멀티미디어 데이터에서 객체 식별을 위한 딥러닝 기반의 시스템 설계 및 구현

고상균*, 김봉재†, 김정동†
*선문대학교 컴퓨터공학부
{highsg19101*, bjkim0422†, kjdvhu† }@gmail.com

Design and Implementation of Deep Learning based System for Object Identification of Multimedia Data

Sang-Gyun Ko*, Bongjae Kim†, Jeong-Dong Kim†
*School of Computer Science and Engineering, Sun Moon University

요 약

최근 CCTV나 블랙박스 등 멀티미디어 데이터를 생성해내는 장치의 사용이 늘어나고 있다. 이러한 대용량 멀티미디어 데이터가 증가함에 따라 사용자가 동영상과 같은 멀티미디어 데이터 내의 객체를 식별하기 위해서는 많은 시간을 할애하여 매뉴얼하게 일일이 찾아야 하는 한계점이 있다.

본 논문에서는 사용자가 동영상 및 이미지에서와 같은 멀티미디어 데이터에서 객체를 자동으로 식별할 수 있 수 있는 딥러닝 기반의 객체 식별 및 검색 모델을 제안한다. 제안하는 객체 식별 검색은 이미지 검색과 동영상 검색을 지원한다. 이미지 검색에서는 이미지에 존재하는 동일한 객체를 검색 대상 이미지들에서 객체를 식별하고, 이미지에 존재하는 객체를 검색하여 결과로 반환한다. 또한 동영상 검색에서는 동영상에서 검색하고자 하는 객체를 식별하고 객체가 출현하는 시간을 전처리과정을 통해 기록하며, 검색하고자 하는 동영상 내에 존재하는 객체의 검색이 가능하다. 따라서 사용자가 동영상에서 객체의 검색 시 키워드 검색이 가능하여 동영상을 모두 재생하서 객체를 식별해야 하는 번거로움을 해결할 수 있다.

1. 서론

최근 인공지능(Artificial Intelligence: AI), 빅데이터(Big Data), IoT(Internet of Thing), 클라우드(Cloud) 등의 기술 발전은 4차산업혁명을 주도하고 있으며, 인터넷 및 스마트 기기의 보급으로 스마트 사회가 도해하였다. 이러한 기술의 발전은 딥러닝(Deep Learning)의 GPU(Graphics Processing Unit)를 비롯한 서버 환경이 급속도로 발전함에 따라 컴퓨터 비젼, 자연어처리, 음성인식, 객체 분류, 동작 식별 등 다양한 분야에서 딥러닝을 적용한 연구가 진행되고 있다[1-2].

특히 Convolutional Neural Network 기반 모델이 객체 분류 속도와 정확성에 있어 큰 성과를 거두어 객체 분류에 관한 연구에서 딥러닝 기술은 큰 성과를 내고 있다[3]. 또한 멀티미디어 데이터의 실시간 객체 식별 및 분류를 위해 최근 Darknet-YOLO가 오픈소스로 발표되어 많은 주목을 받고 있으며, Darknet-YOLO의 활용 사례는 1)고속도로에서 주행하는 자동차를 실시간으로 탐지[2], 2)보행자 검출 시스템[3], 3) CCTV 영상에서의 쓰러짐 검출[4] 등 실시간 객체 분류에서 연구가 활발한 연구가 진행중에 있다.

† 교신저자: 김정동, 김봉재

최근 CCTV나 블랙박스와 같은 대용량의 동영상들의 생성이 많아지고 있는 가운데 대용량의 동영상에서 사용자가 찾고 싶은 장면을 찾아내기 위해서는 동영상을 처음부터 끝까지 열람해야 하는 에로사항이 있다. 예를 들면, 20시간짜리 CCTV 동영상에서 특정 차량을 검색하고자 한다면 동영상에서 차량이 기록된 부분을 사용자가 매뉴얼하기 찾아야 하며, 이때 소요되는 시간은 일반재생 시 20시간이이 소요되는 한계점이 있다.

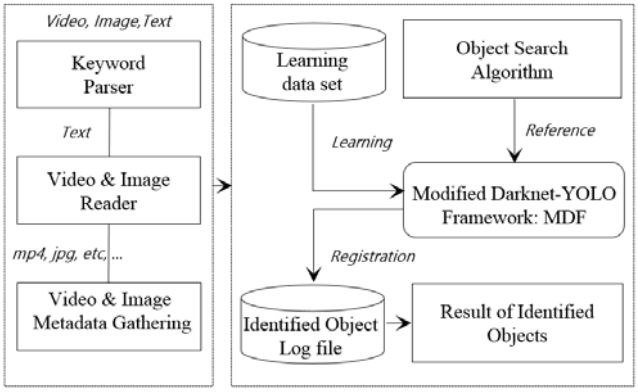
본 논문에서는 이와 같은 한계점을 해결하기 위한 방안으로 멀티미디어 데이터에서 사용자가 검색하고자 하는 객체를 자동으로 식별하고 검색할 수 있는 딥러닝 기반의 객체 식별 및 검색 모델을 제안한다. 제안하는 객체 식별을 위한 딥러닝 기반 알고리즘[5] 및 시스템의 핵심 기술은 실시간 다중 객체 분류 프레임워크인 YOLO(You only look once: Real-Time Object Detection) 이다 [1].

