

이미지 패치 연결 방식을 통한 데이터 증강 기술의 경향성에 관한 연구

(A Study on Tendencies of Data Augmentation
using the Image Patching Method)

구 승 연, 노 시 동, 정 기 석*

한양대학교

(Seung-Yeon Koo, Si-Dong Roh, Ki-Seok Chung)

(Hanyang Univ., Seoul)

Abstract : The convolutional neural networks (CNNs) for image classification tasks are getting larger and deeper. While CNNs have good representation power using numerous parameters, overfitting may occur if the size of the training dataset is insufficient. Therefore, various data augmentation techniques such as rotation, flipping, and cropping have been used to prevent overfitting. Recently, novel data augmentation techniques such as CutMix and RICAP achieved state-of-the-art performance by combining multiple images into one image. Even though he performance is known to depend on the number of images for composition, such dependency has not been sufficiently studied. In this study, we discuss tendencies of the data augmentation technique using the image patching method. We experimented validation error rates for various image patches and patch counts. Experiments are conducted using two convolution-based networks: ResNet, WideResNet on the CIFAR-10 and mini-ImageNet datasets.

Keywords : data augmentation, RICAP, regularization, label smoothing, image patch

1. 서 론

Residual neural network (ResNet)이 괄목할만한 성능을 보여준 이후, 제안되는 많은 Convolutional Neural Network (CNN)들은 이전에 비해 매우 크고 깊어졌다[1,2]. CNN이 수많은 파라미터를 사용하여 훌륭한 표현력을 가지는 반면, 학습 과정에서 네트워크의 크기에 비해 활용할 수 있는 학습 데이터 셋의 크기가 충분하지 못한 경우 overfitting 문제가 발생할 수 있다. 이를 해결하기 위한 한 방법으로 회전, 반전, 자르기 등의 비교적 단순한 변환으로 데이터를 증강시켜 성능을 향상시킬 수 있다.

*Corresponding Author (kchung@hanyang.ac.kr)

구승연, 노시동, 정기석: 한양대학교

※ 이 논문은 2021년도 정부 (과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.2021-0-00131, 제조검사장비 경량화를 위한 지능형 엣지컴퓨팅 반도체 개발)을 받아 수행된 연구임.

Overfitting을 해결하기 위한 방법으로 Dropout, Cutout [3] 등 네트워크의 일부를 누락시키는 방법 또한 널리 사용되고 있다. 그 중에서 Cutout은 이미지의 일부를 제거하여 데이터를 증강시킬 수 있는 방법이다. 반면, 두 개의 이미지를 혼합하는 방식으로 Mixup [4]이 제안되었으며, 이후 Cutout 및 Mixup 방식에서 파생된 다양한 데이터 증강 방식이 제안되었다. CutMix [5]와 RICAP [6]은 이미지 혼합 증강 방식의 하나로, 발표될 당시에 각각 CIFAR-100 분류 문제와 CIFAR-10 분류 문제에서 state-of-the-art 성능을 달성함으로써, 이미지의 일부만 사용하더라도, 여러 이미지의 정보를 이용해 데이터 증강을 하는 것이 강건한 모델을 생성한다는 것을 보여주었다.

본 논문에서는 RICAP과 같이 이미지 패치를 사용한 데이터 증강 방법이 다수의 이미지를 사용하는 점에 착안하여, RICAP 알고리즘을 변형하여 입력 이미지에 사용될 이미지 패치의 수를 다양화하고 이에 따른 모델의 성능의 변화 양상을 실험하였

다. CIFAR-10, mini-ImageNet 데이터셋을 사용하여 ResNet 및 WideResNet으로 실험하였으며, 이미지 패치 연결 방식의 데이터 증강 기술에서 유의미한 경향성이 있음을 밝힌다.

