

# ffmpeg과 YOLO를 이용한 동영상 내 객체 탐지 자동화

김지민<sup>1,\*\*</sup> · 원태호<sup>1,\*\*</sup> · 심정용<sup>1,\*\*</sup> · 윤기범<sup>1,\*\*</sup> · 주종화<sup>1</sup> · 성완용<sup>2</sup>

<sup>1</sup>동국대학교 · <sup>2</sup>(주)테크노니아

## Automating object detection in videos using ffmpeg and YOLO

Kim Ji Min<sup>1,\*\*</sup> · Won Tae-ho<sup>1,\*\*</sup> · Sim Jeong Yong<sup>1,\*\*</sup> · Yoon Ki Beom<sup>1,\*\*</sup> ·

Jong Wha J. Joo<sup>1</sup> · Sung Wonyong<sup>2</sup>

<sup>1</sup>Dongguk University · <sup>2</sup>Technonia

E-mail : kluiop1@naver.com / 96tlawjddy@gmail.com / jwjoo1204@gmail.com

### 요 약

본 논문에서는 동영상에서 일련의 과정을 거쳐 얻었던 학습데이터를 보다 간편하고 빠른 속도로 획득하는 방법을 제안한다. 음성과 영상 스트림을 처리하는 ffmpeg을 이용해 영상을 프레임화하고, 딥 러닝 기반의 YOLO 알고리즘을 사용하여 객체를 검출한다.

### 키워드

YOLO, Data, Filtering, Object, Labeling

## I. 서 론

디지털 시대, 4차 산업 혁명 시대에서 AI는 미래 산업 분야의 중요한 역할을 하고있다. 한국신용정보원에서 예측한 세계 인공지능의 2025년까지 연 평균 시장규모 성장률은 38.4%에 달한다.마켓앤마켓이 발표한 '2026년까지 컴퓨터 비전 시장의 AI(AI in Computer Vision Market, Global Forecast to 2026)' 보고서에서 언급된 컴퓨터 비전 시장 규모 또한 2021년 159억 달러에서 2026년까지 513억 달러로 증가할 전망이다. 이처럼 전세계의 기업은 AI를 도입하고 관련 기술 투자 및 지출을 확대하며 다양한 분야에 활용하고 있다. 이러한 모델을 고도화하기 위해 모델 학습에 사용되는 데이터를 가공, 수집, 라벨링 하는 산업이 더불어 발전하고 있다. 그 중 데이터를 수집하고 라벨링 하는 작업은 전문적인 인력을 고용할 정도로 그 중요도가 갈수록 높아지고 있다. AI 모델을 더욱 발전시키기 위해 이러한 과정은 필수적이라 할 수 있다. 이와 관련된 여러 톨이 발전하고 있지만 완전히 자동화된 학습 데이터 생성 서비스는 찾기 힘들다. 대부분의 톨은 사용자가 이미지에서 직접 바운딩 박스를 그리면 그에 따른 라벨링 정보가 함께 출력되는 방식이며, 해당 방식은 사용자가 한 장의 이미지에서 직접 객체를 파악해야 한다는 불편함이 있다. 또한 영상 입력에 대해서는 필요한 객체가 등장하는 부분을 이미지로 출력해 라벨링을 진행해야 한다. 본 논문에서는 기존

의 수동적인 라벨링 톨로 객체 학습데이터를 얻어내는 과정을 자동화함으로써 간편하고 빠른 속도로 대량의 입력 영상에서 양질의 학습데이터를 추출할 수 있는 솔루션을 제안한다.

## II. 선행연구의 고찰

선행연구로 기계학습을 위한 학습데이터 자동 생성 방법 및 장치 특허<sup>[3]</sup>와 자동 학습 데이터 생성을 통한 모델 학습 시스템 및 방법 특허<sup>[4]</sup>가 있다.

[3]에서는 한계점으로 학습데이터가 너무 작은 경우, 인식할 객체의 일부분이 포함되지 않는 경우가 많다는 점을 지목하고 있다. [4]의 경우, 사용자가 입력한 검색어를 기반으로 이미지 크롤링을 진행함으로써 부정확하거나 목적과 부합하지 않는 이미지 데이터가 수집될 위험이 있다.

