$ 와  $F^{st}$ 의 픽셀 값들의 불연속성을 완화하기 위해 간단한 컨볼루션 모듈을 추가한다. 이 모듈은 서로 다른 백본으로부터 비롯된 특징들이 혼합될 수 있도록 할 뿐만 아니라 교사와 학생 모두의 지식을 학습함으로써 용량 차이를 최소화하는 역할을 한다<sup>[7]</sup>. 여기서 컨볼루션 모듈  $W$ 는 다음과 같다.

$$W(F) = \hat{F} = \text{Conv}_{3 \times 3}(\text{ReLU}(\text{Conv}_{3 \times 3}(F))). \quad (3)$$

여기서  $F$ 는 입력 특징 맵이다.  $W$ 를 통과한  $F^{ts}$ 는 학생의 배경이 교사의 전경을 힌트로 하여 교사의 배경을 학습할 수 있도록 하고,  $W$ 를 통과한  $F^{st}$ 는 학생의 전경이 교사

의 배경 정보를 힌트로 하여 교사의 전경을 예측할 수 있게 한다. 컨볼루션 모듈을 통과한 두 특징 맵에 대해 공간 별 및 채널 별 양방향 KL divergence 손실을 사용한다. 제안된 손실의 수식은 아래와 같다.

$$\sigma_c(F) = \frac{\exp\left(\frac{F_{c,i}}{\tau_1}\right)}{\sum_{i=1}^{HW} \exp\left(\frac{F_{c,i}}{\tau_1}\right)}, \quad (4)$$

$$\sigma_{hw}(F) = \frac{\exp\left(\frac{F_{hw,j}}{\tau_2}\right)}{\sum_{i=1}^C \exp\left(\frac{F_{hw,j}}{\tau_2}\right)}, \quad (5)$$

$$L_c(F_1, F_2) = KL(\sigma_c(F_1) \parallel \sigma_c(F_2)), \quad (6)$$

$$L_{hw}(F_1, F_2) = KL(\sigma_{hw}(F_1) \parallel \sigma_{hw}(F_2)). \quad (7)$$

이때,  $\tau_1$ 과  $\tau_2$ 는 각각 채널 방향, 공간 방향의 KL divergence 손실을 위한 temperature이다.  $C, H, W$ 는 특징 맵의 채널, 높이, 너비에 해당하는 인덱스를 나타낸다. 간단한 컨볼루션 모듈은 작은 수용 필드를 가지므로 교환된 특징들이 충분히 혼합되지 않는다. 따라서 모든 영역의 혼합을 위해 소프트맥스 기반의 KL divergence를 사용한다. 게다가  $\hat{F}^{ts}$ 와  $\hat{F}^{st}$  사이의 양방향 KL divergence 손실에 의해 문맥 정보의 일관성을 규제한다. 이를 통해 컨볼루션 모듈은 교사의 문맥 정보를 학생에게 효과적으로 넘겨주도록 학습되며, 전체 손실은 수식 (8)과 같이 표현된다.

여기서  $\alpha, \beta$ 는 전경, 배경 영역에 대한 손실의 균형 변수이다. 위 과정을 통해 학생의 전경은 교사의 배경 정보를 바탕으로 교사의 전경을 학습하게 되고, 학생의 배경은 교사의 전경 정보를 바탕으로 교사의 배경을 학습하게 된다. 또한, 특징 문맥 브리징 모듈을 통해 교사의 전경-학생의 배경 사이의 관계와 교사의 배경-학생의 전경 사이의 관계

$$L_{KD}(F^{ts}, F^{st}) = \alpha(L_c(\hat{F}^{ts}, \hat{F}^{st}) + L_{hw}(\hat{F}^{ts}, \hat{F}^{st})) + \beta(L_c(\hat{F}^{st}, \hat{F}^{ts}) + L_{hw}(\hat{F}^{st}, \hat{F}^{ts})). \quad (8)$$

를 일치시킴으로써 교사의 문맥 정보를 간접적으로 전달할 수 있게 한다.

