

윤곽 보강 기법을 적용한 다운 샘플된 데이터 셋 학습 신경망 모델 성능 분석

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

Performance Analysis of Neural Network Model for Learning Down-sampled Datasets Applying Edge Enhancement Technique

Eun-Soo Park, Seunghwan Kim, Eun-Seok Ryu
Department of Computer Education

요 약

딥러닝 관련 연구의 가속화로 인해 다양한 타입의 데이터 셋이 존재할 수 있다. 본 논문에서 다운 샘플된 데이터 셋을 학습한 경우 원본 데이터 셋과의 성능 비교를 한다. 추가적으로 다운 샘플된 데이터 셋을 학습할 때의 성능을 높이기 위해 제안하는 윤곽 보강 기법을 적용하여 실험한다. 제안하는 윤곽 보강 기법은 원본 데이터에서 윤곽 데이터를 추출한 후 다운 샘플 후 원본 사이즈로 업샘플링 한 데이터에 적용하여 보강한다. 본 논문에서 윤곽 좌표에 픽셀 값을 0, 255, 원본 픽셀 값을 치환하는 3가지 실험을 수행했고 결과를 분석하였다. 윤곽 픽셀 좌표에 원본 픽셀 값을 치환하는 것이 가장 성능이 높았고 다운 샘플링을 크게 적용할수록 윤곽 강화 기법의 성능이 원본 데이터 셋 학습에 비해 최대 12.2%(CNN), 2.8%(CNN-LSTM) 향상되었다.

1. 서 론

최근 고성능 GPU의 사용으로 처리 가능한 연산량이 대폭 증가함에 따라, 많은 연산이 필요한 딥 러닝 관련 기술들이 연구가 활발히 진행되어 오고 있다. 이미지 처리에 특화된 Convolution Neural Networks(CNN)의 발달로 인하여 객체 인식, 행동 인식, 이미지 캡셔닝 등과 같은 이미지 프로세싱 연구가 진행되어 오고 있다[1]. 이 중에서 행동 인식 관련 연구는 스키텐 데이터, 이미지 데이터, 시퀀셜 데이터 등 여러 방향을 통하여 활발히 진행되어 오고 있으며, 상당히 어려운 주제로 알려져 있다. 행동 인식은 영상 데이터를 입력하므로 많은 양의 데이터를 갖는다.

영상 데이터는 주파수 영역으로 분석할 때 고주파(윤곽선)와 저주파(질감)로 나눌 수 있다. 신경망 모델은 영상 학습 시 초기 레이어에서 고주파에 초점을 두고, 후기 레이어에 저주파를 초점으로 학습을 수행한다[2]. 기존 연구에도 신경망 모델의 성능 향상을 위해 윤곽 데이터를 이용한 연구가 있다. 이미지 슈퍼레졸루션에 윤곽 데이터를 이용한 연구, 담배 연기 인식을 위해 윤곽 및 텍스처 데이터를 퓨징한 연구 등이 있다[3, 4].

본 논문에서 영상의 주파수 영역이 모델 학습 시 어떠한 영향을 끼치는지에 대해 분석하기 위해 학습 데이터를 다운 샘플링하여 신경망 모델 학습 시 원본 학습 성능과 비교 및 분석한다. 또한, 고주파 영역의 정보 보강을 위해 다운 샘플된 데이터의 윤곽 보강 기법을 제안하여 다운 샘플링된 학습 데이터 셋으로 학습한 모델의 성능을 향상시킨다.

2. 윤곽 보강 기법이 적용된 데이터 셋 학습

본 장에서 윤곽 보강 기법이 적용된 데이터 셋 학습에 관해 기술한다. 1절에서 다운 샘플된 데이터 셋 학습 방법에 대해 설명한다. 2절에서 윤곽 보강 기법에 대해 설명한다. 3절에서 실험 및 결과를 분석한다.

2.1 다운 샘플된 데이터 셋 학습 방법

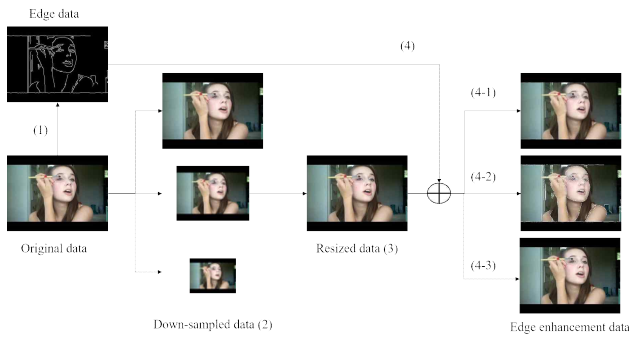
본 논문에서 행동인식 데이터 셋인 UCF-101을 사용한다 [5]. 신경망 모델은 공간 정보 측면과 시간 정보 측면 모두를 분석하기 위해 CNN과 CNN-LSTM 모델을 사용한다. 학습 데이터에 대한 다운 샘플링은 1/4, 1/16, 1/64 단계로 수행한다.

2.2 윤곽 보강 기법

본 논문에서 제안하는 윤곽 보강기법은 3가지 방법이다. 윤곽 보강 기법의 절차는 그림 1과 같다. (1) 원본 이미지에서 윤곽 검출 기법(Canny edge detection)을 적용하여 윤곽 픽셀 좌표 저장 (2) 영상 다운 샘플링(1/4, 1/16, 1/64) (3) 다운 샘플링된 영상을 원본 사이즈로 리사이징 (4) 리사이징 된 영상에 윤곽 보강 (4-1) 윤곽 픽셀 값을 0으로 치환 (4-2) 윤곽 픽셀 값을 255로 치환 (4-3) 윤곽 픽셀 값을 원본 픽셀 값으로 치환. 신경망 모델은 Inception V3를 적용했는데 Inception V3에 maxpooling연산이 적용되어 있다[6]. maxpooling은 마스크 내 가장 높은 픽셀 값을 추출하는 연산이기 때문에 윤곽 픽셀 값을 255로 입력한 실험을 수행했다. 윤곽 픽셀 값을 0으로 치환한 이유는 기존 영상에서 윤곽선의 값은 일반적으로 0값에 가깝기 때문에 0으로 치환한 실험을 수행했다.

