

2020 POSTECH Global Sports AIX Conference에서는 스포츠와 관련되어 진행 된 연구들에 대한 설명을 들었습니다. 모든 주제들이 굉장히 유익했으며, 새로운 시야를 가질 수 있게끔 구성이 되어 있었습니다. 그 중에서도 가장 인상 깊었던 것은 "Listen to Look: Action Recognition by Previewing Audio"라는 연구에 대해 설명해주신 POSTECH 오탉현 교수님의 강연이었습니다. 강연의 시작은 교수님이 왜 이 연구를 진행 했는지에 관하여 동기를 설명해주셨습니다. 저도 컴퓨터 비전 분야를 연구하고 있는 사람으로서 영상이라는 매체가 얼마나 흥미로운 소재인지 잘 알고 있었습니다. 사람들은 개인적인 경험들을 남기기 위해, 혹은 정보를 기록하기 위하여 영상이라는 매체를 사용합니다. 영상은 이미지의 단계를 넘어서 역동적인 시각적 데이터를 표현할 수 있습니다. 그렇기에 영상이라는 매체는 사람들에게 더욱 직관적이고 생생하게 다가올 수 있는 것입니다. 이러한 영상이 스포츠 분야에서 왜 중요한가 했을 때 스포츠 매체에는 수많은 흥미롭고 기억에 남는 순간들이 담겨 있습니다. 사람들은 그 순간을 통해 감동과 열정을 느끼고 재미를 찾게 됩니다. 결국 스포츠와 영상의 만남은 보는 이로 하여금 수많은 감정을 표현하고 전달할 수 있습니다.

최근 SNS 시장과 영상 매체 플랫폼의 영향력이 올라가면서 인터넷 상에는 수많은 동영상 존재합니다. 이렇게 넘쳐나는 동영상들은 어떤 사람들에게 흥미로운 연구 소재가 됩니다. 동영상을 자동으로 분류하거나, 동영상 속 관심 있는 분야를 찾을 수도 있게 됩니다. 교수님이 하시는 연구는 동영상이 하나 주어지게 되면 recognition model을 디자인하여 비디오 레벨의 action class를 예측하는 것입니다. 비디오를 다루게 되었을 때 challenge 포인트들이 몇몇 존재합니다. 첫번째로는 다듬어지지 않은 분량이 긴 비디오를 분석하는 경우에는 연산 자원의 소비가 심하다는 것입니다. 이럴 때는 긴 동영상을 여러 개의 클립으로 나누어 분석을 진행하는 방법이 주로 사용이 됩니다. 두번째로는 눈 여겨볼 만한 시각적인 사건들은 종종 긴 동영상에 드문드문 나온다는 것입니다. 그래서 올바른 동작을 구분하기 위하여 적합한 구간을 찾는 것도 중요한 포인트입니다. 교수님은 이러한 문제들을 해결하기 위하여 여러 클립들 사이의 내용이 크게 유사하다는 것과 하나의 클립 내에 유사성에 대해서 관심있게 살펴보았습니다. 그리하여 이 연구의 목적은 다듬어지지 않은 긴 영상에서 클립 단위의 영상과 비디오 단위의 영상 속 효율적이고 정확한 action을 인식하는 것입니다. 이 과정 속 오디오 파일을 사용해 동적인 사건을 효율적으로 근사 시키도록 했습니다. 그리하여 이미지와 오디오 쌍으로 묶어 클립을 예측하고 프레임과 오디오가 주어지면 뉴럴 네트워크를 통해 학습을 시키고 모델을 디자인하게 됩니다.

동영상 세상 속에서 이미지와 오디오 정보는 굉장히 중요한 데이터입니다. 특히 스포츠 분야에서 영상 속 동적인 이미지와 역동적인 사운드는 팬들로 하여금 더욱 스포츠를 생생하고 즐겁게 즐길 수 있게 하는 요소입니다. 이러한 연구들이 팬들에게 더욱 즐거운 경험을 제공해주고, 더 나아가 스포츠 선수들의 멋지고 역동적인 장면들을 널리 퍼트릴 수 있는 좋은 기회입니다. 이러한 미디어 매체 연구는 앞으로도 지속되어야 하며, 팬들에게 가장 확실하게 다가올 수 있는 분야가 바로 동영상 매체라고 생각합니다.