이러한 선행 연구들의 문제점을 보완하기 위해서 본 연구에서는 데이터 필터링 기능을 제안한다. 추가적으로 [4]에서의 사용자가 입력한 검색어 기반 크롤링 방법에 착안하여 사용자가 연고자하는 객체 클래스를 직접 추가하는 방법을 제안한다.

\* speaker

\*\* These authors contributed equally to this work

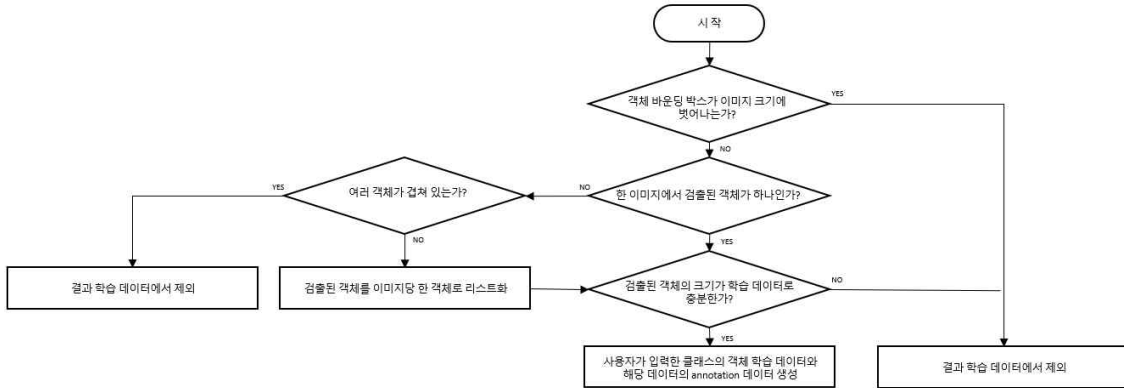


그림 1. 데이터 필터링 과정 순서도

### III. 영상에서의 학습데이터 추출 방법

본 논문에서는 딥 러닝 기반의 YOLO 알고리즘을 이용하여 영상에서 객체를 검출한다. 객체 주변에 박스 형태의 라벨을 표시함으로써 검출된 객체를 나타내며, 검출된 객체 데이터 중 학습데이터로 적합하지 않은 데이터를 필터링한다. 이후, 사용자가 요청한 객체 클래스의 학습데이터와 사용자가 설정한 형식의 어노테이션 파일을 함께 출력한다.



그림 2. 바운딩 박스가 이미지를 벗어난 학습데이터

#### 3.1 학습데이터 추출

본 논문에서 제시하는 방법은 사용자가 입력한 정보를 기반으로 한다. 사용자는 학습데이터를 추출할 영상, 추출할 학습데이터의 객체 클래스, 어노테이션 파일 형식을 입력으로 제공한다. 사용자가 입력한 영상을 초당 2프레임으로 이미지화하고, 해당 이미지에서 객체를 검출하여 학습데이터를 생성한다. 이미지에서 객체를 검출하기 위해서 실시간 객체 검출에 유리한 YOLO 알고리즘을 사용한다. YOLO에 사용되는 Cfg와 Weight 파일은 YOLOv4[1]에서 제공하는 파일을 사용했다. 그 중 Weight 파일은 COCO 데이터를 기반으로 하여 80개의 객체 클래스를 지원한다. 자세한 객체 클래스 항목은 [2]에서 확인할 수 있다. 사용자는 여러 객체 클래스에서 원하는 객체 클래스를 선택하여 학습데이터를 획득할 수 있다.

표 1. 그림 2의 바운딩 박스 좌표

xmin	ymin	xmax	ymax
0.702383	0.632035	1.001355	0.785793

YOLO 알고리즘은 이미지 좌표를 0에서 1 사이 값으로 치환하여 객체 바운딩 박스의 위치를 표시한다. 표 1에서 xmin, xmax는 각각 바운딩 박스의 왼쪽, 오른쪽 면의 x좌표를 나타내며, ymin, ymax는 각각 위, 아랫면의 y좌표를 나타낸다.

