

RGB-csb를 활용한 제한된 CNN에서의 정확도 분석 및 비교

공준배* · 장민석** · 남광우*** · 이연식***

Accuracy Analysis and Comparison in Limited CNN using RGB-csb

Jun-Bea Kong* · Min-Seok Jang** · Kwang-Woo Nam*** · Yon-Sik Lee***

요약

본 논문은 대부분의 변형된 CNN(: Convolution Neural Networks)에서 사용하지 않는 첫 번째 컨볼루션 층(convolution layer)을 사용해 정확도 향상을 노리는 방법을 소개한다. GoogLeNet, DenseNet과 같은 CNN에서 첫 번째 컨볼루션 층에서는 기존방식(3x3 컨볼루션연산 및 배규정규화, 활성화함수)만을 사용하는데 이 부분을 RGB-csb(: RGB channel separation block)로 대체한다. 이를 통해 RGB값을 특징 맵에 적용시켜 정확성을 향상시킬 수 있는 선행연구 결과에 추가적으로, 기존 CNN과 제한된 영상 개수를 사용하여 정확도를 비교한다. 본 논문에서 제안한 방법은 영상의 개수가 적을수록 학습 정확도 편차가 커 불안정하지만 기존 CNN에 비해 정확도가 평균적으로 높음을 알 수 있다. 영상의 개수가 적을수록 평균적으로 약 2.3% 높은 정확도를 보였으나 정확도 편차는 5% 정도로 크게 나타났다. 반대로 영상의 개수가 많아질수록 기존 CNN과의 평균 정확도의 차이는 약 1%로 줄어들고, 각 학습 결과의 정확도 편차 또한 줄어든다.

ABSTRACT

This paper introduces a method for improving accuracy using the first convolution layer, which is not used in most modified CNN(: Convolution Neural Networks). In CNN, such as GoogLeNet and DenseNet, the first convolution layer uses only the traditional methods(3x3 convolutional computation, batch normalization, and activation functions), replacing this with RGB-csb. In addition to the results of preceding studies that can improve accuracy by applying RGB values to feature maps, the accuracy is compared with existing CNN using a limited number of images. The method proposed in this paper shows that the smaller the number of images, the greater the learning accuracy deviation, the more unstable, but the higher the accuracy on average compared to the existing CNN. As the number of images increases, the difference in accuracy between the existing CNN and the proposed method decreases, and the proposed method does not seem to have a significant effect.

키워드

Accuracy, Convolution Neural Network, Deep Learning, Kernel, Layer, Learning Time
정확도, CNN, 딥러닝, 커널, 레이어, 학습 시간

* 군산대학교 석사과정(rhdwnsqo12@kunsan.ac.kr)

** 교신저자 : 군산대학교 컴퓨터정보통신공학부

*** 군산대학교 컴퓨터정보통신공학부

(kwnam@kunsan.ac.kr, yslee@kunsan.ac.kr)

• 접수일 : 2019. 10. 29

• 수정완료일 : 2019. 12. 22

• 게재확정일 : 2020. 02. 15

• Received : Oct. 29, 2019, Revised : Dec. 22, 2019, Accepted : Feb. 15, 2020

• Corresponding Author : Min-Seok Jang

School of Computer & Information & Communication Engineering, College of Engineering, Kunsan National University,

Email : msjang@kunsan.ac.kr

I. 서론

영상처리에 있어 영상 인식 분야는 예전부터 꾸준히 연구되었다. 특히 최근 딥러닝을 적용한 CNN의 영상인식의 등장으로 가장 활발히 연구되고 있다. CNN은 GPU를 이용한 영상처리 연구가 활발하게 이뤄지고 있으며[1], 추가적으로 컴퓨터의 성능 향상, 특히 그래픽 카드의 발전에 따라 더욱 더 심오한 구조의 CNN들이 제안되고 있다. GoogLeNet, ResNet, DenseNet, Mobile-Net은 앞서 말한 연구의 대표적인 CNN구조들이다. CNN의 구조적인 연구 외에 CNN의 과적합 및 연산량을 줄이기 위한 연구들도 진행되고 있으며[2, 3] CNN의 깊이가 증가될 경우 발생하는 vanishing gradient 문제를 해결하기 위한 방법도 제안되었다[4]. 딥러닝의 영상 인식 분야는 이와 같이 다양한 방법으로 정확도를 향상시키고 있으며 특히 변형된 CNN구조들이 많이 나오며 좋은 성능을 보이고 있다. 하지만 앞서 서술한 CNN을 포함한 대부분의 CNN구조는 CNN의 첫 번째 컨볼루션 층에서 그들이 제안한 구조를 적용하지 않고 컨볼루션 연산 및 정규화, 활성화함수만을 수행한다.

