

애니메이션 영화 식별을 위한 캐릭터 인식 방법에 관한 연구

서용석, 임동혁, 박지현

한국전자통신연구원(ETRI)

yongseok@etri.re.kr

A Study on Character Recognition Method for Identifying Animation Movies

Yongseok Seo, Donghyuck Im, Jihyun Park

Electronics and Telecommunications Research Institute

요 약

원본 영상의 핑거프린트 없이 딥러닝 기반 객체 검출 및 인식 기술을 이용하여 애니메이션 영화를 식별하는 방법을 제안한다. 1-stage 방식의 YOLOv3를 이용하여 인간형과 비인간형 캐릭터로 클래스를 단순화하여 검출 정확도를 높이고, 검출된 이미지들에 대해 Deep CNN을 통해 학습된 특징벡터를 추출하고, 일반적인 딥러닝 분류 모델에서 주로 사용하는 Softmax 함수 대신 Angular 기반의 loss 함수를 적용하여 클래스 간 분별력을 최적화한 인식모델을 구성한다. 400종의 애니메이션 캐릭터 인식이 가능한 모델을 생성하여 애니메이션 영화 100편을 대상으로 실험하였을 때 정확한 식별 결과를 보임을 확인하였다.

I. 서 론

콘텐츠 인식 기술 분야에서는 전통적으로 오디오 또는 비디오 신호의 저수준 특징인 핑거프린트를 추출하여 DB로 구축된 원신호와의 유사도 비교를 통해 콘텐츠를 식별하였다. 그러나 원본 콘텐츠를 필요로 하는 기존 식별 방법들은 콘텐츠 공개 시점과 핑거프린트 DB 구축 시점 사이에 공백이 발생하고 이로 인해 초기 불법유통 차단에 기술적 한계가 존재한다.

최근 딥러닝 기술을 이용하여 원본 콘텐츠의 핑거프린트 추출과정 없이 얼굴 인식 기술을 이용하여 영화나 드라마를 식별하는 방법이 제안되었다[1]. 그러나 애니메이션 영화를 식별하기 위해서는 실사 영상에서 배우 얼굴을 검출하고 인식하는 방법과는 다른 접근이 필요하다. 애니메이션 영상은 다양한 그림체와 과장되거나 생략된 형태의 각종 캐릭터 객체들이 등장하기 때문에 일반화된 검출 모델을 가지고 있지 않다. 특히 의인화된 비인간형 캐릭터의 경우는 인간형 캐릭터보다 훨씬 더 다양한 형태로 이루어져 있고, 이로 인하여 실사 영상에 적용되는 객체 검출 기술들이 원활하게 작동하지 않는 문제점을 가지고 있다.

최근, 이러한 애니메이션 영상의 특수성에 주목하여 모든 프레임에서 무작위로 객체를 검출하고 객체의 검출 빈도와 시각적인 중요도(visual saliency)를 비교하여 캐릭터 사진을 생성하는 방법도 제안되었다[2].

본 논문에서는 원본 콘텐츠의 핑거프린트 DB를 사용하지 않고 딥러닝 기반 인간형/비인간형 캐릭터 검출기와 검출된 캐릭터의 검증과정을 통해 애니메이션 영화를 식별하는 방법을 제안한다.

II. 제안 방법

일관성없는 형태를 보이는 애니메이션 캐릭터들을 빠르고 정확하게 검출하기 위해서는 비인간형 캐릭터만이 갖는 문제점에 대한 다각도 분석이 필요하고 인간형과 비인간형 캐릭터 모두에 적용할 수 있는 단일 검출기 개발이 필수적이다.

애니메이션 영상에서 캐릭터를 검출하기 위해 캐릭터 사진을 생성하는

기존 연구[2]에서는 전통적인 기계학습만을 이용하고 가지치기를 위해 복잡한 알고리즘을 중첩 사용한 것과 달리, 제안 방법에서는 딥러닝 기반 객체 검출기를 이용하여 ILSVRC-2014 데이터셋으로 사전학습 후 별도로 구축한 애니메이션 영화 전용 데이터셋을 이용하여 캐릭터 검출기를 미세 조정 학습하였다.

