DBpia

논문 21-46-06-19 The Journal of Korean Institute of Communications and Information Sciences '21-06 Vol.46 No.06
https://doi.org/10.7840/kics.2021.46.6.1087

이미지 분류 네트워크에서의 효율적 훈련 기법

배 은 지◆, 이 성 진°

Efficient Training Methodology in an Image Classification Network

Eunjee Bae◆, Sungjin Lee°

요 약

최근 영상인식 기술의 성능적 발전은 수많은 데이터 축적과 딥러닝 네트워크의 심층화에 기인하여 왔다. 하지만, 이런 다양한 데이터들을 딥러닝 네트워크에 훈련시키는 것은 다양한 문제들을 유발시킨다. 적은 데이터양에서 기인하는 오버피팅, 클래스 간의 데이터양 차이에서 오는 클래스 불균형 (Imbalance), 멀티클래스 훈련 문제 등이 그것이며 본 논문은 이런 소규모 데이터 셋들에서 발생하는 이런 문제들을 발견하고 분석하였으며 해결책을 제시하고 실험으로 성능을 분석해 보았다. 이를 위해 오픈된 소규모 데이터 셋들을 비교 분석해보고 이들 간의 차이점과 각 기술들에 좋은 성능을 내는 훈련기법들을 선정 분석해 보았으며 소규모 데이터 셋에서 서로 공통적이면서 일반적으로 사용될 수 있는 훈련기법들을 상호 교차 실험을 통해 선정 검증하였다.

키워드 : 이미지 분류, 과적합, 클래스 불균형, 멀티클래스, 훈련기법
Key Words : Image Classification, Over-Fitting, Class Imbalance, Multi-Class, Training Scheme

ABSTRACT

The recent advance of image recognition technology comes from the accumulation of numerous data and deepening of neural network. However, training these various data on a deep neural network causes various problems. Overfitting caused by a small amount of data, class imbalance resulting from the difference in the amount of data between classes, and multi-class training problems. This paper found and analyzed these problems occurring in such small data sets, and suggested solutions and analyzed the performance through experiments. For these goals, we compared open small data sets and the differences between them and selected the training techniques that perform well for each dataset.

Ⅰ. 서 론

최근 컴퓨터 기술의 급속한 발전으로 인해 기존의 텍스트 위주의 사용자 환경에서 벗어나 이미지, 그래픽, 오디오 및 비디오 데이터 등을 제공하는 멀티미디어 사용자 환경으로 변화하고 있다.최근 들어, IT산업뿐 아니라 대부분의 산업 분야에서 딥러닝 기반 영상인식 기술이 주목받고 있으며, 이러한 기술들은 ImageNet, Open Images Dataset, MS COCO 와 같은 수많은 데이터 셋 구축[1-3] 과 강력한 GPU 기반 컴퓨팅 능력, 딥러닝 네트워크의 발전[4-6,22]에 기반하고 있다.

※ 이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임 (No. NRF-2019R1F1A1062878)
◆ First Author : Dong-Seoul University Department of Electric Engineering, oe_oej@yonsei.ac.kr
° Corresponding Author : Dong-Seoul University Department of Electric Engineering, sungjinlee@du.ac.kr, 정회원
논문번호 : KICS202103-044-B-RE, Received Feburary 28, 2021; Revised March 14, 2021; Accepted March 15, 2021

www.dbpia.co.kr

하지만 실제로 이런 딥 뉴럴 네트워크를 커스텀 데이터 셋에서 훈련시켜 문헌에서 제공된 성능만큼을 얻어내기란 쉽지 않다. 보통 문헌에서 제공되는 성능은 큰 데이터 셋에서 다양한 클래스를 분류해 내는 챌린지, 예를 들면 ILSVRC (ImageNet Large Scale Visual Recognition Competition)에 우수한 컴퓨팅 자원에서 최적화된 매우 큰 심층 네트워크를 사용하지만, 실제 산업에서 사용되는 데이터 셋과 분류 클래스는 그에 비해 매우 작고 클래스 간 데이터 불균형 문제를 포함하고 있게 되어 데이터에 적합한 네트워크 훈련에 실패하게 된다. 이런 실패들은 데이터 과적합 (Overfitting) 문제, 클래스 불균형 (Class Imbalance) 문제들에서 기인한다.

본 논문은 이런 소규모 데이터 셋들에서 발생하는 이런 문제들을 실험적으로 발견하고 분석하였으며 해결책을 제시하고 실험으로 성능을 분석해 보았다. 이를 위해 오픈된 소규모 데이터 셋들을 비교 분석해보았고 이들 간의 차이점과 각 기술들에 최고성능을 내는 훈련기법들을 분석해 보았으며 소규모 데이터 셋에서 서로 공통적이면서 일반적으로 사용될 수 있는 훈련기법들을 상호 교차 실험을 통해 선정 검증해 보았다.

이를 위해 소규모 데이터 셋 훈련에 많이 적용되는 다양한 데이터 증식 기법들, MixUp[7], Cutout[8], CutMix[9]를 중심으로 실험으로 검증하며 분석한다.

데이터 셋으로는 ImageNet[1], Cifar10[10], Cifar100[10], Pascal VOC[11], Tiny ImageNet[12], CINIC-10[13], STL-10[14], WebVision-1000[15], Clothing 1M[16], iNaturalist[17], Small NORB[18], Food-10N[19], MURA[20] 등을 비교하였으며 이들 중 실험으로 사용할 소규모 데이터 셋 으로 Pascal VOC, Cifar100 그리고 클래스 당 랜덤하게 40장을 뽑아서 사용하는 reduced Cifar 100를 대상으로 성능 분석을 진행하였다.