YOLO를 이용하여 사진, 동영상과 같은 멀티미디어에서 여러 가지 객체를 추출 할 수 있으며 사용자가 사진을 업로드 하면 사진내의 가장 많이 출현된 객체들을 식별하고 검색결과로 반환하는 기능을 제공한다. 또한 사진과 같이 동영상 내의 객체를 찾기 위해 동영상과 동영상 내의 찾고자 하는 객체의 이름을 입력하면 자동으로 그 객체가 출현한 동영상 내의 위치를 반환 해주는 웹 기반 영상 및

이미지 검색 시스템을 제안한다.

3. 제안 모델

그림 1은 본 논문에서 제안하는 멀티미디어 데이터에서 객체 식별을 위한 시스템의 전체 모델을 나타낸다.



(그림 1) 객체 식별을 위한 딥러닝 기반 시스템 모델

본 논문에서 제안하는 영상 및 이미지 데이터에서 실시간 객체 식별을 위한 시스템 모델에서는 먼저 멀티미디어 데이터(동영상, 이미지, 텍스트)를 사용자로부터 입력받는다.

입력받은 멀티미디어 데이터의 전처리를 위해 그림1의 “Keyword Parser”, “Video & Image Reader”, 그리고 “Video & Image Metadata Gathering”의 단계를 거쳐 멀티미디어 데이터의 포맷과 데이터가 가지고 있는 메타데이터를 추출한다.

“Modified Darknet-YOLO Framework: MDF”에서는 기존의 YOLO 프레임워크에서 실시간 분석된 멀티미디어 데이터의 객체의 로그 정보를 기록 및 저장할 수 있도록 수정하였다. MDF는 멀티미디어 데이터의 객체의 식별을 위해 데이터의 학습을 지원한다. 이러한 학습을 위해서는 “Learning Data Set”에 저장된 많은 데이터를 활용하게 된다.

“Object Search Algorithm”는 MDF에서 식별한 객체의 로그 정보를 데이터베이스에 저장하며, 사용자로부터 입력된 키워드 검색에서 객체의 검색을 지원한다. 즉, 사용자로부터 검색하고자 하는 텍스트나 이미지를 입력으로 받으면, 딥러닝을 통한 기 색인되어 있는 다양한 종류 멀티미디어 데이터에서 결과를 도출한다.

또한 동영상 데이터를 입력으로 받으면, 동영상의 FPS(Frames Per Second)와 재생시간 정보를 추출하고, MDF에서 추출된 객체를 “Object Search Algorithm”에 의해 객체가 검색된 시간(초) 정보를 사용자에게 제공한다.

“Result of Identified Objects”는 객체의 분류로 검색결과의 데이터 형식은 텍스트와 이미지이다. 텍스트 검색에서는 동영상에서 식별한 객체의 시간정보를 반환하며, 이미지 검색에서는 동영상 내에서 동일한 객체를 식별하고 식별된 객체의 시간정보를 검색 결과를 제공한다.

4. 구현

4.1 실험 환경과 데이터 셋

본 논문에서 제안하는 객체 식별을 위한 딥러닝 기반 시스템의 개발 환경은 표1에 나타낸다. 개발 환경의 특징은 실시간 객체 검출의 성능을 높이기 위해 멀티 GPU 환경을 구축하였고, 웹 프레임워크인 Flask를 사용하여 웹 기반 서비스를 개발하였다.

<표 1> 제안 시스템의 개발 환경.

환경		설명
GPU	GTX 1080 ti * 4	멀티 GPU 환경
CPU	i7-6850K Broadwell-E	프로세서
RAM	DDR4 PC4-192000 8G * 8	Main memory
S/W	HTML, JS, CSS	UI 구현
	Flask	웹 서버 구현
	Darknet-YOLOv3	딥러닝 프레임워크
	openCV 2.4	영상처리
	cuDNN 5.1	딥러닝 가속화
	CUDA 8.0	GPU 프로그래밍
	python 2.4	서버 알고리즘 구현
	Ubuntu 16.04	운영체제

객체 분류를 위해 사용한 데이터 셋으로는 직접 수집한 국산 차종 i40, Santafe, Ray 로 총 2500장을 수집하여 학습을 진행하였다. YOLO 에서 기본적으로 제공하는 coco dataset[6]이 학습되어진 실행파일을 포함하여 93가지의 객체를 분류한다.

4.2 구현 결과

그림 2는 동영상 데이터에서 객체의 검색을 지원하기 위한 웹 기반 인터페이스의 구현 결과의 스냅샷을 나타낸다. 사용자가 동영상 데이터와 검색하고자 하는 객체의 키워드를 입력으로 제공하며, 만약 동영상 내에 찾고자 하는 객체가 존재한다면 동영상에서 객체가 등장하는 시간을 결과로 제공한다.

