합성곱 신경망의 장르 분류를 위한 색 공간 처리 방법

김승환, 박은수, 류은석
성균관대학교 컴퓨터교육과
whitekomani@skku.edu, espark804@skku.edu, esryu@skku.edu

Color Space Processing Method on Genre Classification via Convolutional Network

SeungHwan Kim, Eun-Soo Park, Eun-Seok Ryu
Sungkyunkwan University, Department of Computer Education

요 약

CNN 모델은 영상과 이미지를 다루는 문제에서 필수적인 요소가 되었다. CNN 모델은 데이터에 따라 적절한 색 공간이 달라지는데, 본 논문에서는 실시간으로 영화의 장르를 분류하는 CNN 모델에서 적절한 색 공간을 탐색한다. 또한, 이 결과를 바탕으로 데이터의 특성을 파악하고, 결과를 개선하는 방법을 제시한다.

1. 서 론

ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 AlexNet[1]이 이미지 분류에서 오류율을 16%로 낮추는 우수한 결과를 얻는 것을 시작으로 Convolutional Neural Network(CNN) 모델은 영상과 이미지를 다루는 최신 모델에서 필수적인 요소가 되었다.

영화의 장르를 분류하는 대부분의 최신 기법들도 CNN 모델을 포함하고 있다. 이미지를 입력으로 하는 모델 [2], 영상을 입력으로 하는 모델[3] 모두 CNN 모델을 사용하고 있다. 본 논문 또한 VGGNet 모델[4]의 gradient 소실, 네트워크의 깊이에서 발생하는 gradient 폭발 문제를 개선한 ResNet 모델[5]을 사용해 장르 인식을 구현한다.

Gowda S.N.는 [6]에서 CIFAR-10에 대해 HSV, LAB, YIQ, YPbPr, YCbCr, YUV 등의 색 공간에 대한 실험에서 LAB 색 공간의 정확도가 가장 높지만, 차이의 폭은 작다고 제시한다. 각각의 색 공간은 클래스에 따라 다른 정확도를 나타냈기에, 클래스에 따라 최적의 색 공간은 달라진다는 것을 밝혔다.

데이터에 대한 적절한 색 공간을 찾는 연구도 다양하게 진행되었다. [7]은 차량의 색상정보를 효율적으로 처리하기 위해 HSV모델과 CIELab 모델을 RGB 모델과 비교했으며, [8]은 자율주행 환경의 실시간처리를 위해 YUV 이미지를 사용하는 모델을 제안했다. 본 논문에서는 실시간 장르 인식 모델에서 적절한 색 공간을 연구한다.

2. 실험 내용

본 절에서는 기존의 데이터 셋과 본 논문에서 제작한 데이터 셋과 모델의 구조를 설명하고, 결과를 제시하는 순서로 구성된다.

2.1 데이터 셋

영화의 장르 인식에 대한 공개된 데이터 셋은 대부분 예고편 영상을 사용하는데[9], 실시간으로 장르를 인식하는 문제에는 이러한 데이터를 사용하기 부적절한 특징들이 있다.

영화 일부분의 장르는 영화 전체의 장르와 다를 수 있다. 또한, 기존 데이터에 존재하는 Action, Comedy 등의 장르는 주관적인 요소가 큰 장르로 한 장면에 대해서 평가하기 어렵다.

이 논문에서는 가장 객관적으로 알 수 있는 장르인 Horror, Animation, Sports 3가지에 대해 실험을 진행한다. 데이터 셋은 그림 1과 같은 형태로, 각 장르에 해당하는 영화 일부분의 이미지로 구성되며, 구성 비율은 표 1과 같다.

(그림 1) 구성한 데이터의 예시

(표 1) 데이터 셋의 비율

	Animation	Horror	Sports
Training	874	904	920
Validation	48	48	48

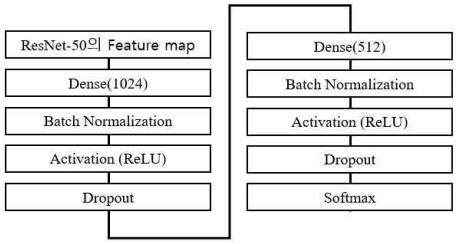
2.2 모델

모델의 구조는 ResNet-50에서 Classifier의 구조를 변경한 구조로 Fully-Connected layer가 주를 이룬다. 모델의 구조는 그림 2와 같다.

상대적으로 적은 데이터를 학습하는 과정에서 Overfiting을 방지하기 위해 [10]에서 소개한 대로 Batch Normalization (BN) 이후에 Dropout을 배치한다.