우선 Pascal VOC 데이터 셋은 ImageNet 데이터 셋에 비해 클래스 별 이미지 양이 적고 클래스 간 이미지 양의 차이가 크며 한 이미지 내에 여러 클래스가 포함되어 있는 경우가 많기 때문에 훈련시키기 까다로운 편이여서 실제 산업 현장에서 사용하는 데이터와 유사도가 높은 편이다. 이에 본 논문은 이런 Pascal VOC 데이터 셋에 대해 분석하고 해당 데이터 셋을 훈련시키는데 발생되는 어려운 점들, 오버피팅, 클래스 불균형, 멀티 클래스을 분석하였다. 그 후 각 문제점들을 위한 해결책들을 제시하고 이를 실험적으로 검증하였다.

반면 Cifar100 데이터 셋은 50K 가 훈련 이미지로 제공되어 이미지 수가 비교적 많은 편이지만 이를 기존에 제공되는 데이터 50K 중에 4K 로만 사용하여 (클래스 당 40장) 소규모 데이터 셋으로 만들어 사용하여 기존 대규모 Cifar100 와 소규모 Reduced Cifar100 를 비교 실험하며 데이터 증식 이외에 데이터 보충이 성능에 얼마나 많은 영향을 미치는지 간접적으로 분석해보았다.

Ⅱ. 문제점 분석

2.1 데이터 셋 분석

표 1은 일반적으로 많이 사용되는 오픈된 데이터 셋들의 특징을 나타내었다. 참고로 이미지 형상 분류가 아닌 글자 분류 기술 (e.g. OCR) 을 대상으로 하는 SVHN, MNIST와 확장 데이터 셋들 (EMNIST-Digit, EMNIST-letter, Noisy MNIST)은 대상에서 제외 하였다.

본 논문에서는 위의 데이터 셋들 중에 실험 데이터 셋으로 Pascal VOC 와 Cifar100, 이 Cifar100 의 50K 이미지 중 일부 4K 이미지를 사용하는 Reduced Cifar100를 선택하고자 한다. 그 이유로 분류 대상으로 일반적인 대상을 우선하는 것이 좀 더 범용 이미지 분류 네트워크의 성능을 알아보는데 도움이 될 것이라 생각하였고 또한 Pascal VOC 가 위 데이터 셋 들 중에 클래스 간 불균형이 가장 심하여 훈련시키기가 가장 까다로운 데이터 셋 이여서 우선적으로 선정하

표 1. 오픈 데이터 셋들의 특징들
Table 1. Features of open datasets.

	클래스 수	총 이미지 수	데이터 부류
ImageNet	1000	1200 K	일반
CIFAR-10	10	50 K	일반
CIFAR-100	100	50 K	일반
PASCAL VOC	20	10 K	일반
Tiny ImageNet	10	90 K	일반
CINIC-10	10	90 K	일반
STL-10	10	5 K	일반
WebVision	5000	1600 K	일반
Clothing 1M	14	49 K	옷
iNaturalist	8000	448 K	동식물
Small NORB	5	1.65 K	3D 일반
Food-10N	101	75.75 K	음식
MURA	7	39.9 K	x-ray 신체부위

www.dbpia.co.kr

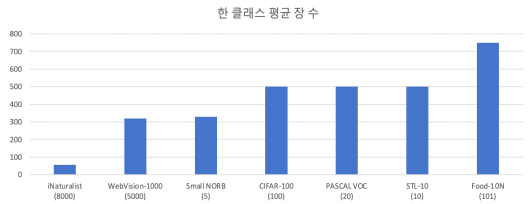

그림 1. 소중 규모 이미지 분류 데이터 셋
Fig. 1. Small-to-Medium Scale dataset for Image Classification

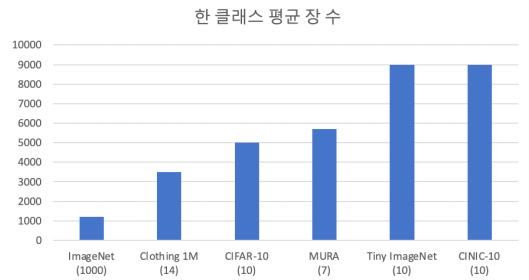

그림 2. 대규모 이미지 분류 데이터 셋
Fig. 2. Large-Scale dataset for Image Classification

였다. 반대로 클래스 개수가 많으면서도 데이터 셋이 상당히 잘 만들어진 Cifar100을 기준으로 일부러 손상을 가하는 Reduced Cifar100 데이터 셋을 만듦으로써 데이터 셋의 품질이 이미지 분류 정확도에 미치는 영향도를 분석하기 위해 Cifar100 를 선택하였다.

Pascal VOC 는 사람, 동물, 탈것, 실내와 관련된 20개의 클래스를 가지고 있는 데이터 셋이다. 그러나 이 데이터 셋을 그대로 이용하여 학습하면 좋은 성능을 얻기 힘들다. 우선 Pascal VOC 데이터 셋을 살펴본다.

2.1.1 과적합 문제 (Over-Fitting)

그림 3에서 보듯이 Pascal VOC의 데이터 셋은 특정 클래스의 경우 이미지 개수가 매우 적은 것을 확인할 수 있다. 이런 데이터 부족은 해당 클래스에 Overfitting 문제를 유발시켜 train과 validation accuracy 의 차이를 발생 시킨다. 이 문제를 해결하기 위해 해당 클래스의 이미지가 최소 250여장이 되도록 추가적인 데이터 약 2500여장 보충을 수행하였다. 또한 해당 데이터 들은 기존 이미지와는 겹치지 않게 새로운 데이터로 크롤링하여 보충하였다. 그림 4는 추가된 데이터 셋의 예시를 보여주고 그림 5는 원래 데이터 셋 (ORG) 과 추가된 데이터 셋 (DS) 의 클래스 별 추가된 데이터 분포 차이를 보여준다.