본 논문에서는 첫 번째 컨볼루션 층을 변형해 정확도를 높일 수 있는 방법을 제안하고, 이를 기반으로 영상 수에 따른 제한된 영상에서의 정확도 분석을 통해 본 논문에서 제안한 방법과 CNN구조 사이의 정확도를 분석 및 비교한다.

II. CNN 기법 제안

2.1 관련 연구

CNN은 딥러닝의 신경망 중 하나이며 DNN(Deep Neural Network)의 한 종류이다. DNN은 기본적으로 하나의 입력층과 여러 은닉층, 하나의 출력층으로 구성되어 있으며 CNN도 동일하다[5]. 영상인식 분야에서 주로 CNN을 사용하며 TensorFlow, Caffe, Torch, Theano등을 통해 구현한다[6].

GoogLeNet은 Inception이란 모델을 통해 입력영상을 다양한 방법으로 컨볼루션 연산을 수행하고 합치는 것을 반복한다[7]. Inception은 다양한 버전이 존재하지만 버전들은 각각 1x1, 3x3, 5x5 컨볼루션 연산을 통해 나온 결과 값들을 concatenation 연산을 통해 이어준

다. 추가로 3x3, 5x5컨볼루션 연산 전에 1x1컨볼루션을 통해 연산할 입력 영상의 크기를 줄이기도 한다. 다소 복잡한 해당 방법은 첫 번째 컨볼루션 층에서는 사용하지 않고 일반적인 컨볼루션 연산과 LRN(Local Response Normalization)과 같은 정규화를 수행한다.

ResNet은 깊은 층의 부작용인 vanishing gradient를 해결하기 위해 skip connection을 추가하였으며 이를 통해 residual network를 설계해 1000계층 이상의 깊은 CNN 층 구조를 선보였다[8]. skip connection을 통해 기존의 신경망 학습 방법과 다른 Residual학습을 유도했고 이를 통해 깊은 층의 구조를 선보였으나 GoogLeNet과 동일하게 첫 번째 층에서는 7x7 컨볼루션 연산만 수행할 뿐 skip connection은 추가되지 않았다.

DenseNet은 Dense Block을 통해 Dense connectivity를 제안했으며 이를 통해 gradient vanishing 문제 해결 및 낮은 층의 신경망으로도 높은 정확도를 보여주었다. DenseNet은 입력 영상 값을 출력 값의 채널 뒤에 합쳐 쌓아 올리는 방식이며 이를 통해 초기 입력 값의 특징 맵을 마지막 출력 값까지 가져갈 수 있는 장점이 있다 [9]. 그러나 앞서 언급했던 신경망과 동일하게 첫 레이어에서는 Dense Block을 적용하지 않는다.

세 가지 기술들 모두 정확도 향상을 위해 다양한 방법들을 추가하였으나 이러한 기술들은 학습효과나 성능저하 문제로 인해 CNN의 첫 번째 컨볼루션 층에서는 그들이 제안한 구조를 사용하지 않는다. 본 논문에서는 첫 번째 컨볼루션 층을 RGB-csb로 대체하여 다른 CNN들의 정확도 성능을 향상시킬 방법을 제안하며 실험을 통해 기존CNN과 비교한다.

2.2 RGB-channel separation block

본 논문에서는 RGB-csb를 첫 번째 컨볼루션 층에 대체하여 사용한다. 입력된 컬러 영상은 기본적으로 RGB의 3색 정보가 포함되어 있다. 본 논문에서는 영상의 특징을 초반부분에서 강조하기 위해 입력 영상의 RGB 정보를 특징 맵에 각각 더하여 사용한다. 입력 영상은 16개의 필터를 거치며 각 필터는 3x3 컨볼루션 연산을 수행해 특징을 더욱 더 견고하게 나타내도록 유도한다. CNN을 통과하는 하나의 RGB 영상 X를 가정한다. 영상X는 224 x 224 크기이며 X_r , X_g , X_b 의 3가지 색 정보를 포함하고 있다. 영상 X가 일반적인 컨볼루션 연산만 수행하는 첫 번째 컨볼루션

층을 통과하게 하는 대신, 첫 번째 컨볼루션 층을 RGB-csb로 대체한다. 다음 그림 1은 RGB-csb구조이다. C.L(: Convolution Layer), SUM, C.R(: Channel Reduction)으로 구성되며 R, G, B에 따라 총 3번 C.L, SUM, C.R이 반복되는 구조이다. 각 C.L은 모두 16개의 3x3 컨볼루션 연산을 수행한다[10].