II. 관련 연구

1. 이미지 패치 방식의 데이터 증강

Cutout[3]은 이미지에서 무작위로 정해진 위치의 직사각형 영역을 0으로 채운 상태를 입력으로 사용하는 데이터 증강 기술이며, 구성요소를 누락시킴으로써 regularization을 가능하게 한다는 점에서 Dropout과 비슷하여 Regional Dropout이라고 불린다. 또 다른 중요한 데이터 증강 기술로 Mixup[4]이 있다. Mixup은 두 이미지를 보간법을 사용하여 섞고, $\alpha : 1 - \alpha$ 의 비율로 soft label을 부여하는 방식으로 데이터를 증강하는 기술이다. 이 두 가지 데이터 증강 기법의 특징적인 아이디어에서 영향을 받은 것이 CutMix와 RICAP이다. 본 연구에서 언급하는 ‘이미지 패치 방식의 데이터 증강 기술’이란, CutMix와 RICAP과 같이 여러 장의 원본 이미지에서 일부분을 떼어내고, 이들을 이어 붙여 새로운 입력을 만들어내는 방식의 데이터 증강 기법을 의미한다.

CutMix는 Cutout된 영역을 다른 이미지로 채우는 알고리즘이다 [5]. 예를 들어 개의 사진에서 개의 얼굴 부분이 지워졌다면 그 자리에 고양이의 얼굴 사진이 들어가며 개와 고양이를 혼합한 soft label로 학습되는 것이다. RICAP은 서로 다른 4장의 이미지에서 일부분을 가져와서 격자모양으로 합쳐지며 새로운 입력이 된다 [6]. Label은 각 이미지의 면적 비율로 정해지는데 자세한 이미지 연결 방식과 labeling 방식에 대해서는 3장에서 다룬다. 이미지 패치 방식의 데이터 증강 기술은 데이터 증강 전후의 이미지 크기가 같다는 특징이 있다.

2. Label Smoothing

일반적인 분류 문제에서는 하나의 data에 하나의 class만 one hot vector의 형태로 labeling이 된다. 이 경우 이미 지정되어있는 label에 대한 신뢰도가 지나치게 높기 때문에 class 매칭이 잘못 되어있는 noisy 데이터 셋에서는 성능이 떨어질 수 있다. 이러한 문제를 해결하기 위하여 1과 0으로 labeling하는 대신 label을 실수 값으로 분산시켜 사용하는 것이 label smoothing이다 [7]. 예를 들어, 고양이, 개, 개구리, 황소를 분류하는 문제에서

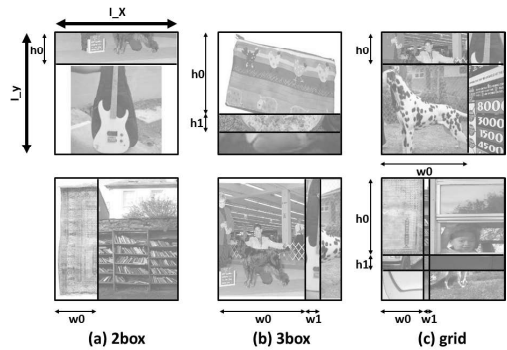


그림 1. 5가지 종류의 변형된 RICAP에 대한 개요이다.
 Fig 1. Overview of 5 types of modified RICAP.

고양이 이미지에 대한 label은 [1,0,0,0]인 반면, label smoothing을 거친 경우는 smoothing parameter가 0.4일 때 [0.7, 0.1, 0.1, 0.1]으로 labeling 된다. 앞의 예시와 같이 실수 값을 갖는 label을 soft label이라고 한다.

III. 변형된 RICAP의 구현

하나의 입력을 만들기 위해서 사용하는 이미지 데이터의 개수가 많아짐에 따라 성능이 변화하는 경향성을 관찰하기 위해 그림 1과 같이 RICAP에 사용되는 이미지 패치의 개수가 달라지도록 변형된 RICAP을 실험에 사용한다. 각각의 패치 결합 방식은 그림 1(a) 2개의 서브이미지를 수평선을 기준으로 결합시킨 2boxhor (위), 2개의 서브이미지를 수직선을 기준으로 결합시킨 2boxver (아래), 그림 1(b) 3개의 서브이미지를 수평선을 기준으로 결합시킨 3boxhor (위), 3개의 서브이미지를 수직선을 기준으로 결합시킨 3boxver (아래), 그림 1(c) 4개의 서브이미지를 격자 모양으로 결합시킨 4grid(위)가 있으며 이 경우는 [6]이 제안한 RICAP과 동일하다. 그리고 9개의 서브이미지를 격자 모양으로 결합시킨 9grid (아래)가 있다.