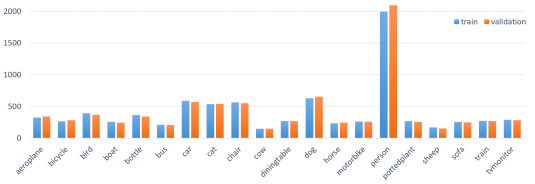

그림 3. Pascal VOC 데이터 셋의 클래스 별 데이터 수
Fig. 3. The amount of data per class of Pascal VOC dataset

그림 4. 크롤링 데이터 셋의 예시
Fig. 4. Example of Crawled Dataset

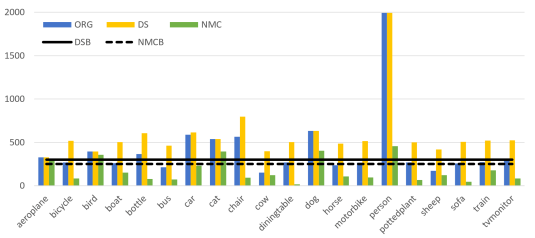

그림 5. 클래스 별 데이터 개수 분포
Fig. 5. Per-Class Data Distribution of evaluated Dataset Variation

2.1.2 클래스 불균형 문제 (Class Imbalance)

훈련 시 클래스 간 이미지 개수를 유사하게 맞추는 클래스 균등화 (Class balancing) 는 인식 시에 클래스 간 편향성을 줄여줄 수 있는 전처리 단계이다. 그림 3에서 보듯이, Pascal VOC 데이터 셋은 클래스 간 불균형이 심하다. 학습 데이터 (training data) 가 총 8,331장으로 이루어져 있으나 그 중 1,994장이 person class에 속해있었으며, 가장 data가 적은 class (151장) 와는 13배가량 차이가 난다. 본 논문에서는 20개 클래스 간 이미지 개수를 유사하게 맞추기 위해 이미지가 부족한 클래스는 위에서 설명한 것처럼 추가적인 데이터 보충 단계를 진행하였고 일부 데이터가 너무 많은 클래스는 300여장으로 균일하게 줄여서 전처리 작업을 수행하였고 이에 대한 영향도 분석하였다.

2.1.3 멀티 클래스 데이터 문제 (Multi-Class)

원 데이터 셋인 Pascal VOC 에서는 한 이미지 내 여러 클래스 (multi- class) 가 다수 존재한다. 이런 다

www.dbpia.co.kr

중 클래스를 포함하는 이미지는 딥러닝 네트워크 훈련 시 인식률에 어려움을 미칠 수 있다. 그래서 이런 다중 클래스를 포함한 이미지를 제거하고 훈련함으로의 인식률 변화를 확인하였다. 이런 단일 클래스만을 포함시킨 데이터 셋의 클래스별 데이터 셋 분포도는 그림 5에 나타내었다. 또한 위에서 언급한 클래스 불균형 문제도 동시에 고려한 (250여장으로 제한) 데이터 셋 (NMCB)도 실험하였고 이들의 클래스 별 데이터 분포도는 그림 5에 나타내었다.

우리는 이러한 어려운 점들을 분석하고 해결하기 위해 데이터 전처리 과정에서 다음의 시도를 하여 성능을 분석해 보았다.

- 데이터 보충 (DS) : 과적합을 해결하기 위해 부족한 훈련 데이터를 보충하여 학습
- 클래스 균등화 (DSB) : 클래스 불균형 문제를 해결하기 위해 클래스 간 데이터 비율을 300여장으로 평탄화하여 데이터 셋 전처리
- 멀티 클래스 제거 (NMC) : 멀티 클래스 인식의 어려움으로 멀티 클래스 데이터를 제거해서 학습
- 멀티 클래스 제거+클래스 균등화 (NMCB) : 멀티 클래스와 클래스 불균등 문제 (250장으로 제한)를 제거한 데이터 셋

2.2 성능 분석

본 논문에서는 위에서 언급된 데이터 셋의 문제점들을 면밀히 관찰하기 위해 보편적으로 사용되는 ResNet에서 좀 더 축약한 그림 6의 ResNet 37 계층 모델을 사용하였고 전이학습을 사용하지 않고 처음부터 학습시켰다.

분석 데이터 셋의 조합은 다음과 같다.

- ORG : Pascal VOC 기본 데이터 셋
- DS : ORG 데이터 셋에 각 클래스 별로 250~1000장 이 되도록 데이터 보충
- DSB : DS 데이터 셋에서 모든 클래스들이 균일하게 300장이 되도록 조정
- NMC : ORG 데이터 셋의 훈련 데이터에서 여러 클래스가 아닌 한 개의 클래스 만 포함하도록 조정한 데이터 셋
- NMCB : NMC 데이터 셋에서 모든 클래스들이 균일하게 250장이 되도록 추가

그림 7에서 알 수 있듯이 ORG 데이터 셋에 비해 DS 데이터 셋에서 성능이 많이 향상된 것을 알 수 있다. 이로써 데이터양이 확보되는 것이 정확도 향상에 중요하다는 것을 알 수 있다.

