# 딥러닝 표정 인식을 통한 운동 영상 유튜브 하이라이트 업로드 자동화(RPA) 설계

신동욱, 문남미
호서대학교 벤처대학원 융합공학과,
sdw1904@naver.com, mnm@hoseo.edu

# Design of Automation (RPA) for uploading workout videos to YouTube highlights through deep learning facial expression recognition

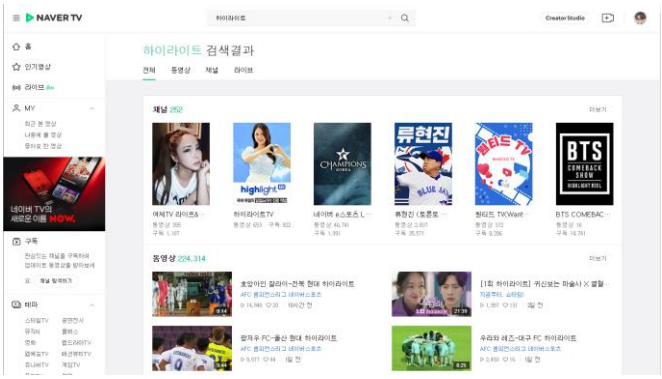
Dong-Wook Shin, NamMee Moon
Dept. of Convergence Engineering, Hoseo Graduate School of Venture, Hoseo Univ.

**요 약**

본 논문은 유튜브에 업로드 된 운동 영상을 시청하는 사람의 얼굴 영역을 YoloV3을 이용하여 얼굴 영상에서 눈 및 입술영역을 검출하는 방법을 연구하여, YoloV3은 딥 러닝을 이용한 물체 검출 방법으로 기존의 특징 기반 방법에 비해 성능이 우수한 것으로 알려져 있다. 본 논문에서는 영상을 다차원적으로 분리하고 클래스 확률(Class Probability)을 적용하여 하나의 회귀 문제로 접근한다. 영상의 1 frame을 입력 이미지로 CNN을 통해 텐서(Tensor)의 그리드로 나누고, 각 구간에 따라 객체인 경계박스와 클래스 확률을 생성해 해당 구역의 눈과 입을 검출한다. 검출된 이미지 감성 분석을 통해, 운동 영상 중 하이라이트 부분을 자동으로 선별하는 시스템을 설계하였다.

## 1. 서론

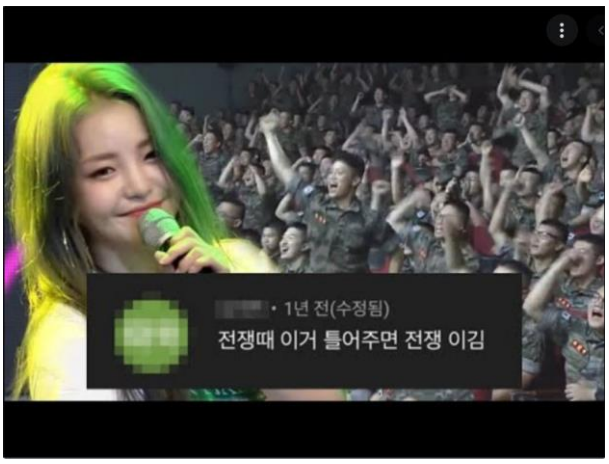
코로나 바이러스로 인해 비대면이 일상화됨에 따라 동영상 컨텐츠가 기하급수적으로 생산되는 시대에 우리는 살고 있다. 영상 스트리밍 시장의 규모도 증가함에 따라 시청자의 서비스 이용 시간도 날이 갈수록 길어지고 있습니다. 동영상 데이터 시청은 시간에 비례하는 시간이 소요된다. 이로 인해, 긴 동영상의 경우 영상을 축약한 형태의 편집된 영상이 점점 많아지고 있다. 편집된 영상은 하이라이트 영상이라고 불리며, 하이라이트 영상을 편집하기 위해 많은 시간과 노력이 필요하다. 하이라이트 추출 연구는 예로부터 많이 진행되어 왔다. 그러나 이전의 하이라이트 추출 연구들은 데이터에 의존적인 도메인 지식을 사용하는 경우가 많았다. 실제 예를 들어보면 2020 년 KBO 포스트 시즌에 맞춰 선보인 'AI 주요장면 하이라이트' 서비스가 제공되고 있지만, 철저히 운동 종류에 따라 특화되어 있을 뿐만 아니라 공급자 중심의 서비스이다. 본 연구는 이러한 도메인 지식을 이용하지 않고, 시청자의 얼굴 표정 데이터와 실시간 채팅 데이터를 통해 딥 러닝을 이용한 하이라이트 추출 모델을 제안한다.




