

# Yolov4 알고리즘을 사용하여 식품 성분에 대한 실시간 다중 클래스 감지를 위한 딥러닝 모델

시아플리 쇼빈<sup>1</sup>, 트리스티안티 유스니타사리<sup>2</sup>, 테디 오스와리<sup>3</sup>, 레니 디아 쿠수마와티<sup>4</sup>,  
누라시아<sup>5</sup>  
{[syaflishovin@gmail.com](mailto:syaflishovin@gmail.com) <sup>1</sup>[tyusnita@staff.gunadarma.ac.id](mailto:tyusnita@staff.gunadarma.ac.id) <sup>2</sup>, [toswari@staff.gunadarma.ac.id](mailto:toswari@staff.gunadarma.ac.id)  
<sup>3</sup>[reni\\_dk@staff.gunadarma.ac.id](mailto:reni_dk@staff.gunadarma.ac.id) <sup>4</sup>, [nurasiah@staff.gunadarma.ac.id](mailto:nurasiah@staff.gunadarma.ac.id) }<sup>5</sup>

인도네시아 자카르타 구나다르마 대학교 컴퓨터 과학부<sup>1,2,5</sup> 인도네시아 자카르타  
구나다르마 대학교 경제학부<sup>3,4</sup>

**Abstract.** 쌓여있을 수 있는 식재료의 위치와 식재료의 모양, 색상, 질감 등의 유사성은 식재료에 대한 최적의 검출 작업을 수행할 수 있는 딥러닝 모델을 구축하는 데 있어 난제가 됩니다. 따라서 성능이 우수한 YOLOv4 알고리즘을 사용하여 식재료를 학습하고 검출하게 됩니다. 이번 연구에서는 수집된 데이터셋의 다양성과 양을 늘리기 위해 YOLOv4 알고리즘의 데이터 증강 기법도 적용했습니다. 본 연구에서는 학습의 효율성을 높이기 위해 전이학습 기법을 활용하여 YOLOv4 사전 학습 모델의 지식을 도입했습니다. 이 연구에서 사용된 접근 방식은 상당히 우수한 성능으로 식재료에 대한 실시간 다중 클래스 검출을 위한 딥러닝 모델을 성공적으로 생성합니다. 이 모델의 성능은 mAP@0.50 값 84.90%, 평균 IoU 72.77%로 나타났습니다.

**키워드:** 딥 러닝; 전이 학습; YOLOv4; 식품 성분 감지.

## 1 소개

딥러닝은 얼굴 인식, 사물 감지 등 다양한 목적으로 자율주행 자동차에 널리 적용되고 있습니다. 특히 객체 검출을 위한 다양한 딥러닝 모델이 여러 분야에서 널리 개발되고 있습니다. 물체 검출을 목적으로 하는 딥러닝이 많이 개발되고 있는 분야 중 하나는 식품 분야에서의 적용입니다. 딥러닝이 식품 분야에 적용되고 있는 이유는 기존 기법에 비해 장점이 많기 때문입니다[7]. 딥러닝 모델을 적용하여 객체 검출을 수행하고 식재료의 이름을 예측하는 것은 다양한 목적을 가지고 있으며, 그 중 하나는 식품 품질 검사를 수행하는 것입니다 [6].

음식 감지 작업을 수행하기 위해 딥러닝을 구현하는 것은 특히 어려운 과제입니다. 쌓여있는 식재료의 위치, 식재료의 모양, 색상, 질감의 유사성 등 여러 가지 요인에 따라 음식 물체를 감지하는 데 어려움을 겪을 수 있으며, 식재료 감지를 적용하는 데 있어서는 식재료의 유사성이 문제가 될 수 있습니다. 식재료 이미지에서 특징을 최적으로 추출하여 식재료를 감지할 수 있는 딥러닝 모델을 생성하려면 효율적인 알고리즘이 필요합니다.

이 연구에서는 전이 학습 기법을 활용하여 우수한 성능으로 식품 성분을 감지할 수 있는

딥러닝 모델을 생성합니다. 이 연구에서는 사전 학습된 모델로서 YOLOv4 모델의 아키텍처를 사용합니다. 실시간 다중 클래스 객체 검출을 수행하고자 하는 본 연구의 목적을 위해서는 실시간 다중 클래스 객체 검출에서 YOLOv4의 성능이 필수적입니다.