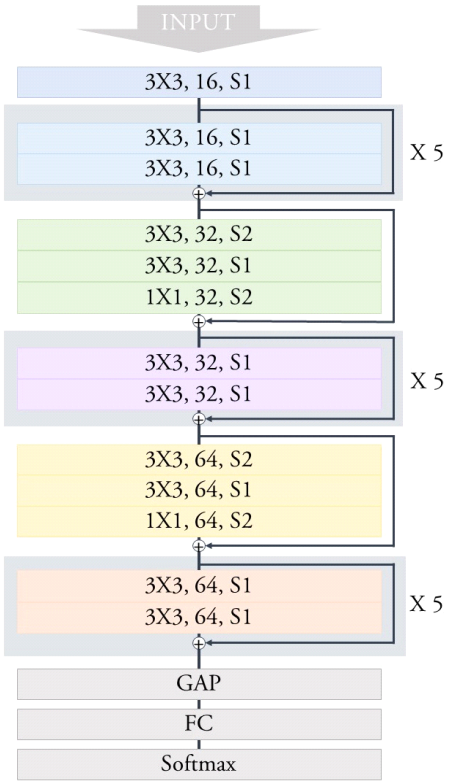

그림 6. 축약된 ResNet
Fig. 6. Compressed ResNet

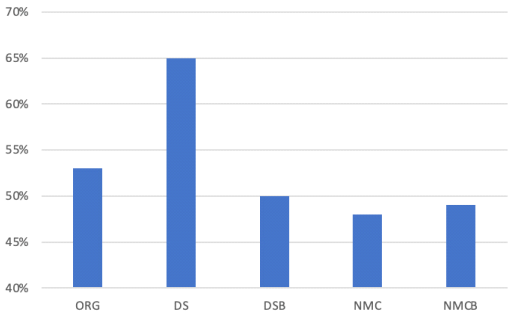

그림 7. Pascal VOC 성능
Fig. 7. Performance of Pascal VOC

두 번째로 클래스 별 불균형을 해결하기 위한 DSB 성능을 보면 오히려 조금 줄어든 것을 확인할 수 있다. 그 원인은 데이터양이 많았던 일부 클래스의 데이터양이 줄어듦으로써 해당 클래스의 정확도 감소가 일어났기 때문이다. 또한 검증 (validation) 용 데이터에서도 클래스 별 불균등이 존재하기 때문에 실제 인식 시에 사용되는 데이터 셋의 특성을 반영한 클래스 분포도를 갖는 것이 중요하다 하겠다.

www.dbpia.co.kr

세 번째로 ORG 훈련 이미지에서 멀티 클래스 이미지를 제거한 NMC 의 성능이 제일 줄어든 것을 알 수 있다. 그 이유는 NMC에서 데이터양이 줄어들기 때문에 거기서 오는 성능 열화가 상당한 것으로 분석된다. 두 번째의 경우와 마찬가지지만 데이터양이 적으면 클래스 균등화 혹은 멀티 클래스 제거와 같은 효과는 크지 않거나 오히려 열화의 원인이 된다.

네 번째로 NMCB 의 성능이 DSB 의 성능보다 작다는 것에 주목하자. NMCB 와 DSB 의 데이터는 동일하게 클래스 당 250장이지만 멀티클래스 데이터의 존재가 오히려 성능 향상에 도움이 된다는 것을 알 수 있다.

위 성능 분석들을 통해 우리는 다음과 같은 중간 결론들을 도출 할 수 있다.

1) 훈련 데이터의 양을 증가시키면 인식 정확도는 향상된다.
2) 훈련 데이터의 클래스 별 분포도는 검증 데이터의 클래스 별 분포도에 의존해야 한다.
3) 훈련 데이터에서 멀티 클래스 데이터들은 인식 정확도를 향상시킬 수 있다.

특히 데이터 양 증가는 정확도 향상에 제일 큰 영향을 미치므로 이에 대한 연구가 해당 관련 제품들의 실용화에 필수적이라 하겠다. 하지만, 실제로 이렇게 새로운 데이터를 구하기란 꽤 많은 비용을 치러야 할 수 있다. 다양한 사생활 이슈와 저작권 이슈가 존재하기 때문에 이를 어느 정도 해결한 데이터 증식 기술로 해결하는 것이 효과적이라 하겠다. 이제 다음 장에서는 이런 이미지 분류 네트워크에서 효과적인 데이터 증식 기술을 알아보고 어떤 기술이 해당 소규모 데이터 셋에 적합한지 실험으로 분석하여 선정한다.

Ⅲ. 데이터 증식 (Data Augmentation)

3.1 기본 데이터 증식 (Baseline Data Augmentation)

위에서 제시된 문제점들 중 가장 어려운 문제는 사실 상 과적합 이며 이는 곧 데이터 부족과 밀접하게 연관되어 있기 때문에 이런 데이터 증식과 관련된 여러 연구들이 진행되어 왔다. 우선 ResNet에서 ImageNet 훈련용도로 사용하였던 기본 데이터 증식 방법은 다음과 같다[6]. 단 아래 6번은 데이터 셋 마다 달라질 수 있음에 유의한다. 예를 들어 Cifar10 의 경우 mean = [125.307, 122.95, 113.865], std = [62.9932, 62.0887, 66.7048] 의 값을 갖는다.