또한, 충분한 학습을 위해서 learning rate는 0.01에서 0.001까지 단계적으로 감소시키며 60 epoch 학습한다.

(그림 2) 모델의 전체 구조



2.3 색 공간별 실험 결과

표 2는 위의 데이터와 모델에 대해 각 색 공간의 학습 결과다. YCbCr 모델이 가장 좋은 결과를 보이며, HSV 모델이 가장 좋지 않았다. [7]은 CNN 모델에서 색상정보를 중심으로 처리하는 문제에는 RGB 모델이 HSV나 YUV 모델보다 유리하다고 밝혔다. 반대로 이 결과를 통해 장르 인식에서는 색상 정보보다 밝기(Y) 값, 혹은 형태 정보가 중요하다고 추측할 수 있다.

(표 2) 색 공간별 학습 결과

	RGB	HSV	LAB	YCbCr
loss	0.85	1.10	0.95	**0.61**
accuracy	**74.29**	69.31	55.32	74.01

2.4 압축된 색 공간에 관한 결과

YCbCr 모델은 주로 동영상의 인코딩 과정에서 사용된다. 대부분의 동영상 코덱은 4개의 Y값에 Cb, Cr 값 하나를 사용하는 YUV420 모델을 통해 색 정보를 압축한다. 이 과정에서 YUV420은 Y값의 비율이 더 높아진다. 표 3은 전체 데이터에서 Y값이 차지하는 비율에 따른 학습 결과를 나타낸다.

(표 3) 압축된 YUV모델의 학습 결과

	YCbCr	YUV420	Y
loss	0.61	0.72	1.056
accuracy	74.01	75.26	**78.03**
Y의 비율	33%	66%	**100%**

YUV420을 사용할 때와 Y값만을 사용했을 때 loss값이 높게 나왔지만, 정확도는 향상되었다. 이를 통해 2.3에서 추론한 내용처럼 장르 인식에서는 밝기와 형태의 데이터인 Y값이 가장 중요한 요인이라고 할 수 있다.

3. 결 론

본 논문은 특정한 색 공간이 모든 경우에 최적이 아님을 밝히고, 실시간으로 장르를 인식하는 특정 문제에 대해 최적의 색 공간은 YCbCr 모델임을 제시한다. 이를 바탕으로 색상보다 밝기와 형태가 중요한 데이터일수록 Y값의 비율을 높이면 정확도가 향상된다는 것을 밝힌다. 추후 연구에는 공개된 데이터 셋을 통해 색 공간을 선택하는 일반적인 방법을 연구할 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2020-2017-0-01630). 본 논문은 또한 한국전력공사의 2016년 선정 기초연구개발과제 연구비에 의해 지원되었음 (과제번호 : R17XA05-68).

참 고 문 헌

[1] Alex Krizhevsky and Sutskever, Ilya and Hinton, Geoffrey E “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25 (NIPS2012), pp.1097-1105, 2012.

[2] Jônatas Wehrmann, Rodrigo, C. Barros, “Movie genre classification: A multi-label approach based on convolutions through time”, Applied Soft Computing. Vol.61, pp.973-982, 2017.

[3] 강상연, 조현, 황원준, “딥러닝 기반 동영상 장면 분류 기법”, 한국통신학회 추계종합학술발표회 논문집 Vol.67 No.01, pp.311-312, 2018.

[4] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv cs.CV, 2014.

[5] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016.

[6] Gowda S.N., Yuan C. ,“ColorNet: Investigating the Importance of Color Spaces for Image Classification” Asian Conference on Computer Vision 2018 pp.581-596, 2018.

[7] Reza Fuad Rachmadi and I Ketut Eddy Purnama, “Vehicle Color Recognition using Convolutional Neural Network” arXiv, 1510.07391, 2015.

[8] Thomas Boulay and Said El-Hachimi and Mani Kumar Surisetti and Pullarao Maddu and Saranya Kandan “YUVMultiNet: Real-time YUV multi-task CNN for autonomous driving” arXiv, 1904.05673, 2019.

[9] G. S. Simões, J. Wehrmann, R. C. Barros and D. D. Ruiz, “Movie genre classification with Convolutional Neural Networks,” 2016 International Joint Conference on Neural Networks (IJCNN), pp.259-266, 2016.

[10] Xiang Li, Shuo Chen, Xiaolin Hu, Jian Yang “Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift” CVPR 2019, pp.2682-2690, 2019.