식품 성분에 대한 탐지. YOLOv4의 또 다른 장점은 1080 Ti 또는 2080 Ti와 같은 단일 GPU를 사용하여 모델을 훈련할 수 있다는 것입니다. 이러한 장점으로 인해 YOLOv4는 다른 최첨단 검출기 모델보다 속도와 정확도 성능이 더 뛰어난 모델을 생성할 수 있습니다[1].

## 2 자료 및 방법

### 2.1 이미지 데이터 세트

이 연구에 사용된 식재료 클래스는 과일, 채소, 동물성 단백질을 포함한 11개 클래스입니다. 11가지 클래스는 바나나, 양배추, 당근, 오이, 감자, 새우, 토마토, 달걀, 레몬, 브로콜리, 오렌지입니다. 수집된 데이터는 확장자가 .jpg인 이미지 파일 형태입니다. 데이터 세트는 OpenImages, Google Image, iStock과 같은 여러 오픈 소스 사이트에서 가져옵니다. 검색된 데이터는 다시 정렬되어 사용된 데이터가 연구 목적에 부합하는지 확인합니다. 사용된 데이터 세트의 총 개수는 2750개이며, 각 클래스에는 250개의 이미지가 있습니다. 그런 다음 파이썬 기반 프로그램을 사용하여 데이터셋을 훈련 데이터셋과 검증 데이터셋으로 나눕니다. 학습용 데이터셋은 머신러닝 모델 학습에 사용되는 데이터셋으로, 본 연구에서 전체 데이터셋의 90%를 차지합니다. 나머지 10%의 데이터셋은 과적합을 방지하기 위한 검증 데이터셋에 사용됩니다.

### 2.2 데이터 집합에 레이블 지정 및 주석 달기

다양한 소스에서 수집된 데이터 세트는 사용된 데이터가 이 연구의 목적에 부합하는지 확인하기 위해 정렬됩니다. 정렬이 완료된 데이터는 해당 객체에 바운딩 박스 및 주석 레이블을 지정합니다. 바운딩 박스 및 라벨링 주석 생성은 파이썬 기반의 오픈 소스 프로그램인 OpenLabeling을 사용하여 수행합니다. 그런 다음 주석 결과는 사진 데이터 세트의 파일 이름에 해당하는 파일 이름을 사용하는 확장자 .txt의 파일 형태로 저장됩니다. 이 도구의 주석 결과는 YOLOv4 형식을 따르므로 나중에 데이터 세트를 YOLOv4로 학습할 수 있습니다.



그림 1. 식품 재료는 등급 이름에 따라 라벨링되었습니다.

### 2.3 데이터 증강

이 연구에 사용된 증강 데이터는 광도 왜곡 및 기하학적 왜곡과 같은 YOLOv4 알고리즘에서 채택한 몇 가지 기법으로 구성됩니다. 이 연구에 사용된 사전 학습된 모델에는 이미지 데이터에서 물체의 존재를 시뮬레이션하기 위한 몇 가지 다른 데이터 증강 기법도 채택되었습니다. 적용된 데이터 증강 기법 중 일부는 컷믹스, 모자이크 데이터 증강, 자기 적대적 훈련 (SAT)입니다.

이 데이터의 레이블은 데이터의 혼합에 따라 조정됩니다 [11]. 반면 모자이크 데이터 증강은 4개의 이미지를 하나의 단위로 결합하는 기법입니다. 기존 데이터 세트를 넘어 사물의 상

태를 인식할 수 있도록 모델을 훈련하는 것이 목표입니다. 모델 학습 과정에서 노이즈와 적대적 공격에 대해 좋은 성능을 내기 위해 SAT는 노이즈를 추가하는 데 사용됩니다 [1].

## 2.4 YOLOv4 모델 아키텍처

YOLOv4는 현재 실시간 물체 감지를 위한 가장 빠르고 정확한 모델입니다. YOLOv4는 1단계 검출기 형태를 채택하고 있습니다. YOLOv4가 사용하는 아키텍처 선택은 그림 2에 나와 있습니다.