기본 데이터 증식 방법 (ImageNet+ResNet)
1. 임의로 한 이미지를 뽑아서 [0, 255] 범위의 32-bit 소숫점 값으로 변환한다. 2. [3/4, 4/3] 의 범위에서 임의의 비율을 가지며 [8%, 100%] 의 범위에서 임의의 넓이를 가지는 사각형 영역을 Crop하여 224*224 크기로 resize 한다. 3. 0.5의 확률로 수평으로 뒤집는다. 4. [0.6, 1.4] 으로부터 균일한 확률로 발생된 Coefficient로 Hue (색상), Saturation (채도), Brightness (명도) 값을 Scale 한다. 5. 정규분포 N(0, 0.1) 로부터 뽑혀진 Coefficient를 가진 PCA 잡음을 더한다. 6. RGB 채널들에 평균 = [123.68, 116.779, 103.939] 값을 빼고 표준편차 = [58.393, 57.12, 57.375] 값으로 나누어 정규화 한다.

위 기술에 기반하여 Cifar100 과 reduced Cifar100 를 학습시켰을 때의 성능은 그림 10, 11과 같다. 또한, 비교를 위해 데이터 증식을 적용하지 않고 기본 데이터 셋에 대해서만 훈련을 시켰을때의 성능은 그림 8, 9와 같다.

그림에서 보듯이 데이터 증식을 적용하지 않은 Cifar100과 reduced Cifar100의 성능은 55.2%, 16.6% 이며 기본 데이터 증식을 적용하였을 때 Cifar100의 ResNet 성능은 65.2 %, reduced Cifar100 의 경우

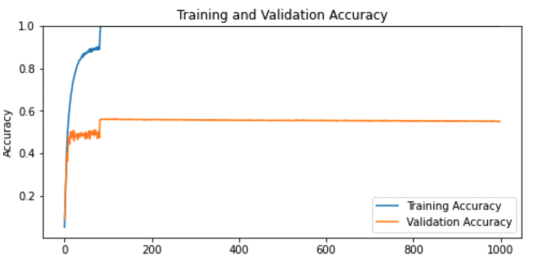

그림 8. Cifar100 의 기본 데이터 성능
Fig. 8. Performance of original dataset for Cifar100

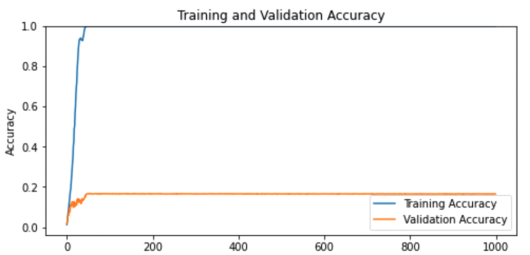

그림 9. Reduced Cifar100 의 기본 데이터 성능
Fig. 9. Performance of original dataset for Reduced Cifar100

www.dbpia.co.kr

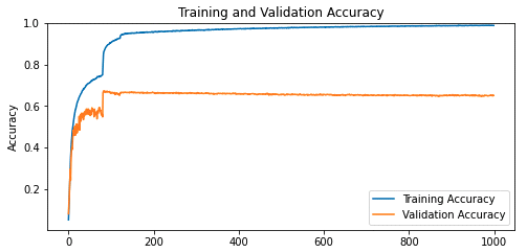

그림 10. Cifar100 의 기본 데이터 증식 성능
Fig. 10. Performance of baseline data augmentation for Cifar100

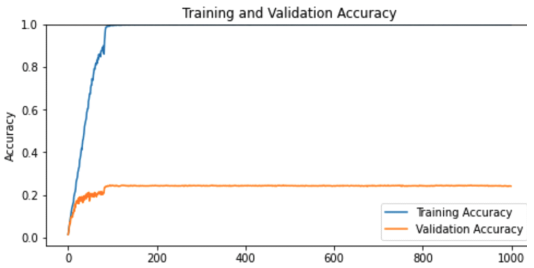

그림 11. Reduced Cifar100 의 기본 데이터 증식 성능
Fig. 11. Performance of baseline data augmentation for Reduced Cifar100

24.9 % 의 정확도 성능이 도출되었음을 알 수 있다. 기본 데이터 증식 기술만으로도 Cifar100의 경우 약 10%, reduced Cifar100의 경우 약 5.3%의 성능향상이 있었음을 알 수 있다.

특히 이런 데이터 증식 기술을 적용하였음에도 reduced Cifar100 의 경우 심각한 과적합 현상이 일어나는데 이는 데이터가 상당히 부족한 상황 (기존 대비 8%) 이고 기존 데이터 증식 기술만으로는 데이터가 부족하므로 과적합 현상이 크게 나타났음을 알 수 있다.

3.2 MixUp

데이터가 부족한 데이터 셋의 경우에 위 기술만으로는 성능 개선에 한계가 있어 더욱 다양한 데이터 증식 방식이 개발될 필요가 있다. 대표적 방법은 MixUp 방식으로 두 이미지의 RGB 값들을 λ : (1-λ) 의 비율로 가중합 하고 정답 Label 역시 두 One-Hot Label 들 역시 λ : (1-λ) 의 비율로 가중합 하여 훈련시키는 방식이다.

