논문 19-44-07-09 The Journal of Korean Institute of Communications and Information Sciences '19-07 Vol.44 No.07
https://doi.org/10.7840/kics.2019.44.7.1299

영상 리사이징이 심층 신경망 기반 영상 분류기 성능에 미치는 영향에 관한 연구

김 윤 형◆, 정 찬 호°, 김 창 익*

Impacts of Image Resizing on the Performance of Deep Neural Network-Based Image Classifiers

Yoonhyung Kim◆, Chanho Jung°, Changick Kim*

요 약

본 논문에서는 영상 리사이징이 심층 신경망 기반 영상 분류기의 성능에 미치는 영향을 분석 및 고찰한다. 고정된 크기의 영상을 입력으로 받아들이는 심층 신경망 기반 영상 분류기를 사용하기 위해서는 입력 영상을 리사이징하는 작업이 선행되어야 한다. 본 논문에서는 5종의 영상 리사이징 기법을 이용하여, 각각의 리사이징 기법이 영상 분류 성능에 미치는 영향을 실험적으로 비교 및 분석한다. 정량적 비교 평가를 위해 ImageNet 영상 데이터셋으로 학습된 5종의 영상 분류기를 활용하여 Top-5, Top-1 정확도를 측정하였다. 본 논문에서 제시한 정량적 분석 결과는 심층 신경망 기반 영상 분류기 활용에 관심이 있는 연구자 및 개발자들에게 유용한 벤치마크가 될 것으로 예상된다.

Key Words : Image resizing, Deep neural network, Visual saliency

ABSTRACT

In this letter, we investigate the impact of image resizing on the performance of deep neural network-based image classifiers. Since most deep neural network-based image classifiers require a fixed-size input dimension, an input image needs to be resized before testing. In this letter, we use five image resizing operators to empirically investigate the impact of each resizing operator on the performance of the image classifiers. For quantitative evaluation, we report Top-5 and Top-1 accuracies of five image classifiers trained by the ImageNet dataset. We believe that this study serves as a practically useful benchmark for researchers and practitioners interested in utilizing deep neural network-based image classifiers.

Ⅰ. 서 론

영상 분류 (Image classification) 과제는 단일 영상을 입력받아 영상에 존재하는 객체 또는 물체의 종류를 인식하는 것으로써, 최근 몇 년간 심층 신경망 기술의 발달과 고성능 GPU의 보급에 힘입어 그 성능이 비약적으로 발전하였다. 1000개의 레이블을 가지는 영상들로 이루어진 ImageNet 영상 데이터셋[1]은 영상 분류 분야의 대표적인 벤치마크 데이터셋이며, 이를 이용한 영상 분류 성능 시합에서 AlexNet[2], GoogLeNet[3], ResNet[4] 등이 각 해의 가장 높은 정확도를 보이는 영상 분류기로 인정받았다. 이들은 모두 심층 신경망을 활용하여 설계된 영상 분류기로써, 입력으로 고정된 크기 (예: $224\times224\times3$)의 영상을 받아들인다. 이 때문에 테스트하고자 하는 영상이 영상 분류기의 규격에 맞지 않을 경우 영상 리사이징을 수행해야 하는데, 이 과정에서 필연적으로 영상 정보의 부분적 손실 또는 왜곡을 초래하여 인식 성능의 저하를 유발한다. 해당 논문들에서는 이러한 단점을 극복하기 위해 원 영상을 잘라내기 기법 등을 통해 여러 개의 패치를 생성(예: 10개[2,4], 144개[3] 등)한 뒤 여러 번의 테스트를 수행한 후 유사도의 평균을 산출하는 방식을 적용하였다. 하지만 이와 같은 방식은 한 장의 영상을 테스트하기 위해 영상 분류기를 반복적으로

◆ First Author : (ORCID:0000-0002-5608-8473)School of Electrical Engineering, Korea Advanced Institute of Science and Technology, yhkim1127@kaist.ac.kr, 학생회원
° Corresponding Author : (ORCID:0000-0003-3145-6732)Department of Electrical Engineering, Hanbat National University, peterjung@hanbat.ac.kr, 정회원
* (ORCID:0000-0001-9323-8488)School of Electrical Engineering, Korea Advanced Institute of Science and Technology, changick@kaist.ac.kr
논문번호 : 201904-045-A-LU, Received April 9 2019; Revised May 1, 2019; Accepted May 3, 2019

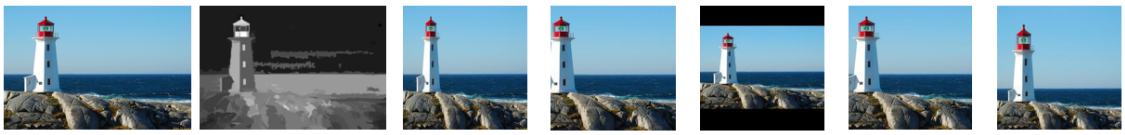

그림 1. 영상 리사이징의 예. (a) 원본 영상, (b) 영상 중요도 지도, (c) 스케일링, (d) 중앙 잘라내기, (e) 레터박싱, (f) 내용 기반 잘라내기, (g) 내용 기반 영상 와핑[5].
Fig. 1. Examples of image resizing. (a) Original image, (b) image importance map, (c) uniform scaling, (d) center cropping, (e) letterboxing, (f) content-based cropping, (g) content-based image warping[5].