객체 검출기는 클래스에 의존적인 검출기와 클래스에 독립적인 검출기로 구분된다. 클래스 의존적인 검출기는 테두리 상자와 클래스명을 표시하는 반면 클래스 독립적인 검출기는 정확한 클래스명은 알수 없으나 객체일 가능성이 높은 전경을 추출하는 차이점이 있다. 다양한 애니메이션 캐릭터들의 형태적 특성상 모든 클래스에 대해 검출과 인식을 한번에 시도하는 방식은 높은 오인식율을 동반하는 것을 실험적으로 확인하였다. 따라서 제안하는 방법에서는 애니메이션 캐릭터 검출기를 인간형 캐릭터와 비인간형 캐릭터로 클래스를 단순화하여 검출 정확도를 높이도록 하고 검출된 캐릭터는 1:1 검증(verification) 과정을 통해 인식하도록 한다.

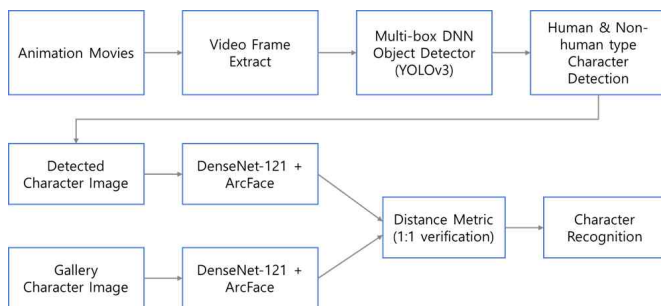


(그림 1) 인간형/비인간형 캐릭터 검출 결과

딥러닝 기반 객체 검출 방법은 후보영역(Region proposal)을 검출하는 단계와 이 후보영역을 기반으로 객체의 종류와 위치를 검출하는 2-stage

방식과 후보 영역의 학습단계없이 미리 정의된 위치의 후보박스(Prior box)를 활용하여 1-stage로 객체의 종류와 위치를 검출하는 방식으로 구분된다. 1-stage 방식의 모델은 객체 위치 후보와 분류기를 같이 학습하여 이미지 전체를 한번에 바라보는 방식으로 검출 속도가 매우 빠르며 클래스 맥락에 대한 이해도가 높은 편이다. 또한 객체에 대한 일반화된 특징을 학습하고 인지하는데 뛰어나서 Artworks나 애니메이션과 같이 형태가 모호한 데이터에 대해서도 객체 검출 성능이 뛰어나도록 확인할 수 있었다. 이러한 점들을 근거로 제안하는 방법에서는 애니메이션 캐릭터 검출기로 가장 적합한 모델로 YOLOv3[3]를 최종 선택하였다.

애니메이션 캐릭터 인식을 위해 YOLO와 같은 DNN 객체 검출기로 검출된 이미지들에 대해 Deep CNN을 통해 학습된 특징벡터를 추출하고, 일반적인 딥러닝 분류 모델에서 주로 사용하는 Softmax 함수 대신에 Angular 기반의 loss 함수를 적용하여 클래스 간 유사도 측정을 모델안에서 직접하는 Metric Learning 방식으로 캐릭터 인식모델을 구성한다. 각 샘플간의 거리(Euclidean distance)만으로 학습이 진행되는 많은 Metric learning 방식들에 비해 Angular 기반 loss 함수는 rotation과 scale에 대한 불변 속성이 보장되어 애니메이션 캐릭터와 같이 동일 클래스라도 캐릭터의 크기, 신체 비율, 꼬리/날개와 같은 다양한 특징들로 인해 feature distance가 유달리 큰 경우에 대해 더욱 우수한 성능을 기대할 수 있다. 본 논문에서는 입력 캐릭터 이미지에 대한 CNN 출력 특징에 대해 클래스 내 compactness와 클래스 간 diversity 향상을 위해 margin을 추가하는 ArcFace[4] 기법을 적용하여 캐릭터 분별력을 최적화하도록 학습한 후 특징간 거리를 1:1 검증(verification)하도록 아래 그림과 같은 애니메이션 캐릭터 인식 시스템을 구성하였다.