#### IV. 실험

##### 1. 실험 세팅

제안하는 방법을 검증하기 위해 일반적인 물체 탐지 작업에서 자주 사용되는 COCO2017 데이터셋에서 정량적 실험을 수행한다. 학습에서는 train2017 set을 사용하고, 검증을 위해 val2017 set을 사용한다. 모든 실험은 위 세팅을 사용하여 수행되며 GPU 당 3개의 샘플에 대해 2개의 NVIDIA RTX A5000에서 학습된다. 그리고 우리의 방법은 Pytorch 기반의 MMDetection 툴킷으로 구현하였으며, 툴킷의 기본 데이터 증강 기법을 사용하였다<sup>[3]</sup>. 교사와 학생의 실험 설정으로는, 3x 및 2x 스케줄에 대한 멀티 스케일 학습으로 훈련된 두 교사에 대한 실험 프로토콜을 채택한다. 학생의 학습 전략은 1x 스케줄로, 12 epoch에 대해 단

일 싱글 스케일 학습을 수행한다. 모든 실험에서 SGD 옵티마이저를 사용하였으며, 초기 학습률(Learning Rate)은  $3.75E-3$ 이다. 우리는 또한 8epoch와 11epoch에 0.1 비율의 step decay 학습 정책을 사용한다.

##### 2. 결과

마스킹 전략 : 본 항에서는 마스킹 방법을 통한 문맥 정보 전달의 효과를 검증하기 위해 여러 마스크에 대한 실험 결과를 설명한다. 표 1에서 No mask 방법은 마스킹 없이 교사와 학생의 특징 맵을 입력하여 양방향 학습이 이루어지는 것을 말한다. BG random 방법은 ground-truth bounding box를 제외한 배경 영역에서 0.5의 비율로 랜덤 픽셀 마스킹을 수행하고, All random 방법은 피쳐 맵의 모든 영역에서 0.5의 비율로 마스킹이 수행된다. GT-Bbox 방법은 ground-truth bounding box를 기반으로 생성된 마스크를 이용하며, Confidence 방법은 제안하는 신뢰 점수 기반 마스킹 방법을 의미한다. No mask 대비 BG random과 All

표 1. 여러 마스크 방법에 따른 성능 평가  
Table 1. Performance evaluation of various mask methods

Mask strategy	mAP	AP50	AP75	APS	APM	APL
No mask	41.7	60.9	44.4	24.7	45.1	54.4
BG random	41.9	61.1	44.8	25.4	45.1	54.6
All random	42.0	61.1	45.1	25.0	45.6	54.5
GT-Bbox	42.3	61.2	45.4	25.7	46.2	55.2
Confidence	42.3	61.3	45.5	26.1	46.1	54.7

표 2. Anchor-free 검출기에서 최첨단 KD 방법과의 비교  
Table 2. Comparisons with state-of-the-art KD methods in anchor-free detector

Method	mAP	AP50	AP75	APS	APM	APL
2xTeacher [13]	42.5	61.8	45.8	25.3	46.7	54.7
Student	38.8	58.4	41.4	23.2	42.3	50.4
FGD [17]	41.6	61.1	44.3	25.3	45.2	54.0
PGD [16]	41.9	61.2	44.5	24.1	45.3	55.0
PKD [1]	42.0	61.2	45.0	24.3	45.6	55.2
Ours	42.0	61.1	45.1	24.6	45.9	54.7
3xTeacher [13]	43.3	62.4	46.7	26.1	47.4	55.4
FGD [17]	41.8	61.1	44.6	25.0	45.3	54.4
PGD [16]	42.3	61.6	45.3	25.1	45.8	55.9
PKD [1]	42.2	61.4	45.3	25.2	46.0	55.2
Ours	42.3	61.3	45.5	26.1	46.1	54.7

표 3. Anchor-based 검출기에서 최첨단 KD 방법과의 비교  
Table 3. Comparisons with state-of-the-art KD methods in anchor-based detector

Method	mAP	AP50	AP75	APS	APM	APL
2x Teacher [13]	44.9	63.1	49.2	28.3	49.6	56.8
Student	39.6	57.7	43.0	24.0	43.1	51.1
FGD [17]	43.2	61.3	46.9	25.7	47.2	55.6
PGD [16]	43.4	61.6	47.3	26.7	47.3	56.3
PKD [1]	43.6	61.7	47.4	26.2	47.6	56.4
Ours	44.0	62.1	48.0	26.1	48.1	57.2

random 방법이 각각 0.2% mAP와 0.3% mAP 성능 향상을 보이는데, 이를 통해 마스크 기반 특징 교환 프레임워크가 문맥 정보가 효과적으로 전달하여 성능 향상을 이룬 것을 확인할 수 있다. 앞서 설명한 랜덤 기반 방법에 비해 전경 기반 방법(GT-Bbox, Confidence)은 0.3% mAP 향상을 보였고, 전경과 배경 사이의 gap을 고려하여 문맥 정보를 효과적으로 전달하는 것을 간접적으로 확인할 수 있다. [18]에서 문맥 정보를 잘 고려한다면 작은 객체의 성능 향상을 이룰 수 있다고 주장하는 바와 같이, GT-Bbox 방법보다 0.4% APS 높은 성능을 갖는 신뢰 기반 마스크링 방법이 더 효과적임을 알 수 있다.