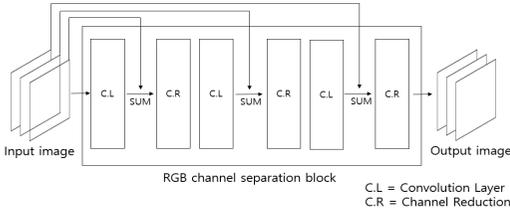


그림 1. RGB-csb 구조
Fig. 1 Structure of RGB-csb

그림 1의 RGB-csb에서 C.L, SUM, C.R에 대해 자세한 내부구조 설명은 그림 2와 같다.

Convolution Layer	3x3 Convolution
	Batch Normalization
	Rectified Linear Unit
SUM	Combine monochromatic images X_r , X_g , and X_b in order on all channels in the feature map
Channel Reduction	1x1 Convolution
	Batch Normalization

그림 2. RGB-csb 내부 구조
Fig. 2 Inner structure of RGB-csb

그림 1과 그림 2를 토대로 RGB-csb의 내부 학습 과정은 다음과 같이 진행된다. 1) 영상 X 를 16개의 작은 수의 필터를 사용해 3x3 컨볼루션 연산을 하고, 배치정규화와 활성화함수인 ReLU를 차례대로 적용하여 $224 \times 224 \times 16$ 의 특징 맵을 얻는다. 2) 앞서 얻은 특징 맵에서 16개 채널들에 대해 각각 SUM연산을 통하여 특징 맵에 X_r 을 더한다. (순차적으로 X_g 와 X_b 에도 동일하게 적용) 3) 16개의 채널을 3개의 채널

로 줄이기 위해 1x1 컨볼루션 연산 및 배치정규화가 포함된 CR(Channel Reduction)을 수행한다.

위의 과정을 총 3번 반복하게 되며 2번에서 첫 번째에서 X_r 을 두, 세 번째에선 X_g , X_b 의 영상을 이용해 더해준다. 해당 학습과정을 통한 RGB-csb는 모두 같은 크기의 결과물이 도출되며, RGB-csb를 모두 통과할 시 최종 결과물의 채널 개수는 3개이다.

2.3 실험 데이터 및 구조

본 논문에서 사용할 실험 데이터는 인터넷에서 검색한 동물 영상데이터이며 약 5천장의 5종류로 실험을 진행하였다. 실험 환경에 사용한 컴퓨터 환경은 다음과 같다.

표 1. 실험 환경
Table 1. Experiment environment

OS	windows 10 pro 64 bit
CPU	i7 - 7700
RAM	16G
GPU	NVIDIA TITAN Xp

영상데이터는 표 2와 같이 데이터를 나눠서 실험을 진행하였다. 표 2에서 제안된 모델명 뒤의 숫자는 카테고리 별 제한된 훈련 및 시험 영상 개수를 나타내며 960의 경우 5종류의 영상을 각각 960장, 총 4,800장으로 학습을 진행한 것이며 다른 숫자를 가진 실험 모델의 경우 최소 800장의 영상을 사용한다.

표 2. 실험 모델 분류 표
Table 2. Experiment model classification table

Model	explanation
CNN-160	using 160 images by each category
rgbCNN-160	
CNN-320	using 320 images by each category
rgbCNN-320	
CNN-640	using 640 images by each category
rgbCNN-640	
CNN-960	using 960 images by each category
rgbCNN-960	

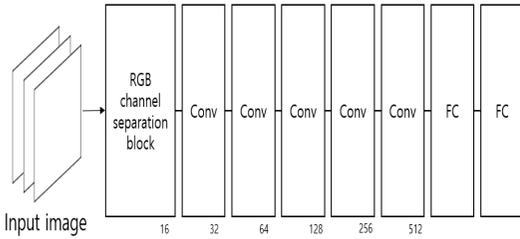


그림 3. 실험에 사용된 CNN 구조
Fig. 3 CNN structure used in experiments

표 2와 같은 방법으로 제한 영상에 대한 모델을 분류한다. 내려갈수록 증가되는 개수는 그 전 모델에서 사용한 제한된 영상을 포함한다. 표 2에서 CNN은 기존 CNN구조를 동일하게 사용한 모델을 뜻하며, 그림 3에서 RGB-csb 대신 일반적인 컨볼루션 층이 추가된 형태이다. 제한된 영상에서의 정확도 성능 측정을 위해 다음 그림 3과 같은 구조를 갖는 신경망을 rgbCNN이라 부르고 rgbCNN은 본 논문에서 제안한 RGB-csb를 포함한 구조를 사용하는 것을 나타낸 것으로 그림 3과 동일하다. 두 가지 모델 모두 6개의 컨볼루션과 2개의 완전연결층(fully connected layer)으로 구성된다.