그림 2의 객체 바운딩 박스는 우측 일부가 이미지 크기를 벗어나 있다. 이 경우, 표 1의 xmax가 1을 초과하게 되며, 이를 통해 객체 바운딩 박스의 각 좌표가 0과 1사이의 값을 가져야 객체의 일부가 소실되지 않았다는 것을 알 수 있다. 이를 판단하기 위해 다음과 같은 식을 제안한다.

#### 3.2 데이터 필터링

성공적인 모델 학습은 학습데이터로 사용되는 이미지의 품질에 좌우된다. 자동화된 학습데이터 추출 과정에서 또한 데이터의 품질은 가장 중요한 고려 사항 중 하나이다. 본 논문에서 제안하는 방법은 추출된 학습데이터의 품질을 높이기 위해 3단계로 구성된 필터링 과정을 포함한다.

$$0 \leq xmin \leq 1 \quad (1)$$

$$0 \leq ymin \leq 1 \quad (2)$$

$$0 \leq xmax \leq 1 \quad (3)$$

$$0 \leq ymax \leq 1 \quad (4)$$

제안하는 데이터 필터링 과정은 그림 1과 같다. 첫 번째 필터링 단계로 YOLO 알고리즘으로 탐지한 객체의 바운딩 박스가 이미지 크기를 넘어가는지 판단한다.

식 (1)에서 (4)를 모두 만족하면 객체 바운딩 박스가 이미지 내에 존재함을 알 수 있다. 만약 식 (1)에서 (4)중, 만족하지 않는 식이 있다면 해당 학습데이터는 결과 학습데이터에서 제외한다.

두 번째 필터링 단계는 한 학습데이터에서 검출된 객체가 하나인지 확인한다. 검출된 객체가 여러개인 경우, 타 객체와 겹치는지에 따라 결과 학습데이터에서 제외할지 각 객체를 분리할지를 결정하게 된다. 이 때, 타 객체와의 위치 중복을 확인하는 방법으로 각 객체 바운딩 박스의 xmin, ymin, xmax, ymax를 사용한다. 객체 중 좌표가 중복되는 경우, 해당 학습데이터는 결과 학습데이터에서 제외된다.

마지막 필터링 단계는 객체의 크기를 점검한다. 영상에서 객체의 크기가 너무 작으면 학습에 영향을 미칠 수 있다. 따라서 너무 검출은 되었지만 객체가 너무 작은 객체의 경우 필터링을 진행한다. 작은 객체를 거르는 방법은 전체 영상에서 검출된 객체의 바운딩 박스가 차지하는 면적의 비율을 계산하여 결정한다. 그로부터 다음과 같은 식을 제안한다.

$$\frac{\text{객체 바운딩 박스 면적}}{\text{이미지 면적}} < \text{임계값} \quad (5)$$

식 (5)에서 임계값은 0.05로 임의 설정하여 진행하였다. 식 (5)를 만족하지 않은 학습데이터는 객체의 크기가 충분하지 못하다고 판단하여 결과 학습데이터에서 제외한다.

```
<annotation>
  <filename>test-2</filename>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>608</width>
    <height>608</height>
    <depth>3</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>car</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>3.1960630416870117</xmin>
      <ymin>193.35568046569824</ymin>
      <xmax>403.4572563171397</xmax>
      <ymax>524.0368137359619</ymax>
    </bndbox>
  </object>
</annotation>
```

그림 3. 출력된 XML 형식의 어노테이션 파일

```
2 0.702383 0.632035 1.001355 0.785793
```

그림 4. 출력된 TXT 형식의 어노테이션 파일

필터링 과정을 거쳐 생성된 결과 학습데이터는 사용자가 선택한 형식에 따라 그림 3, 4와 같이 어노테이션 파일이 생성되며, 사용자는 이미지 학습데이터와 더불어 해당 어노테이션 파일을 획득하게 된다.

### 3.3 객체 클래스 추가

COCO 데이터에서 제공하는 80개의 객체 클래스에 사용자가 원하는 객체 클래스가 없는 경우에도 학습데이터를 제공하기 위해 객체 클래스 추가 기능을 제공한다.