다른 지식 증류 방법들과 비교 : 제안한 방법이 1단계 물체 검출기에 미치는 영향을 확인하기 위해, 우리는 COCO val2017 set에서 anchor-free 방법인 FCOS와 anchor-based 방법 ATSS를 여러 가지 최첨단 지식 증류 방법과 비교한다<sup>[13,19]</sup>. 표 2에서 보이는 것과 같이, 제안하는 방법은 3x 스케줄로 학습된 교사를 사용한 실험에서 APS와 APM가 다른 방법들을 능가했고, 2x 설정에서는 최신 방법들과 비슷한 결과를 얻었다. 특히, APS는 교사의 성능과 동일한 수치를 보여주며 최신 방법들과 대등한 결과를 보여준다. 이는 분류 시 객체의 중심에 대한 위치 정보를 포함하는 중심 기반 방식 FCOS를 사용하기 때문에 정확도가 높아졌음을 시사한다. 표 3의 결과에 따르면 anchor-based 검출기는 학생보다 4.4% mAP 우수한 성능을 보였고, 기존 방법들의 성능을 모두 능가한다. 또한, AP75는 객체의 크기에 상관없이 우수한 성능을 갖는다. 이것은 높은 IOU 임계 값을 갖을 때 모델의 경계 상자 예측 성능이 높다는 의미로, 보다 정교한 경계 상자를 생성한다는 것을 의미한다.

따라서 위의 결과들은 문맥 정보가 효과적으로 전달됨에

따라 정교한 예측 값을 갖는 것을 보여준다. 주목할 만한 점으로, 제안하는 방법은 학습할 때 약간의 파라미터가 추가되지만, 실제 추론 시에는 추가된 파라미터들이 사용되지 않으므로 기본 검출기와 동일한 추론 시간과 더 나은 정확도를 갖는다.

## V. 결론

본 논문에서는 마스크 기반 특징 맵 교환을 통한 문맥 중심 지식 증류 프레임워크를 제안한다. 문맥 정보와 함께 특징 맵 정보를 증류하기 위해, 신뢰 기반 마스크를 사용하여 전경 특징을 추출한다. 그리고 전경-배경 사이의 gradient gap을 반영하기 위해 교사와 학생 각각의 특징 맵에 마스크를 적용한 후 한 모델의 전경 특징과 다른 모델의 전경 특징을 교환한다. 마지막으로, 특징 문맥 브리징 부분에서는 교환된 특징 맵의 이질적인 차이를 완화하기 위해 컨볼루션 모듈을 통과한 후, 양방향 KL divergence 손실을 통해 문맥 정보를 효과적으로 전달한다. 수행된 실험은 최첨단 방법과 비교할 수 있는 결과를 보여주며, 특히 AP75 수치는 다른 방법들보다 더 나은 성능을 보여줌으로써 보다 정교한 예측이 가능하다는 것을 증명한다.

## 참고 문헌 (References)

- [1] Weihan Cao, Yifan Zhang, Jianfei Gao, Anda Cheng, Ke Cheng, and Jian Cheng. Pkd: General distillation framework for object detectors via pearson correlation coefficient. arXiv preprint arXiv:2207.02039, 2022.
- [2] Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan