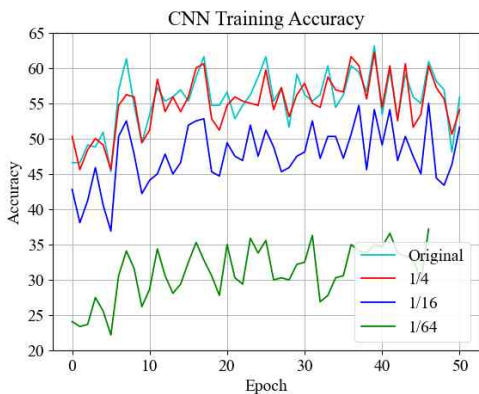
2.3 실험 및 결과 분석

실험은 앞서 언급한 CNN(Inception V3)과 CNN-LSTM 모델로 학습 수행했고 하이퍼파라미터는 다음과 같다. Augmentation은 normalization, horizontal flip과 random crop을 적용했다[7]. learning rate는 초기 0.5로 설정하고 스케줄링 기법을 적용했다. 스케줄링 기법은 5 에폭마다 learning rate를 1/2하였다. 배치 사이즈는 32, 인풋 사이즈는 299x299로 설정했다. CNN-LSTM의 경우 시퀀스 길이를 20



(그림 1) 윤곽 보강 기법 개념도

프레임으로 설정했다. 실험 결과는 테스트 셋의 정확도를 평가 지표로 설정하였다. CNN 모델의 경우 표 1, CNN-LSTM의 경우 표 2와 같다.



(그림 2) CNN 학습 정확도

(표 1) 다운 샘플링 데이터 셋과 윤곽 보강 기법 성능 실험 결과(CNN) (단위: %)

	Original	Edge 0	Edge 255	Edge original
Original	56.6	50.0	52.2	-
1/4	54.7	55.0	51.6	52.5
1/16	49.4	50.6	51.6	50.0
1/64	35.3	47.5	46.9	41.6

(표 2) 다운 샘플링 데이터 셋과 윤곽 보강 기법 성능 실험 결과(CNN-LSTM) (단위: %)

	Original	Edge 0	Edge 255	Edge original
Original	74.5	54.9	59.9	-
1/4	73.3	63.6	64.4	72.5
1/16	67.7	55.5	54.2	70.4
1/64	54.5	24.4	29.3	57.3

그림 2와 같이 학습 시 테스트 정확도 결과를 볼 때 원본과 1/4는 정확도의 차이가 크지 않았고 1/16부터 원본 데이터 셋 학습에 비해 정확도가 현저히 떨어졌다. 윤곽 보강 기법의 결과를 볼 때 CNN 모델에서 다운 샘플링을 크게 적용할수록 윤곽 보강 기법을 적용한 성능이 더 높아지는 것을 볼 수 있었다. 그러나 CNN-LSTM 모델에서 윤곽 좌표 픽셀에 0, 255 값을 입력할 경우 성능이 더 낮아지는 경향이 있었다. 이는

UCF-101 데이터 셋 특성상 원본 데이터 셋에 블록현상이 있는데 블록현상으로 인해 윤곽 검출 기법 성능이 낮아져서 윤곽 보강 기법의 결과가 노이즈로 작용된 것으로 볼 수 있다. 윤곽 좌표 픽셀 값에 원본 픽셀 값을 입력했을 때 1/16 다운 샘플링 데이터부터 원본 데이터보다 높은 성능을 보였다.

3. 결론 및 향후 연구

본 논문에서 다운 샘플링 데이터 셋으로 학습할 때 윤곽 보강 기법을 적용하여 성능을 높인 실험을 수행했다. 실험 결과로 윤곽 좌표에 0 또는 255 값을 치환하는 것은 오히려 노이즈 현상을 일으켜 성능이 낮아지는 것을 확인할 수 있었고 원본 픽셀 값을 치환하는 것이 안정적으로 성능이 향상되는 것이 보였다. 본 논문의 한계점은 UCF-101 데이터 셋으로만 실험을 수행했다는 점이다. 추후연구로 KTH 데이터 셋, HMDB51과 같은 추가 데이터 셋으로 실험을 수행해 더욱 일반적인 추론을 할 수 있다.

Acknowledgement

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019R1A2A2C1010476). "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT 연구센터육성지원사업의 연구결과로 수행되었음"(IITP-2020-2017-0-01630)

참고 문헌

- [1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
- [2] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.
- [3] Yang, W., Feng, J., Yang, J., Zhao, F., Liu, J., Guo, Z., & Yan, S. (2017). Deep edge guided recurrent residual learning for image super-resolution. IEEE Transactions on Image Processing, 26(12), 5895-5907.
- [4] Yuan, F., Li, G., Xia, X., Lei, B., & Shi, J. (2019). Fusing texture, edge and line features for smoke recognition. IET Image Processing, 13(14), 2805-2812.
- [5] Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv preprint arXiv:1212.0402.
- [6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
- [7] Park, Eun-Soo, Kim, SeungHwan, Ghulam Mujtaba, Ryu, Eun-Seok, "Analysis of Action Recognition Model Performance Differences According to Preprocessing Methods", Summer Workshop on Computer Communications (SWCC), Aug. 27, 2020.