III. 실험

실험에 사용한 모든 신경망의 하이퍼파라미터는 동일하며, 입력 영상은 224 x 224 픽셀이다. batch size는 32로 설정해 한 번에 32개의 이미지씩 학습하도록 하였으며 모든 신경망은 40 에폭 만큼 훈련시켰다. 또한 초기 시작 학습 속도는 0.01로 설정했으며 10에폭

마다 5로 나눠 학습 속도를 변경하였으며, 6개의 컨볼루션 층과 2개의 완전연결층(fully connected layer)을 사용하였다. 6개의 컨볼루션 층에는 비교대상에 따라 RGB-csb를 포함하거나 포함하지 않는다. 학습은 비교모델마다 정확도의 신뢰도를 확보하기 위해 각각 10번씩 진행해 비교하였다.

표 2는 제안한 모델들이 정확도에 주는 영향을 세부적으로 파악하기 위해 실험했던 데이터이며 정확도 성능을 비교 분석한 결과, 몇 가지 의미 있는 결과들이 도출되었다. 영상의 개수가 많아질수록 평균 정확도가 높게 측정되었으나, 정확도의 증가폭이 영상의 개수가 적은 경우에 비해 작아지므로 영상 수와 정확도의 상관관계가 로그함수로 유도됨을 보였다. 또한 rgbCNN-160에서 CNN-160보다 약 5%이상 크게 정확도 차이가 날 때도 있는데, 이는 rgbCNN이 변형된 특정 영상을 학습에 추가해서 사용하기 때문에 data augmentation과 유사한 구조를 보이기 때문인 것으로 보인다. 즉 더 많은 데이터로 학습을 수행했기 때문이다.

10번 수행된 학습의 평균치를 보면 rgbCNN-160의 경우 CNN-160보다 약 2.3% 높은 정확도가 나왔다. RGB-csb를 첫 레이어로 입력된 영상들은 첫 입력 영상의 특징이 RGB색깔 별로 더욱 강조되어 추가되기 때문에 비록 편차에 의한 성능 저하의 원인이 될 수 있으나 평균적으로 봤을 때 RGB-csb를 적용할 경우 더 나은 성능을 보임을 알 수 있다. 나머지 rgbCNN-320, 640, 960의 경우 영상의 수가 많아질수록 각각의 학습 정확도 결과의 편차는 줄어들고, 기존 CNN과의 정확도 성능 차이도 줄어들어 비슷한 성능을 보인다.

표 3. 실험 정확도
Table 3. Test Accuracy

model	Test Accuracy (10 times)										avg.
CNN-160	61.88	62.50	61.87	60.00	58.75	60.00	59.38	61.25	58.52	61.25	60.54
rgbCNN-160	58.12	67.50	61.25	62.50	65.63	65.00	63.75	60.50	61.88	62.50	62.86
CNN-320	70.00	70.94	72.81	71.25	72.50	72.81	71.88	71.50	70.00	72.19	71.59
rgbCNN-320	74.06	74.06	73.13	74.37	70.00	72.81	72.25	75.62	71.87	73.13	73.13
CNN-640	79.53	78.91	81.09	81.41	83.28	82.97	82.66	81.56	80.78	81.41	81.36
rgbCNN-640	80.38	83.31	83.72	80.81	82.34	82.81	82.19	80.94	82.97	83.44	82.29
CNN-960	83.65	83.23	83.31	83.33	82.81	83.02	82.60	81.98	82.08	81.98	82.80
rgbCNN-960	84.27	84.15	83.02	84.73	84.06	83.32	83.40	83.08	82.46	83.88	83.64

IV. 결론

본 논문에서는 RGB-csb를 첫 번째 컨볼루션 층 대신 사용해 입력 영상의 RGB값을 각 특징 맵에 추가적으로 입력해 정확도를 높이는 방법을 제안했다. 또한 이러한 방법을 기반으로 영상 수에 따른 정확도 평균 및 비교분석을 실시하였다.