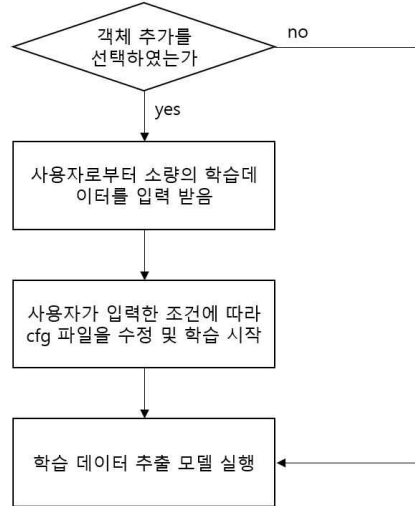


그림 5. 객체 추가 기능의 순서도

객체 클래스 추가 방법은 그림 5와 같은 순서를 따르며, 단기간에 학습된 YOLO 모델의 학습 파일을 사용해 새로운 학습데이터를 추출하는 방식으로 구현된다. YOLO 모델의 Cfg 파일을 수정하고 사용자가 입력한 소량의 학습데이터를 이용해 학습을 진행한다. 본 논문에서는 객체 클래스 추가 방법을 구현하기 위해 YOLOv3 tiny, YOLOv3, YOLOv4중 가장 빠른 학습 속도를 보여준 YOLOv3 tiny를 채택하였다. 학습 결과로 생성된 Weight와 Cfg 파일을 이용해 학습데이터 추출 모델을 실행한다.

예시를 위해 YOLOv3 tiny 모델을 이용하여 각각 800장의 서있는 사람, 누워있는 사람을 fallen\_body, normal\_body 2개의 클래스로 분류하여 4000 epoch로 학습을 진행하였다. 학습 시간은 약 70분이 소요되었으며, Weight 파일이 정상적으로 생성되었다. 해당 Weight와 Cfg 파일을 이용해 학습데이터 추출 모델을 실행하여 학습데이터를 추출하였다. 추출한 결과 예시는 다음과 같다.

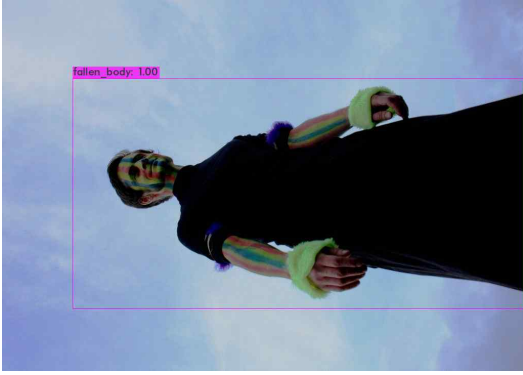


그림 6. 생성된 fallen\_body 클래스의 학습데이터

0 0.604843 0.520723 0.773749 0.488947

그림 7. 생성된 TXT 어노테이션 파일

이미지에서 사람의 몸이 가로로 되어 있는 경우 fallen\_body 클래스를 출력하게 된다. 그림 6에서 fallen\_body 객체가 정상적으로 검출되었고, 그림 7의 어노테이션 파일 또한 함께 출력되었다.

#### IV. 결 론

본 논문에서는 수동적인 방식의 학습데이터 생성의 비효율적인 생성 방식에 주목하여 자동화 방식의 학습데이터 생성 방법을 제안하였다. 이를 통하여 학습데이터 생성 방식에 비약적인 성장을 가져올 것으로 기대된다. 또한 이를 통해 많은 양의 질 좋은 학습 데이터를 생성함에 따라 인공지능 시장 규모 및 인공지능 산업 발전에 큰 영향을 끼칠 것으로 기대된다. 향후 연구에서는 한정적으로 지원하는 어노테이션 파일을 여러 형식으로 지원할 수 있는 기술에 대한 연구를 진행할 예정이다.

#### Acknowledgement

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) (2016-0-00017)

#### References

- [1] <https://github.com/Tianxiaomo/pytorch-YOLOv4> (2020-2021)
- [2] [https://github.com/leggedrobotics/darknet\\_ros](https://github.com/leggedrobotics/darknet_ros) (2017~2021)
- [3] 한동대학교 산학협력단 안재경, 기계학습을 위한 학습데이터 자동 생성 방법 및 장치 특허,

- 1020180166601, 2018.12.20, 2021.02.18  
 [4] 주식회사 아임 클라우드, 자동 학습 데이터 생성을 통한 모델 학습 시스템 및 방법 특허, 1020170183940, 2017.12.29, 2021.03.09