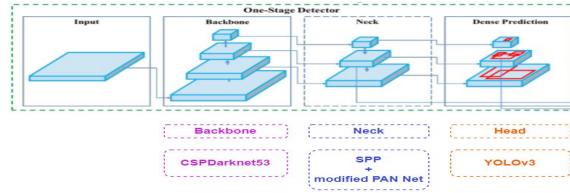


그림 2. YOLOv4 모델 아키텍처

YOLOv4 모델 아키텍처와 가중치는 알렉스 보코브스키의 GitHub 리포지토리에서 다크 넷 리포지토리를 사용하여 얻습니다. 다크넷 신경망은 그래픽 처리 장치(GPU)를 활용하여 구축되며, 사전 학습된 YOLOv4 모델의 가중치를 채택하여 본 연구의 목적에 맞게 사용할 것입니다.

## 2.5 미세 조정 프로세스

이 부분에서는 본 연구의 목적에 부합하는 딥러닝 모델을 생성하기 위해 파라미터를 조정하여 모델을 미세 조정할 예정입니다. 조정되는 파라미터는 데이터 증강, 학습 배치, 입력 크기, 학습 속도, 합성곱 계층의 필터, 각 YOLO 계층의 객체 클래스 파라미터를 설정하는 것으로 진행됩니다. 사전 학습된 YOLOv4 모델에서 모델을 미세 조정하기 위해 표 1에서 일부 파라미터를 설정했습니다.

표 1 네트워크 매개변수

모델	입력 배치 크기	하위 $s_1$ 학습 모멘텀	속도	YOLO 레이어 이전의 각 Conv. 최대.	배치
YOLOv4	416 x 416 64	16	0.001	0.949	0.0005 48
					22000

## 2.6 교육 모델

모델 아키텍처와 파라미터가 조정되면 다음 단계는 미리 학습된 YOLOv4 모델의 지식을 활용하여 전이 학습 방법을 활용하여 모델을 학습하는 것입니다. 학습은 구글 콜라보레이션을 사용하여 다크넷 플랫폼의 딥러닝 프레임워크를 기반으로 진행됩니다. 구글 콜라보레이션은 엔비디아 테슬라 K80과 같은 그래픽 처리 장치(GPU)를 무료로 제공합니다.

## 2.7 모델 평가

평가 프로세스는 획득한 여러 가중치의 성능을 비교하는 것을 목표로 합니다. 비교 후 가장 우수한 성능을 가진 가중치가 선택됩니다. 평가 프로세스는 과적합 성능을 생성하는 모델을 사용할 가능성을 피하기 위해 수행됩니다. 사용되는 평가 지표는 평균 평균 정밀도(mAP) 및 교차점 간 합집합(IoU) 값입니다.

Intersection Over Union 값은 0에서 1 사이의 범위로, 예측 경계 상자(BB)가 기준 실측 경

계 상자(BB)와 얼마나 일치하는지를 나타냅니다. 수식과 시각화의 형태로 IoU를 그림 3에 표시할 수 있습니다.


$$IoU = \frac{area(BB_{prediction} \cap BB_{groundTruth})}{area(BB_{prediction} \cup BB_{groundTruth})}$$


그림 3. IoU 공식 및 시각화

빨간색 상자는 예측된 바운딩 박스를 나타내고 녹색 상자는 지상 실측 바운딩 박스를 나타냅니다. IoU 값이 클수록 예측된 바운딩 박스와 지상 실측 바운딩 박스 사이의 거리가 짧다는 것을 의미합니다[4].

평균 정밀도(mAP)는 객체 감지에서 모델의 성능을 측정하는 데 자주 사용되는 메트릭입니다. mAP 계산에는 예측 결과를 정답으로 테스트할 가치가 있는지 여부를 결정하기 위한 IoU 값이 포함됩니다. 이름에서 알 수 있듯이 mAP는 해당 IoU 임계값을 가진 모든 클래스의 평균 정밀도 수치입니다[10]. mAP는 정밀도 및 재검출 값을 포함하는 평균 정밀도(AP)에서 계산됩니다. 여기서 정밀도와 재검출은 각각 방정식 (1)과 (2)에 정의되어 있습니다,

$$\text{정밀도} = \frac{TP}{FP+TP} \quad (1)$$

$$\text{리콜} = \frac{TP}{FN+TP} \quad (2)$$

정밀도 및 회수율 값에서 (3)에 정의된 공식으로 AP 값을 구하고 공식 (4)에서 mAP를 계산하는데 사용할 수 있습니다,

$$AP_k = \int_0^1 P_k(R_k) dR_k \quad (3)$$

$$mAP = \frac{1}{K} \sum_{i=1}^K AP_i \quad (4)$$

여기서 k는 이 연구에서 각 식재료 등급에 할당된 값을 나타내며, 표 2에 나와 있습니다.