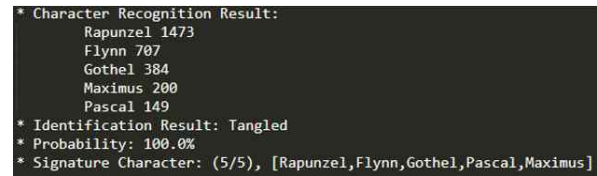
(그림 2) 애니메이션 캐릭터 인식을 위한 시스템 흐름도

III. 실험 결과

캐릭터 검출기 학습을 위해 2D, 3D 애니메이션을 포함한 50개의 애니메이션 영화로부터 수집한 3만 여장의 이미지를 이용하였다. 캐릭터 검출기로 선정된 YOLOv3는 서로 다른 3가지 scale로 객체들을 예측함으로써 기존 버전에 비해 작은 객체들에 대한 검출 성능이 향상되었다. 애니메이션 영화 전용 데이터셋으로 미세 학습한 결과 인간형에서는 91.21%, 비인간형에서는 92.37%로 평균 검출 정확도(mAP) 91.79%를 얻을 수 있었다. 캐릭터 인식 모델 학습을 위해서는 캐릭터 검출기를 이용하여 100편의 애니메이션 영화에 등장하는 156종의 인간형 캐릭터와 244종의 비인간형 캐릭터 총 400만 여장을 수집하여 사용하였다. DenseNet과 ArcFace loss를 이용하여 캐릭터 인식기를 학습한 후 각 캐릭터 당 한 장씩 뽑은 대표 이미지와 학습에 이용하지 않은 캐릭터 이미지들과 1:1 검증과정을 거쳐 가장 가까운 캐릭터로 분류하도록 실험하였다. 실험결과 총 400종 캐릭터의 평균 인식 정확도(mAP)는 91.0%를 보였다.

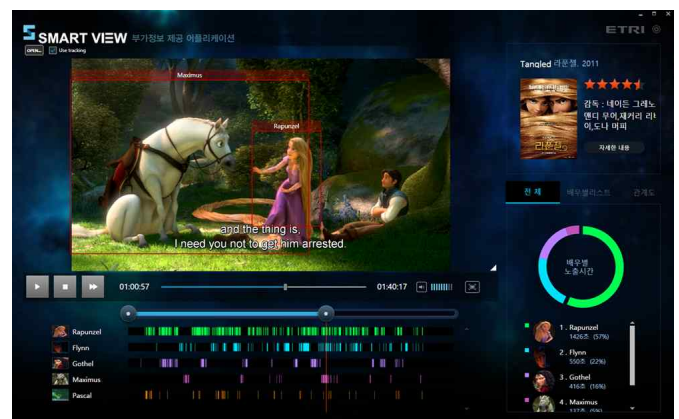
애니메이션 영화별 특징 캐릭터 집합(Signature character set)을 구성한 뒤, 인식된 캐릭터명과 빈도를 입력받아 애니메이션 영화를 최종 식별할

수 있다. 아래 예시는 저작물 영상 식별 API를 통해 “라푼젤(Tangled)” 영상에 대해 캐릭터 인식 기반 식별 테스트 결과이다.



(그림 3) 캐릭터 인식 기반 애니메이션 식별 결과

100편의 애니메이션 영화에 대해 테스트한 결과 모두 정확한 저작물 식별이 가능함을 확인할 수 있었으며, 그림 4와 같이 캐릭터 인식 과정에서 객체 트래킹을 이용하지 않고 캐릭터 인식 로그 분석만으로 캐릭터별 출현구간을 생성하여 다양한 UI/UX로 활용 가능함을 확인하였다.



(그림 4) 캐릭터 인식 및 애니메이션 영화 식별 결과

IV. 결론

본 논문에서는 원본 콘텐츠의 고유한 신호적 특징 정보인 핑거프린트를 이용하지 않고 딥러닝 기반 객체 검출 및 인식 기술을 이용하여 애니메이션 영화를 식별할 수 있는 시스템을 구현하였다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국저작권위원회의 2020년도 저작권보호 및 이용활성화 연구개발 지원사업으로 수행되었음

참고 문헌

- [1] Y. Seo, D. Im, W. Yoo, "Video identification using deep learning based face recognition." Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, 2017.
- [2] K. Somandepalli, N. Kumar, T. Guha and S. S. Narayanan, "Unsupervised Discovery of Character Dictionaries in Animation Movies," IEEE Transactions on Multimedia, 2018.
- [3] J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," Tech Report, arXiv:1804.02767, 2018.
- [4] J. Deng, J. Guo, N. Xue, S. Zafeiriou. "Arcface: Additive angular margin loss for deep face recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019.