rgbCNN모델은 기존의 CNN 모델보다 영상의 개수가 적을수록 평균적으로 약 2.3% 높은 정확도를 보였으나 정확도 편차는 5% 정도로 크게 나타났다. 반대로 영상의 개수가 많아질수록 기존 CNN과의 평균 정확도의 차이는 약 1%로 줄어들고, 각 학습 결과의 정확도 편차 또한 줄어든다. 즉, 제안 방법에서 훈련 및 시험 영상 개수가 증가할수록 기존 CNN모델에 비하여 정확도의 차이가 적어지면서 로그함수의 상관관계를 나타냄을 알 수 있다. 이는 RGB-csb가 입력 영상의 개수를 어느 정도 늘려주는 효과로 볼 수 있으며 이는 data augmentation과 비슷한 효과를 보이며 실제로 data augmentation에서 색반전등을 통해 데이터를 늘리기도 한다. 본 논문에서 제안한 RGB-csb는 RGB 3색 영상을 모두 활용해 커널 결과 값에 추가적인 연산을 수행하는데 이를 통해 영상의 특징 및 정보를 늘려 적은 데이터일수록 좋은 효율을 보이는 것으로 판단된다.

결론적으로 정확도가 기존 CNN에 비해 적은 영상의 개수에서 각각 평균적으로 향상된 것은 RGB-clb의 RGB채널 값에 따른 SUM 연산이 기존 Convolution layer 보다 정확도 측면에서 긍정적인 영향을 준 결과로 판단된다. 실험결과에서 볼 수 있듯 RGB-csb를 통해 첫 번째 Convolution layer를 변형시켜 대부분의 응용 CNN구조에서 정확도 향상이 될 것으로 보인다.

감사의 글

이 논문은 2018년 한국국토정보공사 공간정보연구원 산학협력 R&D 지원사업(자유과제)의 지원과 2018년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2018R1D1A1B07051045)

References

- [1] J. Park, "Comparison Speed of Pedestrian Detection with Parallel Processing Graphic Processor and General Purpose Processor," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 10, no. 2, June 2015, pp. 239-245.
- [2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol.15(1), June 2014, pp. 1929-1958.
- [3] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning*, Lille, France, 2015.
- [4] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines," *International Conference on Machine Learning*, Haifa, Israel, 2010.
- [5] S. Chung and Y. Chung "Comparison of Audio Event Detection Performance using DNN," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 13, no. 3, June 2018, pp. 571-578.
- [6] Y. Lee and P. Moon, "A Comparison and Analysis of Deep Learning Framework," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 12, no. 1, 2017, pp. 115-122.
- [7] C. Szegedy, W. Lium, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Computer Vision and Pattern Recognition*, Boston, U.S.A, 2015.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Computer Vision and Pattern Recognition*, Las Vegas, U.S.A, 2016.
- [9] G. Huang, Z. Liu, K. Weinberger, and L. Maaten, "Densely connected convolutional

networks," *Computer Vision and Pattern Recognition*, Hawaii, U.S.A, 2017.

- [10] J. Kong, Y. Lee, and M. Jang, "Convolution Neural Network with RGB channel separation block," *Proc. of MITA2019*, Hochiminh, Vietnam, 2019.

저자 소개



공준배(Jun-Bea Kong)

2017년 군산대학교 컴퓨터정보공학과 졸업(공학사)

2018년 군산대학교 대학원 컴퓨터정보공학과 재학(공학석사과정)

※ 관심분야 : 딥러닝



장민석(Min-Seok Jang)

1989년 연세대학교 전자공학과 졸업(공학사)

1991년 연세대학교 대학원 전자공학과 졸업(공학석사)

1997년 연세대학교 대학원 전자공학과 졸업(공학박사)

1997년 ~ 현재 군산대학교 컴퓨터정보통신공학부 교수

※ 관심분야 : IoT 시스템, 드론 응용, 딥러닝



남광우(Kwang-Woo Nam)

1995년 충북대학교 전자계산학과 졸업(이학사)

1997년 충북대학교 대학원 전자계산학과 졸업(이학석사)

2001년 충북대학교 대학원 전자계산학과 졸업(이학박사)

2004년~ 현재 군산대학교 컴퓨터정보통신공학부 교수

※ 관심분야 : GeoAI, 빅데이터, 공간 DBMS 및 데이터마이닝



이연식(Yon-Sik Lee)

1982년 전남대학교 전자계산학과 졸업(이학사)

1984년 전남대학교 대학원 전자계산학과 졸업(이학석사)

1994년 전북대학교 대학원 전자공학과 전산응용공학 전공 졸업(공학박사)

~ 현재 군산대학교 컴퓨터정보통신공학부 교수

※ 관심분야 : 에이전트 시스템, S/W 미들웨어