표 2. 각 식재료 클래스의 k 값

클래스 이름	바나나	양배추	당근	오이	감자	새우	토마토	달걀	레몬	브로콜리	오렌지	K 값
												1
												2
												3
												4
												5
												6
												7
												8
												9
												10

### 3 결과 및 토론

본 연구에서는 YOLOv4 알고리즘을 사용하여 실시간으로 식재료에 대한 다중 클래스 검출을 수행하는 딥러닝 기반 모델을 성공적으로 생성했습니다. 이 모델은 mAP@0.50 값 84.90%, IoU 값 72.77%로 상당히 우수한 성능을 보였습니다. 학습에 사용된 데이터는 바나나, 양배추, 당근, 오이, 감자, 새우, 토마토, 달걀, 레몬, 브로콜리, 오렌지 등 11가지 식재료 클래스의 데이터 이미지입니다.

본 연구에서 모델 개발 과정에서 수행된 단계는 데이터셋 수집, 데이터셋 주석 달기, 기존 데이터셋 전처리, 다크넷 구축, 새로운 모델 아키텍처를 만들기 위한 파라미터 미세 조정, 모델

학습 수행, 모델 평가, 마지막으로 학습 및 검증 세트 외부의 데이터로 모델 테스트입니다. 테스트 프로세스



기본 및 보조 데이터 소스를 사용합니다. 1차 데이터는 스마트폰 카메라로 촬영한 이미지에서, 2차 데이터는 인터넷 소스에서 가져옵니다.

이 모델은 그림 4와 같이 실시간 감지 기능을 확인하기 위해 라이브 웹캠으로 테스트 중입니다.



그림 4. 라이브 웹캠에서 YOLOv4 알고리즘을 사용하여 모델이 감지한 식품 성분의 예시

## 4 관련 작업

식재료에 대한 다중 클래스 검출을 위해 관련 연구가 진행되었습니다. 이전 연구로는 '스마트 냉장고를 위한 딥러닝을 이용한 다중 소스 데이터 융합'이라는 연구로 냉장고에 있는 카메라와 딥러닝 모델을 통합하여 냉장고에 있는 과일, 채소 등 식품의 이름을 감지하는 연구가 있습니다. 이 연구에서는 SSD(ResNet), SSD(VGG16), SSD(VGG19)와 같은 여러 딥러닝 모델의 가중치를 통합하는 다중 융합 기법을 사용했습니다[12]. 또 다른 연구에서는 동일한 과제의 객체 감지에서 YOLOv4와 SSD의 성능을 비교했는데, 그 결과 YOLOv4 모델 아키텍처가 동일한 감지 과제를 가진 SSD보다 더 나은 성능을 내는 것으로 나타났습니다 [9]. 오렌지 과수원에서 오렌지 과일을 감지하는 데도 YOLO 알고리즘이 사용되었습니다. 이 연구에서는 YOLO 아키텍처의 여러 버전, 즉 YOLOv2, YOLOv3, YOLOv4를 비교했습니다. 그 결과 YOLOv4가 90.8%의 mAP 값을 얻어 가장 우수한 성능을 보였습니다. 이전 연구에서 YOLOv4 알고리즘이 보여준 엄청난 성능은 본 연구에서 YOLOv4 알고리즘을 선택한 근거가 되었습니다[8].

과일을 분류하기 위해 컨볼루션 신경망(CNN), 순환 신경망(RNN), 장단기 기억(LSTM) 등 다양한 딥러닝 접근 방식이 사용되었습니다. 다양한 접근 방식을 통합한 결과 각 접근 방식을 개별적으로 사용하는 것보다 더 나은 성능을 가진 새로운 모델이 탄생했습니다[3]. 과일과 채소와 같은 식재료를 분류하기 위한 딥러닝 모델을 생성하기 위한 또 다른 연구도 수행되었습니다. 언급된 모든 선행 연구에서 과일과 채소를 식품 성분으로만 감지하는 것을 목표로 했습니다 [2], [5]. 앞서 언급한 선행 연구들에서 YOLOv4 모델 아키텍처를 이용한 전이 학습 기법을 활용하여 과일, 채소, 동물성 단백질 유형을 포함한 식품 성분을 검출하기 위한 딥러닝 기반 모델을 생성하는 연구는 수행되지 않았습니다.