- Chandraker. Learning efficient object detection models with knowledge distillation. *Advances in neural information processing systems*, 30, 2017.
- [3] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155, 2019.
- [4] Xing Dai, Zeren Jiang, Zhao Wu, Yiping Bao, Zhicheng Wang, Si Liu, and Erjin Zhou. General instance distillation for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7842 - 7851, 2021.  
doi: <https://doi.org/10.1109/cvpr46437.2021.00775>
- [5] Jianyuan Guo, Kai Han, Yunhe Wang, Han Wu, Xinghao Chen, Chunjing Xu, and Chang Xu. Distilling object detectors via decoupled features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2154 - 2164, 2021.  
doi: <https://doi.org/10.1109/cvpr46437.2021.00219>
- [6] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
- [7] Lujun Li and Zhe Jin. Shadow knowledge distillation: Bridging offline and online knowledge transfer. *Advances in Neural Information Processing Systems*, 35:635 - 649, 2022.
- [8] Quanquan Li, Shengying Jin, and Junjie Yan. Mimicking very efficient network for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6356 - 6364, 2017.  
doi: <https://doi.org/10.1109/cvpr.2017.776>
- [9] Yufan Liu, Jiajiong Cao, Bing Li, Chunfeng Yuan, Weiming Hu, Yangxi Li, and Yunqiang Duan. Knowledge distillation via instance relationship graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7096 - 7104, 2019.  
doi: <https://doi.org/10.1109/cvpr.2019.00726>
- [10] Wonpyo Park, Dongju Kim, Yan Lu, and Minsu Cho. Relational knowledge distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3967 - 3976, 2019.  
doi: <https://doi.org/10.1109/cvpr.2019.00409>
- [11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.  
doi: <https://doi.org/10.1109/tpami.2016.2577031>
- [12] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. Fitnets: Hints for thin deep nets. arXiv preprint arXiv:1412.6550, 2014.
- [13] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9627 - 9636, 2019.  
doi: <https://doi.org/10.1109/iccv.2019.00972>
- [14] Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1365 - 1374, 2019.  
doi: <https://doi.org/10.1109/iccv.2019.00145>
- [15] Tao Wang, Li Yuan, Xiaopeng Zhang, and Jiashi Feng. Distilling object detectors with fine-grained feature imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4933 - 4942, 2019.  
doi: <https://doi.org/10.1109/cvpr.2019.00507>
- [16] Chenhongyi Yang, Mateusz Ochal, Amos Storkey, and Elliot J Crowley. Prediction-guided distillation for dense object detection. In *Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23 - 27, 2022, Proceedings, Part IX*, pages 123 - 138. Springer, 2022.  
doi: [https://doi.org/10.1007/978-3-031-20077-9\\_8](https://doi.org/10.1007/978-3-031-20077-9_8)
- [17] Zhendong Yang, Zhe Li, Xiaohu Jiang, Yuan Gong, Zehuan Yuan, Danpei Zhao, and Chun Yuan. Focal and global knowledge distillation for detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4643 - 4652, 2022.  
doi: <https://doi.org/10.1109/cvpr52688.2022.00460>
- [18] Mengmi Zhang, Claire Tseng, and Gabriel Kreiman. Putting visual object recognition in context. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12985 - 12994, 2020.  
doi: <https://doi.org/10.1109/cvpr42600.2020.01300>
- [19] Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z Li. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9759 - 9768, 2020.  
doi: <https://doi.org/10.1109/cvpr42600.2020.00978>
- [20] Du Zhixing, Rui Zhang, Ming Chang, Shaoli Liu, Tianshi Chen, Yunji Chen, et al. Distilling object detectors with feature richness. *Advances in Neural Information Processing Systems*, 34:5213 - 5224, 2021.

---

저 자 소 개



**백 건 우**

- 2022년 : 아주대학교 전자공학과 학사
- 2022년 ~ : 아주대학교 인공지능학과 석사과정
- ORCID : <https://orcid.org/0009-0002-8025-968X>
- 주관심분야 : 딥러닝, 컴퓨터비전, 패턴인식



**조 현**

- 2018년 : 아주대학교 소프트웨어학과 학사
- 2018년 ~ : 아주대학교 석박통합과정
- ORCID : <https://orcid.org/0000-0001-7017-803X>
- 주관심분야 : 딥러닝, 컴퓨터비전, 패턴인식



**황 원 준**

- 1999년 : 고려대학교 전자공학과 학사
- 2001년 : 고려대학교 전자공학과 석사
- 2016년 : KAIST 전기 및 전자공학부 박사
- 2001년 ~ 2016년 : 삼성중합기술원 전문연구원
- 2016년 ~ 2020년 : 아주대학교 SW학과 조교수
- 2020년 ~ 현재 : 아주대학교 SW학과 부교수
- ORCID : <https://orcid.org/0000-0001-8895-0411>
- 주관심분야 : 컴퓨터비전, 패턴인식, 딥러닝