이제 위 Cifar100 와 reduced Cifar100에 위의 데이터 증식 기술들을 적용하였을 때의 성능들을 살펴본다. MixUp 방식 적용 시, λ 값은 베타 분포 $\beta(\alpha=0.2,\ \beta=0.2)$ 에서 random하게 추출하여 이

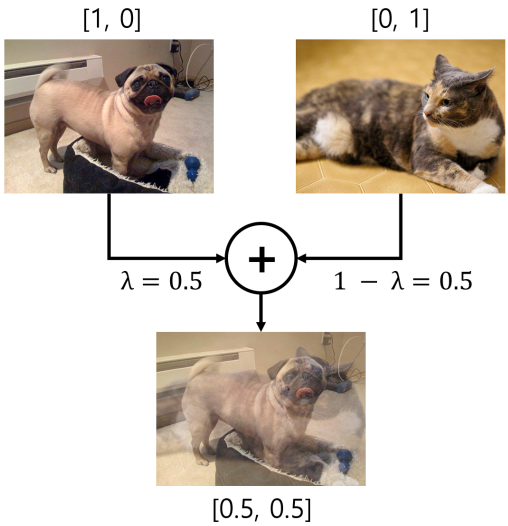

그림 12. MixUp 데이터 증식 기술 적용 예
Fig. 12. Example of MixUp data augmentation

미지와 레이블링에 해당 비율로 MixUp하였다. 또한 MixUp 적용 시 100개의 클래스들이 서로 섞이기 때문에 1/100의 확률로 동일 클래스 들이 섞일 수 있으나 이 또한 동일 클래스에 대한 Augmentation 이기에 그냥 두고 실험 하였다. 그러면 MixUp 방식을 적용하였을 때 Cifar100 의 성능은 71.1%, reduced Cifar100

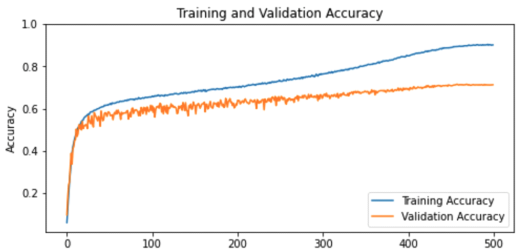

그림 13. Cifar100 의 MixUp 데이터 증식 성능
Fig. 13. Performance of MixUp data augmentation for Cifar100

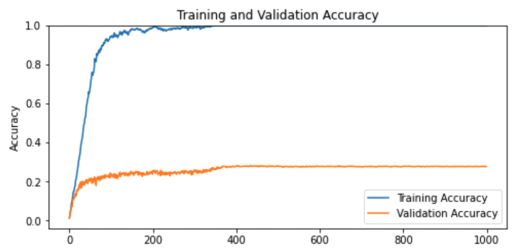

그림 14. Reduced Cifar100 의 MixUp 데이터 증식 성능
Fig. 14. Performance of MixUp data augmentation for Reduced Cifar100

www.dbpia.co.kr

의 성능은 27.9% 로 기존 대비 약 5.9%, 3.0% 의 성능향상이 있음을 알 수 있다.

이처럼 복수개의 클래스 간 MixUp 하는 방식은 성능 향상에 효과적이기 때문에 기존 기술들의 단점을 보완한 여러 변이 기술들이 출현하였다.

3.3 Cutout

반면 데이터 증식 기술에서 기존 이미지에 노이즈를 추가하는 방식으로 발전하는 축이 있다[21]. 해당 [21] 논문에서 Bengio 교수팀은 Salt and Pepper 노이즈를 추가하거나 motion blur, Gaussian smoothing, Gaussian noise, pixel permutation 등을 추가하여 데이터 증식을 적용하였다. 하지만 이런 방식의 노이즈 추가는 성능향상에 제한적이었다.

이런 노이즈를 추가하여 데이터 증식 하는 것을 개선하여 일정부분을 블랍 (Blob) 이나 사각형으로 마스킹 (Masking) 하는 방식의 Cutout 기술이 개발되었다 [8]. 그러면 위에서 설정한 Cifar100와 Reduced Cifar100 데이터 셋과 ResNet 모델에 이 Cutout 기술을 적용하였을 때의 성능을 알아본다.

이 Cutout 을 적용하기 위한 zero mask는 사각형 형태로서 좌상단 좌표는 이미지 전체에서 균일한 분포로 랜덤하게 추출하였다. 그 넓이는 이미지 넓이의 2%~ 40%까지의 범위에서 랜덤하게 추출하였으며 0.3 부터 3.3 의 가로 세로 비율로 masking을 적용하였다.

그림 16과 17에서 보듯이 Cifar100 의 경우 70.1%, Reduced Cifar100 의 경우 28.2% 의 성능이 나왔다. 결과를 통해 확인할 수 있듯이 Cifar100의 경우 4.9%, Reduced Cifar100의 경우 3.3% 의 성능 향상이 있었다.

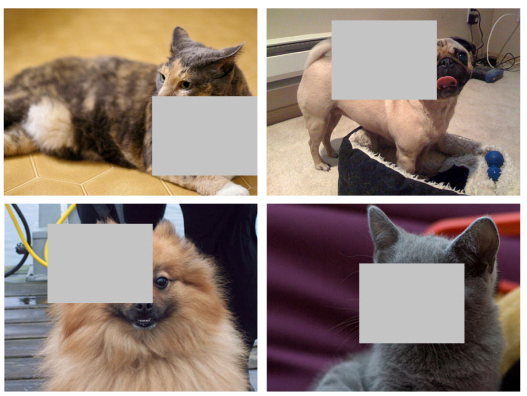

그림 15. Cutout 데이터 증식 기술 적용 예
Fig. 15. Example of Cutout data augmentation

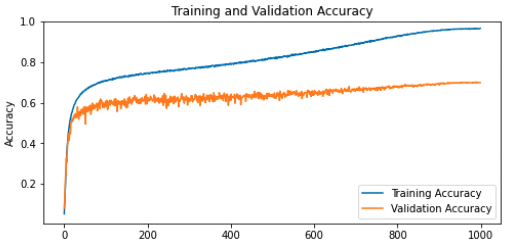

그림 16. Cifar100 의 Cutout 데이터 증식 성능
Fig. 16. Performance of Cutout data augmentation for Cifar100

