

개인 맞춤형 광고를 위한 딥러닝 검출 툴을 이용한 영상 카테고리 분류기

*박진영 **안원진 ***안천수 ****강석주

서강대학교 전자공학과

*wlsdud7907@naver.com **mickeyahn@naver.com

ahncs95@gmail.com *sjkang@sogang.ac.kr

Video Category Classifier for Personalized Advertisements using Deep Learning
Detection Tool YOLO

*Jin-Young Park **Won-Jin Ahn ***Cheon-Su Ahn ****Suk-Ju Kang

Department of Electronic Engineering, Sogang University

요약

최근 인터넷 영상 매체가 발전하고 대중화되며 이를 통한 광고 효과가 커지고 있다. 이들 영상에 관련된 광고를 자동으로 연결할 수 있다면 효과적인 것이다. 본 논문은 딥러닝 검출 툴을 적용한 영상 카테고리 분류 기법을 제안한다. 이 기법은 주어진 영상을 몇 가지 카테고리로 분류하고, 분류 정보를 바탕으로 관련성이 높은 광고를 연결지어, 결과적으로 영상 시청자에게 맞춤형 광고를 제시한다.

1. 서론

최근 각종 SNS, 유튜브 등 영상 기반 매체들의 발전과 대중화로 인하여 영상의 시청자 도달성이 높아지고, 이에 힘입어 영상에 포함된 광고의 파급력 또한 증가하고 있다. 특정 부류의 영상에 그 내용과 관련된 광고를 자동으로 연결시킬 수 있다면 효과적인 것이다.

영상을 카테고리 분류하고자 한 기존의 노력을 소개하면 다음과 같다. 1) 영상 장르의 특징에 주목한 연구[1]로, 만화, 광고, 음악, 뉴스, 스포츠 등 여러 영상의 장르별 특징을 분석하고, 장르 사이의 유사성, 비유사성, 혼동 가능성에 대해 분석하여 영상 분류기를 제작하였다. 2) 영상에 첨부된 텍스트에 주목한 연구[2]로, 유튜브 영상의 제목, 설명, 태그, 댓글 정보를 이용하여 텍스트 분류기를 제작하였다. 이를 바탕으로 영상 분류기와 텍스트 분류기의 분류 결과를 종합하여 향상된 성능의 분류기를 제작하였다. 3) 지도 학습 기법을 이용한 연구[3]로, 지도 학습 모델인 SVM(Support Vector Machine)을 기반으로 하는 영상 분류기를 제작하였다. 4) 영상의 의미론에 주목한 연구[4]로, 영상 전체의 특징만을 분석하는 기존 방식에서 나아가 영상 내 특정 행동이나 사건 등 의미론을 분석하여 영상 분류에 이용하였다. 5) 멀티모달리티를 이용하여 시청자 맞춤형 TV 광고를 전달하는 연구[5]로, 다양한 채널에서 시각 및 텍스트 데이터를 받아서 PLSA(Probabilistic Latent Semantics Analysis) 기법으로 데이터에 적합한 광고군을 파악한 뒤, 시청자에게 이 광고군을 전달하였다.

본 논문은 과거의 연구들과 달리 딥러닝 기반의 실시간 영상 클래스 검출기 YOLO(You Only Look Once) v3[6]를 영상 카테고리 분류기에 적용하였다.

YOLO가 등장하기 전까지 영상 클래스 검출기는 대부분

R-CNN(Region-based Convolutional Neural Network) 방식을 사용했다. R-CNN 방식은 두 단으로 구성된다. 1단은 주어진 이미지에서 수천 개의 후보군을 파악하고, 2단은 각 후보에 대한 분류를 수행한다. 이렇듯 두 단을 거치며 이미지에서 물체의 클래스를 검출한다.

R-CNN 방식에는 단점이 4가지 있다. 1) 두 단을 각각 구축해야 하므로 검출기 전체를 한 번에 최적화시킬 수 없다. 2) 모든 후보에 대하여 각각 분류를 수행하기 때문에 연산량이 많다. 이미지 하나를 처리하는 데에도 수 초 정도의 시간이 걸린다. 3) 이미지 하나를 여러 부분으로 나누어 분석하기 때문에 이미지 전체에 대한 이해도가 낮다. 4) 배경 노이즈를 물체로 검출하는 False Positive(1종 오류)가 발생한다.