(그림 1) 도메인 지식을 이용한 네이버 TV 서비스.

(그림 1)은 네이버 TV 에서 하이라이트라고 검색시

노출되는 데이터들을 보여준다. 긴 러닝타임이 필요한 축구, 야구 등의 동영상들이 주로 하이라이트 영상으로 업로드 되는 것을 확인할 수 있다.



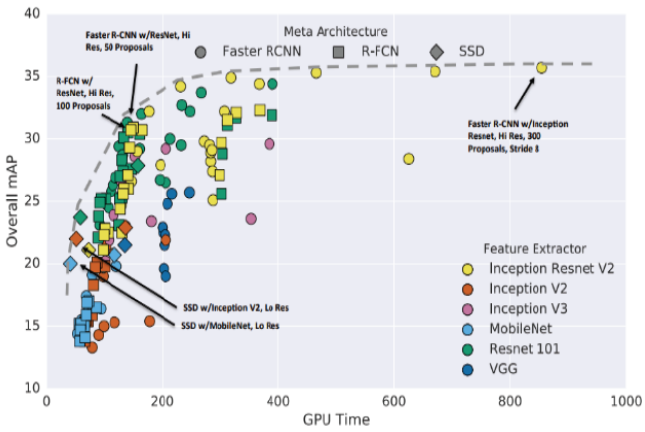

(그림 2) 시청자 얼굴 표정과 실시간 채팅.

(그림 2)는 실제 공연을 관람하는 분들의 반응과 실시간 채팅 데이터 기반의 모델링이 특정 집단에는 현실적일 수 있다는 예를 보여준다.

## 2. 시스템 설계를 위한 관련 연구

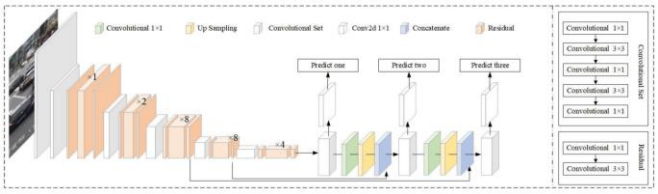
하이라이트 영역을 추출하는 방법은 여러 형태의 데이터를 사용하는 다양한 연구들이 있다.

액션 캠으로 찍은 스포츠 영상들의 이미지만을 사용하여 하이라이트를 추출하는 모델을 제안하였다[1]. 음성 데이터만을 사용하여 여러 장르의 노래 하이라이트를 추출하는 모델을 제시하였다[2]. 스트리밍 서비스의 유저 텍스트를 기반으로 하이라이트를 추출하는 모델을 제시하였다. [3]. 본 연구에서는 시청자의 얼굴 표정 데이터, 실시간 채팅 데이터를 사용한 모델을 제시한다.




(그림 3) 객체 인식 알고리즘 정확도와 시간 비교.[4]

(그림 3)는 동일한 조건에서 실험을 통해 검출 정확도 대비 처리시간을 보여준다.