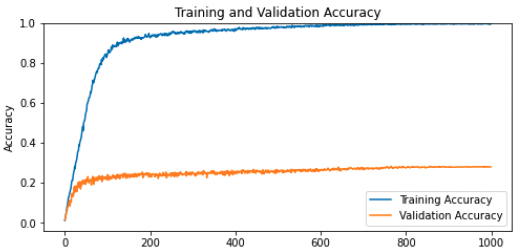

그림 17. Reduced Cifar100 의 Cutout 데이터 증식 성능
Fig. 17. Performance of Cutout data augmentation for Reduced Cifar100

3.4 CutMix

이렇게 복수의 이미지를 섞는 기술인 MixUp 기술과 한 이미지 내에서 임의로 Blob을 생성하여 제거하는 기술인 Cutout 기술 모두 데이터 증식 기술로서 성능 향상을 보이기 때문에 이 두 기술을 결합한 CutMix[9] 즉, 임의로 블랍 (Blob)을 만든 자리에 다른 이미지의 영역을 넣은 방식의 기술이 개발되었다.

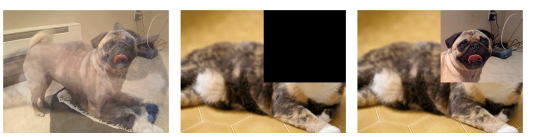

그림 18. CutMix 데이터 증식 기술 적용 예
Fig. 18. Example of CutMix data augmentation

그림 19. Reduced Cifar100 의 CutMix 데이터 증식 성능
Fig. 19. Performance of CutMix data augmentation for Cifar100

www.dbpia.co.kr

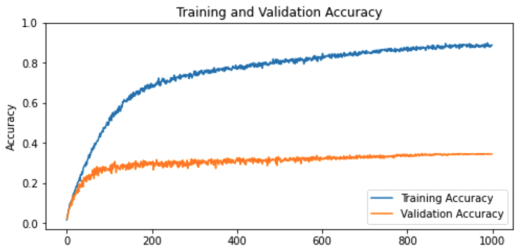

그림 20. Reduced Cifar100 의 CutMix 데이터 증식 성능
Fig. 20. Performance of CutMix data augmentation for Reduced Cifar100

표 2. Cifar100 와 Reduced Cifar100 의 다양한 데이터 증식 성능
Table 2. Various data augmentation performance of Cifar100 and Reduced Cifar100

	Cifar100	Reduced Cifar100
NoAug	55.2	16.6
Baseline	65.2	24.9
MixUp	71.1	27.9
CutOut	70.1	28.2
CutMix	74.0	34.8

위 데이터 증식 기술을 Cifar100와 Reduced Cifar100 에 적용하였을 때의 성능을 살펴본다. 그림 19, 20에서 보듯이 Cifar100는 74.0%, Reduced Cifar100는 34.8 % 의 성능을 얻어 낼 수 있음을 알 수 있다. 이는 기존 기본 데이터 증식 기술 대비 Cifar100의 경우 8.8%, Reduced Cifar100의 경우 9.9% 의 성능 향상을 이끌어 낼 수 있었다.

이제 위에서 언급한 기술들의 성능을 비교해본다.

위의 표를 통해 CutMix 의 성능이 가장 좋음을 알 수 있으며 이는 소규모 데이터 셋 (Reduced Cifar100)에서 특히 좋은 성능을 나타냄을 알 수 있다. 즉, 데이터의 양이 적은 소규모 데이터 셋에서 CutMix처럼 좀 더 어렵고 다양한 데이터 증식 기술을 응축한 데이터로 훈련하는 것이 효과적임을 알 수 있다.

또한 MixUp 기술과 Cutout 기술을 비교하였을 때 대용량 데이터 에서는 MixUp 기술이 좋고 소규모 데이터 셋인 Reduced Cifar100에서는 Cutout 기술이 미세하게 더 좋음을 알 수 있다. 사실 MixUp을 통해 얻어지는 멀티 클래스 데이터의 품질이 그림 12에서 보듯이 좋지는 않다. 즉 두 이미지를 MixUp을 통해 섞으면 두 이미지 안의 인식 객체들이 서로 동일한 위치에 있을 경우 인식에 어려움이 생길 정도의 객체들이 출현하게 되므로 합성 멀티 클래스 데이터가 항상 좋지만은 않음에 주목하자. 하지만 대용량 데이터 셋에서는 데이터의 양이 어느 정도 확보되어 있으므로 이런 불완전한 합성 멀티클래스 데이터가 있음에도 극복 가능한 다른 대체 학습 데이터들이 있어서 큰 폭의 정확도 향상이 가능했지만 소규모 데이터 셋은 이런 불완전한 합성 멀티 클래스를 대체할 충분한 데이터들이 확보되지 않아 정확도 향상이 제한적이었음을 알 수 있다. 이 실험으로 가능한 양질의 멀티 클래스 데이터를 확보하는 것이 정확도 향상에 더 도움이 될 수 있음을 알 수 있다.

Ⅳ. 실 험

이전 장에서 알아보았듯이 소규모 데이터 셋의 인식 성능에 가장 효과적인 CutMix 데이터 증식 기술을 Pascal VOC 데이터 셋에 적용하여 성능을 살펴본다. 우선 기본 실험 셋인 ORG에 위 데이터 증식 기술을 적용한 실험 셋을 ORGA 라 명명하고, 데이터 보충을 실시한 DS 실험 셋에 위 데이터 증식 기술을 적용한 실험 셋을 DSA 라 명명하였다.