YOLO는 GoogLeNet을 변형시킨 검출기와 다수의 합성곱 계층과 완전연결 계층으로 구성된 한 단으로 이루어진다. YOLO에 이미지를 입력하면 $13 \times 13 \times 125$ 특징 맵이 출력된다. 각 특징 맵은 25 앵커박스로 구성되는데, 앵커박스에는 경계박스의 좌표, 높이, 폭과 클래스에 대한 정보가 들어있다. 이를 통하여 YOLO는 이미지 속 물체 둘레에 경계박스를 치는 동시에, 물체가 80개의 YOLO 클래스 중 어느 클래스에 어떤 확률로 일치하는지 분석한다.

YOLO는 기존의 방법에 비해 3가지 장점이 있다. 1) 한 단으로 구성되기 때문에 검출기 전체를 한 번에 최적화시킬 수 있다. 2) 초당 10장 이상의 이미지를 분석하기 때문에 실시간 영상 분석이 가능하다. 3) 이미지를 분할 없이 인식하기 때문에 이미지 전체에 대한 이해도가 높다.

2. 본론

서론에서 논한 것처럼 YOLO는 효과적인 영상 클래스 검출기이므로

로, YOLO로 검출한 클래스의 경향으로 영상의 카테고리르 파악할 수 있다. 이렇게 파악한 카테고리르와 그 카테고리르에 관련된 광고를 연결짓는다면 개인의 영상 재생 목록으로부터 맞춤형 광고를 제시할 수 있다. 본 논문은 개인 맞춤형 광고를 위한 광고 목록 생성 모델을 <그림 1>과 같이 제안한다.

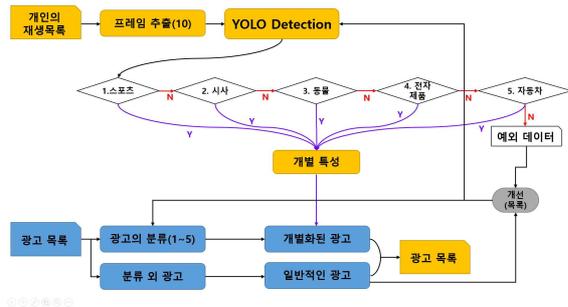


그림 1. 개인 맞춤형 광고 목록 생성 모델

본 논문에서는 <그림 1>과 같이 개인 맞춤형 광고를 제시하기 위해 서 우선 영상을 다음과 같이 다섯 카테고리로 나누었다. 1) 스포츠, 2) 시사, 3) 동물, 4) 전자제품 리뷰, 5) 자동차 리뷰. 이 다섯 카테고리는 분류에서 순서대로 우선순위를 가지도록 하였다. 즉 영상이 스포츠로 분류되는지 우선 판단하여, 스포츠로 분류되었을 경우 뒤의 네 카테고리일 여지를 배제하고, 스포츠로 분류되지 않았을 경우 시사로 분류되는지 판단하는 방식이다. 이 우선순위는 각 카테고리의 검출 빈도와 배제성에 따라 선정되었다. 예를 들어 1순위 카테고리인 스포츠는 하위 카테고리인 시사, 동물, 전자제품 리뷰, 자동차 리뷰에 비하여 검출 빈도는 낮은 대신 검출되는 경우 스포츠에 관련된 영상일 확률이 높았기 때문에 우선 순위를 1순위로 결정하였다. 이렇게 카테고리별로 분류한 영상에 각각 1) 스포츠 브랜드와 스포츠 상품 광고, 2) 비교적 높은 연령층을 대상으로 하는 넥타이나 시계 등의 광고, 3) 반려동물 제품군 광고, 4) 해당 전자제품과 연관하거나 같은 제품군 광고, 5) 자동차, 렌트 서비스, 엔진 오일과 세정제 등 자동차 관련 제품군 광고를 추천하여 광고 효과를 극대화할 수 있을 것으로 기대된다.

다음으로 YOLO가 다섯 카테고리를 서로 배타적으로 분류할 수 있는지 확인하고자 분류 결과를 F1 score로 평가하였다. F1 score를 낼 때 분류 결과는 다음과 네 가지 중 하나의 결과를 갖는다. 1) True Positive, 2) True Negative, 3) False Positive, 4) False Negative. YOLO 검출 결과 발생한 모든 경계박스의 클래스 중 카테고리별로 True로 분석하는 기준 YOLO 클래스는 <표 1>과 같다.