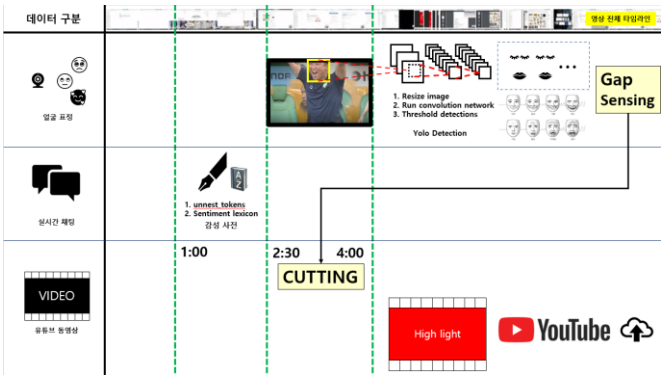


(그림 4) YOLOv3 구조.[5]

(그림 4)는 Darknet-53 백본 네트워크로 사용하며 3가지 척도 예측을 사용한다.

## 3. 제안 시스템 구성

- 시청자들은 유튜브 실시간 방송 플랫폼에 접속 후 운동 동영상을 다 같이 시청한다.
- 카메라를 통해 얼굴 표정 데이터를 실시간으로 수집하여 YoloV3 알고리즘으로 눈썹과 입술을 수집하여 눈썹과 입술의 수치가 크게 변경되는 시점을 센싱하여 수치화한다.
- 채팅 데이터는 실시간으로 수집되며, 1 분 단위로 데이터베이스에 저장 후 긍정, 부정 분석을 진행하여 퍼센트 수치화한다.
- 운동 동영상이 종료되면, 얼굴 데이터와 채팅 데이터의 Thresold 가 특정 수치 이상인 부분을 검출한다.
- 최고 Thresold 바로 전의 최저 Thresold 를 하이라이트 영상 컷팅의 시작점의 시간 데이터로 설정한다.
- 최고 Thresold 직후 평균 Thresold 이하의 수치를 하이라이트 영상 컷팅의 종료점의 시간 데이터로 설정한다.
- 파이썬 moivepy 패키지의 subclip 함수를 이용하여 하이라이트 영상 컷팅을 실행한다.
- 파이썬 selenium 패키지로 유튜브에 자동으로 하이라이트 동영상을 업로드한다.




(그림 5) 제안 시스템 구성도.

## 4. 결론 및 향후 연구

본 논문에서는 기존의 연구 방식인 한가지 데이터 종류만을 사용한 모델이 아닌, 시청자의 얼굴 표정 데이터와 실시간 채팅 데이터를 입력 값으로 사용하였다. 또한, 공급자 중심의 하이라이트 영상이 아닌 감정 상태를 인식하고, 영상의 하이라이트 영역만 컷팅하여 유튜브에 자동으로 업로드 하는 시스템을 설계

하였다.

## 참고문헌

[1] H. Yang, B. Wang, S. Lin, D. Wipf, M. Guo, B. Guo, “Unsupervised Extraction of Video Highlights Via Robust Recurrent” IEEE International Conference on Computer Vision (ICCV), P8, 2015

[2] H. J. Woo, A. Kim, C. Kim, J. Park and S. Kim “Automatic Music Highlight Extraction using Convolutional Recurrent Attention Networks” CoRR abs/1712.05901, P3, 2017

[3] H. G. Nam, “Automatic Generation of Titled Video Highlights From Mass Interaction” Masters dissertation. Chungnam National University, Daejeon, Korea. P2, 2018

[4] Jonathan Huang Vivek Rathod Chen Sun Menglong Zhu Anoop Korattikara, “Speed_accuracy trade-offs for modern convolutional object detectors”, P8, CVPR, 2017

[5] QI-CHAO MAO, HONG-MEI SUN, YAN-BO LIU, AND RUI-SHENG JIA, “Mini-YOLOv3_Real-Time_Object_Detector_for_Embedded_Applications”, IEEE Access, P3, 2019