그림 1에서 I_x 와 I_y 는 모델이 입력으로 받아들이는 이미지의 너비와 높이를 의미한다. CIFAR-10에서는 $I_x = I_y = 32$ 이고 mini-ImageNet에서는 256이다. h_0 는 첫 번째 서브이미지의 높이이고, 2분할 체계에서 두 번째 서브이미지의 높이는 $I_y - h_0$ 로 정해진다. 3분할 체계에서 두 번째 서브이미지의 높이는 h_1 로 정의되며, $I_y - h_0$ 만큼의 길이 안에서 무작위로 정해진다. h_0 역시 I_y 만큼의

길이 안에서 랜덤으로 정해지는데, 이때 랜덤 함수는 그림 2와 같이 α 와 β 가 동일한 베타분포를 사용하며, 이때 β 는 0.3과 2를 사용한다.

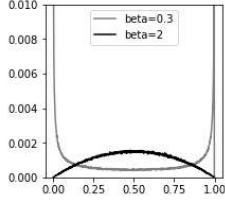


그림 2. 베타분포의 확률밀도 그래프
Fig 2. The graph of the Beta distribution

w 을 결정하는 방법은 h 와 유사하다.

$$\begin{aligned} h_0 &\sim I_Y B(\beta, \beta), & w_0 &\sim I_X B(\beta, \beta) \\ h_1 &\sim (I_Y - h_0) B(\beta, \beta), & w_1 &\sim (I_X - w_0) B(\beta, \beta) \end{aligned} \quad (1)$$

그림 3을 통해 서로 다른 이미지에서 잘린 서브 이미지가 어떻게 결합되는지 알 수 있다. x 와 y 는 0과 $I_X - roi_w$, $I_Y - roi_h$ 사이에서 무작위로 결정되며, 이때는 균등분포를 사용한다.(식 2)

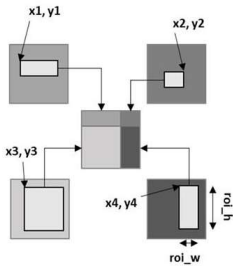


그림 3. RICAP의 결합 방식에 대한 개요
Fig 3. Overview of the RICAP patching method.

$$y_i \sim U(0, I_Y - roi_h), \quad x_i \sim U(0, I_X - roi_w) \quad (2)$$

Label은 모두 soft label로 변환이 되며 이때 label weight c 는 식 3과 같이 입력에서 차지하는 면적의 비율이 결정된다.

$$c_i = \frac{roi_x roi_y}{I_X I_Y} \quad (3)$$

IV. 실험 결과 및 해석

본 연구의 성능 비교 실험에서는 ResNet-50 모델 과 WideResNet28-10 모델을 사용했다. CIFAR-10 이미지는 비교적 단순하기 때문에 2/3 분할의 경우 수평연결만 고려했고, mini-ImageNet 은 ResNet-18 모델로 실험하였다.

1. mini-ImageNet

아래의 표 1는 Cutout의 mask size를 16으로, CutMix과 Mixup의 α 를 1.0으로, 그리고 RICAP의 β 를 0.3으로 80 epochs 만큼 학습시킨 결과이다. 표 1과 그림 4에서 입력 당 이미지의 개수가 많아 질수록 validation error는 감소하는 것을 확인할 수 있다. 특히 9분할을 한 경우 baseline에 비해 4.4677% 향상되었으며, 입력 당 이미지의 개수가 많아질수록 일반화 성능이 좋아지는 경향이 있다는 것을 확인할 수 있다.

표 1. mini-ImageNet에서 $\beta=0.3$ 일 때 평균 오차율이다. (%)

Table 1. Result of the average error rate with $\beta= 0.3$ on mini-ImageNet. (%)

	#img/input	val err.
baseline	1	18.02458
cutout	1	17.87625
cutmix	2	16.50292
mixup	2	16.03104
ricap2boxhor	2	14.6992
ricap2boxver	2	14.30979
ricap3boxhor	3	14.42333
ricap3boxver	3	14.23271
ricap4boxgrid	4	13.58458
ricap9boxgrid	9	13.55688

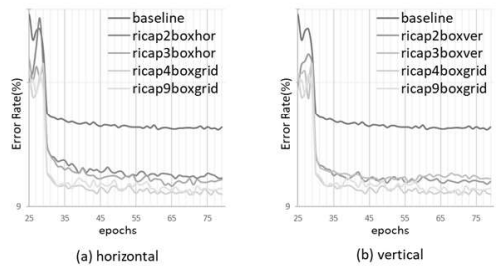


그림 4. mini-ImageNet에서의 변형된 RICAP 결과
Fig 4. Result of the modified RICAP test on mini-ImageNet.