## 5 결론 및 향후 작업

이 연구는 YOLOv4 알고리즘을 사용하면 식품 성분에 대한 실시간 다중 클래스 감지를 수행할 수 있는 새로운 딥러닝 모델을 생성할 수 있음을 성공적으로 보여주었습니다. 이 연구

결과에 따르면 이 모델은 mAP@0.50 점수 84.90%, 평균 IoU 점수 72.77%를 기록했습니다. mAP@0.50 메트릭을 사용한다는 것은 바운딩 박스 영역의 IoU 임계값이 0.5 이상인 경우 모델이 실제 양성 값을 계산한다는 것을 의미합니다. 탐지에서 얻은 IoU 값이 임계값 미만이면, 이를 진양성으로 계산할 수 없습니다. mAP와 IoU를 평가 지표로 적용하면 객체 탐지 모델의 성능 결과를 모호하지 않게 확인할 수 있습니다.

추가 연구를 위해 테스트할 조건에 맞게 조정된 데이터 세트가 있는 기존 식품 클래스의 변형을 추가할 수 있습니다. 또한 연구 결과를 스마트폰과 같은 다양한 기기와 통합하거나 농업 분야에서 사용할 수 있도록 더욱 발전시킬 수 있습니다. 사용자가 음식 레시피를 더 쉽게 찾고, 음식에서 영양 정보를 얻는 등 다양한 요구에 따라 스마트 냉장고에서 추가 개발을 구현할 수도 있습니다.

## 참조

- [1] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M.: YOLOv4: 물체 감지의 최적 속도와 정확도. (2020)
- [2] Gao, X., Tao, Y., Ding, X., & Hou, R.: SSD 타겟 감지 알고리즘 기반 스마트 냉장고의 식품 인식 연구. ACM 국제 학술대회 논문집 시리즈, 303-308. (2019)
- [3] 길, H. S., & 케라, B. S.: 과일 이미지 분류를 위한 CNN-RNN-LSTM을 사용한 통합 접근법. 재료 오늘: Proceedings. (2021)
- [4] 호페스만 E.: 더 나은 탐지 평가 지표, IoU | 작성자: Eric Hofesmann | 데이터 과학을 향하여. (2020)
- [5] Li, S., Lü, J., & Ni, S.: 통합 컨볼루션 신경망과 지능형 냉장고의 과일 및 채소 인식에의 적용. 슈주 차이 지 유출리/데이터 수집 및 처리 저널, 31(1), 205-212 (2016)
- [6] Meenu, M., Kurade, C., Neelapu, B. C., Kalra, S., Ramaswamy, H. S., & Yu, Y.A.: 디지털 이미지 처리를 사용한 식품 품질 평가에 대한 간결한 검토. 식품 과학 및 기술 동향, 118, 106-124. (2021)
- [7] Minz, P. S., Sawhney, I. K., & Saini, C. S.: 식품 색도 측정을 위한 고해상도 이미지 처리 알고리즘. Measurement: 국제 측정 연맹 저널, 158. (2020)
- [8] Mirhaji, H., Soleymani, M., Asakereh, A., & Abdanan Mehdizadeh, S.: 다양한 이미징 및 조명 조건에서 간단한 접근 방식을 통해 YOLO 모델을 사용한 오렌지 과수원의 과일 감지 및 적재량 추산. 농업의 컴퓨터 및 전자, 191, 106533. (2021)
- [9] Morera, Á., Sánchez, Á., Moreno, A. B., Sappa, Á. D., & Vélez, J. F.: 여러 변수 하에서 옥외 도시 광고 패널 감지를 위한 SSD와 Yolo. Sensors (Switzerland), 20(16), 1-23. (2020)
- [10] 솔라베츠 J.: 객체 감지에서 평균 정밀도(mAP)란 무엇인가요? (2020)
- [11] Yun, S., Han, D., Chun, S., Oh, S. J., Choi, J., & Yoo, Y.: CutMix: 지역화 가능한 특징을 가진 강력한 분류자를 훈련하기 위한 정규화 전략. IEEE 국제 컴퓨터 비전 컨퍼런스, 2019-October, 6022-6031. (2019)
- [12] Zhang, W., Zhang, Y., Zhai, J., Zhao, D., Xu, L., Zhou, J., Li, Z., & Yang, S.: 스마트 냉장고를 위한 딥러닝을 사용한 다중 소스 데이터 융합. 컴퓨터 인 인더스트리, 95, 15-21. (2018)