실행해야 하므로 효율성이 크게 저하된다.

본 논문에서는 기 학습된 영상 분류기를 실제 활용할 때의 동작 효율성에 주목하여, 한 장의 입력 영상에 대해 한 차례의 테스트만으로 가장 높은 인식 성능을 도출하는 리사이징 기법에 대해서 실험적인 방법으로 알아본다. 이를 위해 잘라내기, 스케일링, 레터박싱, 내용 기반 잘라내기, 내용 기반 와핑 등 5종의 기법으로 리사이징한 경우에 대한 인식 결과를 비교 및 분석한다.

Ⅱ. 실험 방법

실험에 사용된 영상 분류기는 원 논문의 학습 가이드라인에 따라 ImageNet 데이터셋으로 학습된 AlexNet[2], GoogLeNet[3], ResNet (18, 50, 101 layers)[4]을 이용하였다. 입력 영상 리사이징 기법은 스케일링(Uniform Scaling, US), 중앙 잘라내기 (Center Cropping, CC), 레터박싱 (Letterboxing, LB), 내용 기반 잘라내기 (Content-based Cropping, CBC), 내용 기반 영상 와핑 (Axis-aligned Deformation, AAD[5])을 적용하였다. 내용 기반 접근법은 관심 영역 검출 기법[6,7]으로 영상 중요도 지도를 생성하고, 이를 이용해 리사이징 시 시각적 관심도가 큰 영역은 보존하고 상대적으로 덜 중요한 영역의 변형 또는 삭제를 허용함으로써 리사이징 이후 전반적인 시각적 품질 저하를 경감시키는 기법이다. 본 논문에서는 내용 기반 접근법에 활용될 영상 중요도 지도 생성을 위해 Cheng의 기법[6]을 사용하였다. 상기한 5종의 리사이징 적용 예를 그림 1에 도시하였다. 영상 분류 성능 측정을 위해 ImageNet 검증 (Validation) 데이터셋 5만 장을 이용하여 Top-5, Top-1 정확도를 측정하였다. Top-5 정확도는 영상 분류기가 출력한 1000개의 레이블 유사도에서 상위 5개 이내에 정답 레이블이 있을 경우 올바로 인식한 것으로 판정하는 정확도 측정 방식이며, Top-1 정확도는 가장 유사도가 높은 레이블이 정답일 경우 올바로 인식한 것으로 판정하는 정확도 측정 방식이다.

Ⅲ. 실험 결과 및 분석

표 1은 5종의 영상 분류기 각각에 대해 5종의 영상 리사이징 방법을 적용하는 총 25가지 경우에 대한 인식 정확도를 나타낸다. 전반적인 수치로 보았을 때, 심층 신경망 기반 영상 분류기는 입력 영상을 스케일링과 잘라내기 방법으로 리사이징한 경우에 높은 정확도를 보였다. 스케일링과 잘라내기 기법의 정확도를

표 1. 영상 리사이징 기법에 따른 심층 신경망 기반 영상 분류기의 Top-5 (Top-1) 평균 정확도 (단위: %)
Table 1. Top-5 (Top-1) average accuracies of deep neural network-based image classifiers according to image resizing methods (In percentage)

Image classifier / Resizing method	AlexNet[2]	GoogLeNet[3]	ResNet-18[4]	ResNet-50[4]	ResNet-101[4]
Uniform Scaling (US)	77.92 (54.39)	87.18 (66.01)	86.02 (65.21)	88.55 (68.94)	89.47 (70.44)
Center Cropping (CC)	77.45 (53.57)	86.99 (65.60)	86.91 (66.47)	90.05 (71.75)	91.02 (72.95)
Letterboxing (LB)	71.24 (46.50)	84.42 (61.75)	83.52 (61.48)	86.15 (65.37)	88.01 (68.03)
Content-based Cropping (CBC)	76.58 (52.75)	86.71 (65.15)	86.66 (66.02)	89.93 (71.38)	90.90 (72.86)
Axis-aligned Deformation (AAD[5])	72.71 (48.20)	82.93 (60.74)	82.10 (60.70)	85.43 (65.54)	87.92 (69.63)

(a) (b) (c) (d)

그림 2. 각 리사이징 기법 기준, 타 리사이징 기법에 비해 강인한 성능을 보이는 영상의 예 (ResNet-101). (a) 스케일링, (b) 레터박싱, (c) 내용 기반 잘라내기, (d) 내용 기반 와핑.
Fig. 2. Examples of images that are more robustly recognized by each resizing operator (ResNet-101). (a) Uniform scaling, (b) letterboxing, (c) content-based cropping, and (d) content-based image warping[5].