우선순위	종류	키워드
1	스포츠	sports ball, baseball globe, baseball bat, tennis racket
2	시사	(neck) tie
3	동물	dog, cat
4	전자제품 리뷰	TV monitor, laptop, mouse remote, keyboard, cell phone, toaster
5	자동차	car, motorbike

표 1. 영상 카테고리별 기준 YOLO 클래스

성능 평가 방법으로는 주어진 영상 전체를 11등분 하여 얻은 12개 프레임 중 첫 프레임과 마지막 프레임을 제외한 10개 프레임을 분석한 뒤, 분석 결과를 합산하여 F1 score를 내는 방법을 사용하였다. 이 평가 방법으로 카테고리별로 50개 영상을 분석하고 F1 score를 내었다. 영상은 모두 유튜브에서 획득하였고, 구독자 수가 많은 채널의 영상, 조회 수가 높은 영상, 또는 검색어와 관련성이 높은 영상을 기준으로 하였다. YOLO의 프레임별 검출 결과는 <그림 2>와 같이 나타난다.

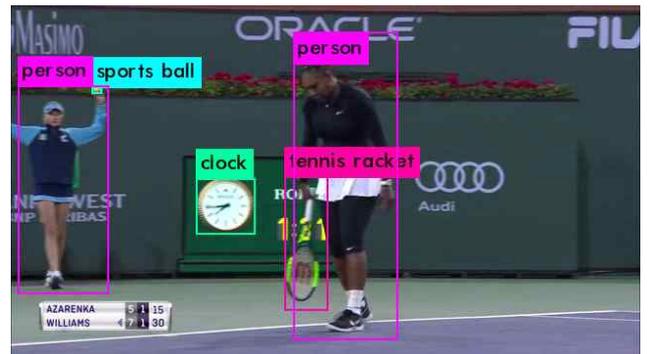


그림 2. YOLO 검출 결과 예시

한편 YOLO 클래스는 있으나 경계박스로 검출되지 않은 물체는 분석에서 제외되었는데, 이 예외 데이터는 광고 목록 중 다섯 카테고리에 포함되지 않는 광고의 데이터와 더해 차후 분류 체계를 개선하는데 쓸 수 있을 것으로 생각된다.

3. 실험 결과

본론에서 논한 것처럼 각 영상의 F1 score를 내고, 카테고리별로 F1 score의 평균을 구하니 <표 2>와 같았다.

주제		F1 Score	주제		F1 Score
스포츠	축구	0.774	전자제품	마우스	0.649
	농구	0.267		리모컨	0.753
	야구	0.936		키보드	0.813
	테니스	0.959		휴대전화	0.893
	배구	0.473		토스터	0.171
	전체	0.682		냉장고	0.96
시사		0.677	시계		0.957
동물		0.925	헤어 드라이어		0
전자제품	TV 모니터	0.908	전체		0.703
	노트북	0.928	자동차		0.902

표 2. 영상 카테고리별 평균 F1 score

나머지 클래스는 해당 카테고리에서 모두 False로 간주한다. 분류

스포츠 카테고리의 경우 축구, 농구, 야구, 테니스, 배구 다섯 종목

에서 10개씩 영상을 선정하였다. 축구, 야구, 테니스의 F1 score가 평균 0.89로 높음을 확인할 수 있다. 이들 종목의 경우 축구는 공이 빠르게 움직이지 않아 sports ball 클래스 검출 빈도가 높고, 야구는 sports ball 클래스 검출 빈도는 낮으나 야구에 특화된 baseball bat과 baseball glove 클래스 등장 빈도가 높으며, 테니스는 sports ball과 tennis racket 클래스 등장 빈도가 높기 때문이다.

시사 카테고리의 경우 CNN, ABC, JTBC, KBS, YTN 다섯 채널에서 10개씩 영상을 선정하였다. ABC, JTBC, KBS 채널의 F1 score의 평균 0.91로 높음을 확인할 수 있다. 이들 채널의 경우 시사 카테고리 기준 클래스인 neck tie의 등장 빈도가 높기 때문이다.

동물 카테고리의 경우 개나 고양이로 다루는 10개 채널에서 5개씩 영상을 선정하였다. 전체 영상의 F1 score가 평균 0.925로 다섯 카테고리 중 가장 높음을 확인할 수 있다.

전자제품 리뷰 카테고리의 경우 YOLO 클래스 중 전자제품 클래스 10종에 해당하는 제품의 리뷰 영상을 5개씩 선정하였다. 하위 2개 클래스인 toaster와 hair drier를 제외한 F1 score가 평균 0.858으로 높음을 확인할 수 있다.

자동차 리뷰 카테고리의 경우 구독자가 많은 10개 채널에서 5개씩 영상을 선정하였다. 전체 영상의 F1 score가 평균 0.902로 다섯 카테고리 중 두 번째로 높음을 확인할 수 있다. 자동차 클래스의 경우 유리에 반사된 사영도 검출하는 등 검출 빈도가 높음을 확인할 수 있다.

4. 한계점

실험 결과와 <표 2>에서 확인할 수 있는 것처럼, 본 모델에는 한계점이 있다.