그림 2-①은 사용자가 찾고자 하는 동영상 파일을 업로드 할 수 있는 파일 선택 창을 나타내며, 그림 2-②는 동영상 내에서 검색하고자 하는 객체의 이름을 키워드로 입력하는 입력 창이다. 그림 2-③은 업로드 버튼으로 입력된 동영상 데이터와 키워드는 서버에 전송되고 동영상에 존재하는 객체가 실시간으로 추출되며, 키워드와 매칭을 통한 객체 식별이 가능하다.

그림 2-④는 입력으로 제공한 “ray” 키워드가 입력된 동영상에서 등장하는 시간 정보를 초단위로 나타낸 결과 화면이다. 본 논문에서 제안하는 객체 식별 모델은 동영상 데이터에 대한 1회의 분석을 통해, 동영상에 등장하는 모든 객체에 대한 식별 정보를 로그 파일로 기록하고 관리함에 따라 동일한 동영상에서 다른 키워드의 객체 검색에서 동영상에 대한 분석 없이 실시간으로 키워드 검색의 결과를 도출할 수 있다.

(그림 2) 동영상 데이터에서 객체 검색 결과의 스냅샷.

그림 3은 이미지 데이터를 입력으로 제공했을 때 동일한 객체에 대한 다른 이미지들의 검색 결과를 제공한다. 만약 입력으로 제공한 이미지 데이터에 다수의 서로 다른객체가 존재한다면, 객체의 출현 빈도와 가장 높은 객체를 검색의 결과로 제공한다.

그림 3-①과 3-②는 입력으로 제공할 이미지 데이터를 업로드하기 위한 버튼이며, 3-③는 객체의 출현 빈도수를 나타내며 가장 많이 출현한 객체를 식별하고, 사용자가 업로드한 이미지 데이터에서 추출한 객체의 출현 빈도를 제공한다. 그림 3-④는 사용자가 업로드 한 이미지 데이터에서 가장 많이 출현한 객체가 포함된 객체와 동일한 이미지의 검색 결과를 제공한다.

(그림 3) 이미지 데이터의 객체 검색 결과의 스냅샷.

5. 결론

본 논문에서는 멀티미디어 데이터에서 객체 식별을 위한 딥러닝 기반의 모델을 제안했다. 제안하는 객체 식별 모델은 동영상과 이미지 데이터를 기반으로 객체 식별 및 검색이 가능하다. 제안 모델은 동영상 데이터에서는 검색하고자 하는 객체가 출연한 시간정보를 결과로 제공하며, 이미지 데이터에서의 객체 검색에서는 입력으로 제공한 이미지 데이터의 가장 많이 출현한 객체의 검색이 가능하다. 또한 딥러닝 기반 객체 식별을 통한 검색이 가능함을 보이기 위해 웹을 활용한 구현 환경을 구축하고 개발하였다. 만약, 사용자가 동영상 데이터에서 찾고자하는 객체의 분석 시간을 동영상의 재생시간이라고 가정할 때, 기존의 동영상을 재생하여 사용자가 수동적으로 객체를 검색하는 시간을 비교한다면 제안 모델을 통한 객체 검색의 성능이 매우 우수함을 확인하였다. 본 논문에서 제안한 객체 식별 모델은 CCTV나 차량의 블랙박스 등에서 사용자의 목적에 따라 활용도가 매우 높으며, 향후 실시간 객체 식별이 요구되는 범죄 및 화재 예방, 그리고 교통단속 등에서 활용 될 수 있다.

향후 연구로는 실시간 객체 식별 알고리즘의 신뢰성과 정확성을 위한 비교 평가 및 검증에 대한 연구가 요구되며, 사용 목적에 따른 식별 가능한 객체 목록에 대해 추가적인 딥러닝에 대한 연구를 진행하고자 한다.

참고문헌

[1] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., “You only look once: Unified, real-time object detection”, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.

[2] Chang-Jin Seo, "The Study of Car Detection on the Highway using YOLOv2 and UAVs," The Transaction of the Korean Institute of Electrical Engineers P, Vol.67P, No.1, pp. 42-46, 2018.

[3] Kyu Min Park, Youngwoo Kim, Seokjun Kang and Dong Seog Han, "Pedestrian recognition system using YOLO detection system," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp.507-507, 2017.

[4] Chulyeon Kim, Taekjin Han, Illo Yoon, Yoonjin Lee, Jiyoung Lee, Gyunghyun Choi, Chung-In Won, and Young-Min Kim, “Fall detection in CCTV using YOLO”, Korea Computer Congres. Vol.3, No.1, pp.785-787, 2018.

[5] Sang-Gyun Ko, Bongjae Kim, and Jeong-Dong Kim, “Deep Learning-based Algorithm for Object Identification in Multimedia”, The 13th International Conference on Ubiquitous Information Technologies and Applications, 2018 (Submission).

[6] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., and Zitnick, C. L, “Microsoft coco: Common objects in context”, In European conference on computer vision, pp.740-755, 2014.