2. CIFAR-10

데이터셋 mini-ImageNet의 결과와는 상반되게도 CIFAR-10에서는 입력에 포함된 이미지가 많아질수록 성능이 하락하는 경향을 보였다. CIFAR-10 데이터 셋은 이미지의 크기가 32x32로 매우 작기 때문에 $\beta=0.3$ 인 베타분포로 경계를 정하면 roi의 넓이가 0인 경우가 발생할 수 있다. 표 2의 zero rate는 크기가 0인 서브이미지의 비율이다. 즉 표 2에서 2분할 실험의 경우, 실제로는 약 32%가 2분할되지 않고 원본 이미지로 학습된다. 3/4분할 실험은 과반수가 2분할 또는 원본 이미지로 학습됨을 유의하여야 한다. $\beta=2$ 인 베타분포를 사용하면 w 와 h 는 높은 확률로 대상 길이의 0.5배가 된다. 따라서 추가로 β 가 2인 환경에서 ResNet-50에서 실험하였다.

표 2. CIFAR-10에 대해 $\beta=0.3$ 인 환경에서 200 epochs로 학습한 평균 오차율 결과(%)

Table 2. Result of the average error rate in 200 epochs with $\beta=0.3$ on CIFAR-10. (%)

	#img/input	val err.	zero rate
baseline	1	0.8675	*
2box	2	8.7482	31.896
3box	3	8.5801	64.949
4boxgrid	4	9.1361	53.879

다음 표 3은 β 를 2로 조정했을 때 error rate를 관찰한 결과이다. 그 결과 2분할에서 3/4분할로 포함된 이미지가 많아질수록 validation error rate는 각각 4.52%, 11.67% 만큼 증가하였다.

표 3. ResNet-50에 대해 $\beta=2$ 에서 300 epochs로 CIFAR-10을 학습한 평균 오차율 결과 (%)

Table 3. Result of the average error rate in 300 epochs with ResNet-50, $\beta=2$ on CIFAR-10.(%)

	#img/input	val err.
baseline	1	28.82
2box	2	27.87
3box	3	32.39
4boxgrid	4	39.54

V. 결론

본 논문에서는 이미지 패치 방식의 데이터 증강 기법에서 패치 개수 및 방식에 따른 성능의 경향성에 대해서 연구하였다. 비교적 큰 입력 이미지에 대해서는 이미지 패치의 개수가 증가할수록 성능이 향상됨을 보였다. mini-ImageNet의 분류 성능은 9

분할로 실험한 결과는 기존 RICAP 보다 0.03% 향상된 성능을 보였고, 2분할 수평결합보다는 1.14% 향상되었다. 반면, CIFAR-10과 같이 이미지의 크기가 작은 경우에는 포함된 패치 수가 증가할수록 성능이 최대 11.67% 하락하는 경향을 보였다. 이미지 패치 방식의 데이터 증강은 매우 간단하고 대부분의 task에 적용할 수 있다. 따라서 본 논문에서 밝힌 경향성을 토대로 이미지 패치 방식의 데이터 증강을 적용할 때 입력에 포함된 패치 개수를 적절히 조절한다면 보다 다양한 task에 대해 성능을 향상시킬 수 있을 것으로 기대한다.

References

- [1] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 770-778, 2016.
- [2] S. Zagoruyko and N. Komodakis. "Wide Residual Networks". In Proceedings of the British Machine Vision Conference (BMVC), pp. 87.1-87.12. BMVA Press, 2016.
- [3] T. DeVries, and G.W. Taylor. "Improved regularization of convolutional neural networks with cutout." arXiv:1708.04552, 2017.
- [4] H. Zhang, M. Cisse, Y.N. Dauphin, & D. Lopez-Paz, "mixup: Beyond empirical risk minimization". arXiv:1710.09412, 2017.
- [5] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, & Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features". In Proceedings of the IEEE/CVF international conference on computer vision, pp. 6023-6032, 2019.
- [6] R. Takahashi, T. Matsubara and K. Uehara, "Data Augmentation Using Random Image Cropping and Patching for Deep CNNs," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 9, pp. 2917-2931, 2020.
- [7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 2016.