비교해보면, 영상 분류기의 인식 성능이 높을수록 잘라내기 기법이 스케일링보다 더 나은 결과를 도출함을 확인하였다. 이는 영상 분류기의 성능이 높아질수록 부분적으로 삭제된 영상 정보로부터 객체의 종류를 추론하는 능력이 향상되기 때문으로 여길 수 있다. 반대로 AlexNet과 같은 비교적 낮은 성능의 영상 분류기는 잘라내기보다 스케일링을 적용한 경우에 더 높은 정확도를 도출하였다. 레터박싱 기법은 모든 영상 분류기에 대해서 대체로 낮은 성능을 도출하였다. 이를 통해 영상 분류기 입력에 무의미한 영상 정보를 삽입하는 것은 인식 성능에 부정적인 영향을 미침을 알 수 있다.

내용 기반 리사이징 기법은 관심 영역 검출 기법으로 생성된 영상 중요도 지도를 활용하므로, 그림 1(f), (g)와 같이 타 기법들에 비해 시각적으로 두드러지는 영역을 잘 보존하는 것을 확인할 수 있다. 그러나 표 1과 같이, 내용 기반 리사이징 기법들은 전통적 기법들에 비해 대체로 낮은 정확도를 도출하였다. 이는 입력 영상에서 인식해야 하는 객체의 시각적 관심도가 타 영역의 시각적 관심도보다 항상 높지는 않음을 암시한다. 내용 기반 와핑 기법은 내용 기반 잘라내기에 비해 낮은 정확도를 도출하였는데, 이를 통해 입력 영상을 비선형적으로 변형하는 것이 영상 분류기의 인식 성능에 부정적인 영향을 미침을 알 수 있다.

입력 영상의 특성에 따른 최적의 리사이징 기법을 찾기 위해 정성적 분석을 진행하였다. 그림 2는 각 리사이징 기법을 기준으로 타 리사이징 기법에 비해 강인한 성능을 도출하는 입력 영상들을 보여준다. 예를 들어, 그림 2(a)의 영상은 스케일링을 적용했을 시에는 영상 분류기가 인식에 성공하지만 타 리사이징 기법들을 적용했을 시에는 인식하지 못하는 영상의 예이다. 스케일링의 경우, 영상 내 객체의 원형이 가로-세로비 변형에 의해 크게 훼손되지 않는 경우에 강인한 양상을 보였다. 레터박싱의 경우, 영상 내 객체가 구조적 요소가 많음과 동시에 입력 영상의 영역 전반에 걸쳐 형성된 경우에 강인한 양상을 보였다. 내용 기반 잘라내기의 경우, 영상 내 객체의 위치가 특정 구역에 편중되어 있는 경우에 잘 인식되는 양상을 보였다. 내용 기반 와핑의 경우, 배경이 단조로운 영상에 대해서 강인한 경향을 보였다.

실험에 사용된 영상 리사이징 기법이 입력 영상의 특성에 따라 미치는 영향을 이해하기 위해 정성적인 분석을 진행하였다.

Ⅳ. 결 론

본 논문에서는 영상 리사이징이 심층 신경망 기반 영상 분류기의 인식 성능에 미치는 영향에 대해서 정량적으로 분석 및 고찰하였다. ImageNet 영상 데이터셋을 이용하여, 총 25가지 경우의 영상 분류 실험을 통해 각각의 리사이징 기법이 주는 영향을 분석하였다. 또한, 입력 영상의 특성에 따른 최적의 리사이징 기법에 대해서도 사례를 통해 정성적으로 분석하였다. 본 논문에서 제시한 분석 결과는 심층 신경망 기반 영상 분류기를 효율적으로 활용하고자 하는 연구자 및 개발자들에게 유용한 벤치마크가 될 것으로 예상된다.

References

[1] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition 2009*, pp. 248-255, Miami, USA, Jun. 2009.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. Adv. in Neural Inf. Process. Syst. 2012*, pp. 1097-1105, Lake Tahoe, USA, Dec. 2012.

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, V. Vanhoucke, and A.

Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition 2015*, pp. 1-9, Boston, USA, Jun. 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 770-778, Las Vegas, USA, Jun. 2016.

[5] D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," *Computer Graphics Forum*, vol. 31, no. 2, pp. 229-236, Wiley Online Library, May 2012.

[6] M. M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569-582, Mar. 2015.

[7] H.-S. Kim, H.-M. Kim, J.-M. Seo, and C.-S. Jeong, "Region of interest extraction using saliency map of vairous resolution," in *Proc. IEIE Summer Conf. 2015*, pp. 757-760, Jeju, Korea, Jun. 2015.