- ORGA : ORG에 CutMix 데이터 증식 기술 적용
- DSA : DS에 CutMix 데이터 증식 기술 적용

그림 21에서 알 수 있듯이 데이터 보충 및 데이터 증식을 수행한 DSA 가 가장 좋은 성능을 보임을 알 수 있다. 물론 원 실험 셋 ORG에 데이터 증식 기술을 적용한 ORGA 역시 ORG 에 비해 성능 향상이 있었지만 데이터 보충과 데이터 증식을 모두 적용하는 것이 성능 향상에 효과적임을 알 수 있다. 또한, 데이터 보충에 많은 비용이 들어가는 것을 생각하면 비용이 들지 않는 ORGA 역시 매우 좋은 해결책이라는 것을 알 수 있다.

두 번째로 NMC 와 ORGA 의 비교를 통해 멀티

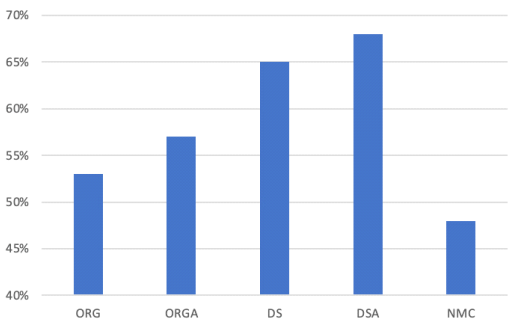

그림 21. Pascal VOC 의 데이터 증식 성능
Fig. 21. Data augmentation performance of Pascal VOC

www.dbpia.co.kr

클래스 데이터가 정확도 향상에 큰 도움이 되는 것을 확인할 수 있다. 즉 적용한 데이터 증식 기술인 CutMix 는 멀티 클래스 데이터를 인공적으로 만드는 기술이므로 이는 단일 클래스 데이터만을 가지고 있는 NMC에 비해 좋은 성능을 나타냄을 알 수 있다.

V. 결 론

본 논문에서는 커스텀 데이터 셋들에서 발생할 수 있는 과적합, 클래스 불균형, 멀티 클래스 인식에 관한 어려운 점들을 분석해 보았고 실험을 통해 데이터 셋 양 확보가 성능에 중요한 요인이라는 것을 보였으며 멀티 클래스 데이터는 정확도 향상에 도움이 된다는 것을 검증하였다.

또한 고비용의 데이터의 보충 방법 외에 저비용의 데이터 증식 기술을 적용하는 것이 성능 향상에 효과적임을 보였고 특히 멀티 클래스 데이터를 만드는 종류의 데이터 증식 기술을 적용하는 것이 성능 향상에 효과적임을 실험을 통해 검증하였다.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” *IEEE CVPR*, 2009.

[2] https://opensource.google/projects/pen-mages-dataset

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” *ECCV,* 2014.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *NIPS,* 2015.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *ICLR,* 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *IEEE CVPR,* 2014.

[7] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “Mixup: Beyond empirical risk minimization,” *ICLR,* 2018.

[8] “Improved Regularization of Convolutional Neural Networks with Cutout,” *arXiv:1708.04552*, 2017.

[9] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, “CutMix: Regularization strategy to train strong classifiers with localizable features,” *ICCV,* 2019.

[10] https://www.cs.toronto.edu/~kriz/cifar.html

[11] http://host.robots.ox.ac.uk/pascal/VOC/

[12] https://www.kaggle.com/c/tiny-imagenet/

[13] L. N. Darlow, E. J. Crowley, A. Antoniou, and A. J. Storkey, “*CINIC-10 is not ImageNet or CIFAR-10*,” Report EDI-INF-ANC-1802.

[14] https://cs.stanford.edu/~acoates/stl10/

[15] W. Li, L. Wang, W. Li, E. Agustsson, and L. Van Gool, “Webvision database: Visual learning and understanding from web data,” *arXiv preprint arXiv:1708.02862*, 2017a.

[16] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, “Learning from massive noisy labeled data for image classification,” *IEEE CVPR*, pp. 2691-2699, 2015.

[17] https://github.com/visipedia/inat_comp/tree/master/2017

[18] https://cs.nyu.edu/~ylclab/data/norb-v1.0-small/

[19] L. Bossard, M. Guillaumin, and L. Van Gool, “Food-101 – mining discriminative components with random forests,” *ECCV*, pp. 446-461, 2014.

[20] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, “Mura: Large dataset for abnormality detection in musculoskeletal radiographs,” in *Conf. Med. Imaging with Deep Learn.,* 2018.

[21] Y. Bengio, et al., “Deep learners benefit more from out-of-distribution examples,” in *Proc. Fourteenth Int. Conf. Artificial Intell. and Statistics*, pp. 164-172, 2011.

[22] E. Bae and S. Lee, “Research of optimal lightweight image classification models in environments of mobile devices,” in *Proc. KICS Summer Conf.*, pp. 549-551, Aug. 2020.

www.dbpia.co.kr

배 은 지 (Eunjee Bae)

2020년 2월 : 동서울대학교 전자공학과 졸업
2021년 3월~현재 : 연세대학교 전기전자공학과 석사과정
<관심분야> 딥러닝, 영상인식, 3D Reconstruction
[ORCID:0000-0001-6226-1490]

이 성 진 (Sungjin Lee)

2011년 8월 : 연세대학교 전기전자공학과 박사 졸업
2012년 9월~2016년 7월 : 삼성전자 DMC연구소 책임연구원
2016년 7월~현재 : 동서울대학교 전자공학과 조교수
<관심분야> 딥러닝, 영상인식, 3D Reconstruction
[ORCID:0000-0003-3159-8394]

www.dbpia.co.kr