스포츠 카테고리의 농구와 배구의 F1 score가 낮는데, 공이 빠르게 움직여 sports ball 클래스 검출 빈도가 낮거나, 공이 화면 밖에 있어 sports ball 클래스 등장 빈도가 낮기 때문이다. 시사 카테고리의 경우 CNN과 YTN 채널의 F1 score가 낮는데, 여성 진행자가 나오거나 인물들이 벡타이를 매지 않는 등 neck tie 클래스 등장 빈도가 낮기 때문이다. 전자제품 리뷰 카테고리의 toaster와 hair drier의 F1 score가 매우 낮는데, YOLO의 학습이 불완전해 클래스 검출 빈도가 낮기 때문이다.

제한된 환경상 실험에 이미 학습된 가중치를 가진 YOLO를 활용했기 때문에, 애니메이션 같은 특정 종류의 영상이나 toaster와 hair drier와 같은 특정 클래스가 등장하는 영상에서 클래스 검출 빈도가 낮은 문제가 공통으로 발생하였다. 실험에 영상 전체가 아닌 등 간격 분할된 몇 개 프레임을 추출해 분석했기 때문에, 일부 프레임에서 클래스 등장 빈도가 낮은 문제가 발생하였다.

5. 차후 연구 방향 및 결론

본론과 실험 결과에서 논한 것과 같이, 카테고리별로 차이를 보이는 F1 score 문제를 해결하기 위하여 각 카테고리의 기준 클래스 검출 빈도가 낮을수록 분류 우선순위를 높게 둬으로써 낮은 검출 빈도의 카테고리에 대한 분류도 정확히 이루어질 수 있도록 하였다.

한계점에서 논한 것과 같이, 제한된 환경상 이미 학습된 가중치의 YOLO를 활용해 실험을 진행하였기 때문에 일부 클래스의 검출 빈도가 낮은 문제가 발생하였다. 본 논문에서 분류한 카테고리에 부합하는

기준 클래스를 YOLO에 추가한 뒤 학습을 수행한다면 대부분의 정확도 문제를 해결할 수 있을 것이다. 또한, 몇 개 프레임을 추출해 분석하며 실험을 진행하였기 때문에 클래스 등장 빈도가 낮은 문제가 발생하였다. YOLO가 실시간 영상 클래스 검출기라는 점을 활용하여 몇 개 프레임이 아닌 특정 구간 분할 영상을 추출하여 분석하면 문제를 해결할 수 있을 것이다.

본론에서 논한 것과 같이, 본 논문은 영상 카테고리 분류에 기준 클래스 검출 여부만을 판단 기준으로 삼았지만, 추가로 특정 영상 카테고리에 등장하는 클래스의 패턴이나 클래스 간 관계를 분석하면 분류의 정확도가 높아질 것이다. 예를 들어 같은 위치, 크기, 클래스로 계속 등장하는 경계박스, 또는 한 쌍으로 등장하는 클래스 등이 있다. 서론에서 논한 것과 같이, 영상의 텍스트 정보를 기존의 시각 정보와 결합해 새로운 분류의 기준으로 삼을 수도 있을 것이다.

지금까지 딥러닝 검출 툴 YOLO를 적용하여 영상을 카테고리별로 분류하는 기법에 대해 논하였다. 이 기법을 통하여 기준 클래스의 검출 빈도에 따라 영상을 카테고리별로 분류한다면, 특정 시청자의 영상 재생 목록을 통하여 영상 시청자에게 광고 효과를 극대화 시킨 맞춤형 광고를 제공할 수 있을 것으로 기대한다.

6. 참고문헌

- [1] B. T. Truong, C. Dorai, "Automatic genre identification for content-based video categorization," *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, September 2000.
- [2] K. Filippova, K. B. Hall, "Improved video categorization from text metadata and user comments," *SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 835-842, July 2011.
- [3] X. Yuan, W. Lai, T. Mei, X. Hua, X. Wu, S. Li, "Automatic Video Genre Categorization using Hierarchical SVM," *2006 International Conference on Image Processing*, October 2006.
- [4] Y. Jiang, Z. Wu, J. Wang, X. Xue, S. Chang, "Exploiting Feature and Class Relationships in Video Categorization with Regularized Deep Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 40, no. 2, pp. 352-364, February 2018.
- [5] J. Wang, L. Duan, L. Xu, H. Lu, J. S. Jin, "TV ad video categorization with probabilistic latent concept learning," *MIR '07 Proceedings of the international workshop on Workshop on multimedia information retrieval*, pp. 217-226, September 2007.
- [6] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," <https://pjreddie.com/>.