creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

**저작자표시**. 귀하는 원저작자를 표시하여야 합니다.

**비영리**. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

**변경금지**. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이 규 중 교수 지도
석사학위 청구논문

# 딥러닝을 위한 이미지 처리 기반의 데이터 증강

2024

성신여자대학교 대학원
미래융합기술공학과
오 주 현

# 딥러닝을 위한 이미지 처리 기반의 데이터 증강

이 규 중 교수 지도

이 논문을 석사학위 논문으로 제출함

2023년 11월

성신여자대학교 대학원

미래융합기술공학과

오 주 현

# 인 준 서

오주현의 석사학위 논문으로 인준함

2023년 11월

심사위원장 변 혜 원 (인)

심 사 위 원 오 장 민 (인)

심 사 위 원 이 규 중 (인)

성신여자대학교 대학원

# 논문개요

이미지를 다루는 심층 학습 분야에서는 CNN 구조의 모델을 많이 사용한다. CNN 모델의 매개변수가 학습데이터보다 많을 때, 훈련 성능은 좋지만 테스트 성능이 좋지 않은 과적합이 발생할 수 있다. 이처럼 일반화 능력이 떨어질 때, 단순히 데이터 수집을 통한 해결은 비용이 들고 효율적이지 못하다. 일반화 성능을 향상시키는 방법으로 데이터 증강법과 규제화이 있다. 데이터 증강법은 원본 데이터를 뒤집기, 자르기, 색 지터(color jitter) 등의 처리를 통해서 데이터의 다양성을 확보하는 방법이다. 규제화은 학습 중 파라미터나 중간 특징(feature)을 규제하는 방법을 통해서 과적합을 방지하는 것이다. 데이터 증강법에서는 Fast R-CNN, Style Gan 등과 같은 심층 신경망을 활용하는 방법이 있지만, 추가적인 모델 학습과 데이터 처리가  비용이 든다. 반면, Random Erasing, Cutout, Hide-and-Seek 같은 방법은 비용이 적게 들지만, 이미지의 일부를 무작위로 지우기에 확률에 의존한다.

본 논문은 이미지 분류 모델에 대한 효율적이면서 효과적인 데이터 증강법을 제안하고자 한다. 기존의 증강법 중에 지우기(erasing) 기법들은 객체나 배경에 상관없이 확률적으로 지우기에 최적화할 요소가 다분하다. 기존의 학습된 분류 모델의 분석을 통해서 각 이미지의 CAM(Class Activation Map)을 통해 추가적인 모델 없이 객체의 추정이 가능하다. 이를 바탕으로 객체의 일부의 폐색하여, 효과적인 지우기 데이터 증강법을 제안하고자 한다. Resnet과 CIFAR10을 활용하여 CAM을 활용한 지우기 기법의 성능을 확인한다.

# 목 차

**논문개요**

**Ⅰ. 서론** ····· 1

**Ⅱ. 관련 연구** ····· 3

1. 규제화 연구 ····· 3
2. 데이터뱅크 증강법 ····· 4
   1) Random Erasing ····· 7
   2) Cutout ····· 8
   3) Hide-and-Seek ····· 9

**Ⅲ. 제안 방법** ····· 11

1. 지우기 증강법 문제점 ····· 11
2. 제안 알고리즘 ····· 12
3. Class Activation Mapping(CAM) ····· 16
   1) CAM ····· 16
   2) Gradient weighted CAM ····· 16
4. 실험 세팅 및 결과 비교 ····· 18
   1) 실험 세팅 ····· 18
   2) 결과 비교 ····· 19

Ⅳ. 결론 ····· 22
1. 요약 ····· 22
2. 한계 및 향후 연구 ····· 22

참고문헌
ABSTRACT

# 그 림 목 차

[그림 1] 데이터 증강법 분류 ····· 6
[그림 2] Random Erasing 예시 ····· 8
[그림 3] Cutout ····· 9
[그림 4 ] Hide-and-Seek 의사 코드 ····· 10
[그림 5] Hide-and-Seek 예시 ····· 10
[그림 6] 기존 지우기 증강법 ····· 11
[그림 7 ] CAM 이미지 예시 ····· 12
[그림 8] 알고리즘 흐름도 ····· 13
[그림 9] 제안한 지우기 방법 ····· 14
[그림 10] Random Erasing 지우기 비율 분포 ····· 15
[그림 11] Cutout 지우기 비율 분포 ····· 15
[그림 12] Hide-and-Seek 지우기 비율 분포 ····· 15
[그림 13] CIFAR10 데이터 예시 ····· 19

# 표 목 차

[표 1] Resnet20 및 cifar10에 대한 성능 비교, error rate(5번 평균) ····· 14
[표 2] 지우는 방법에 대한 각 모델별 성능 비교 ····· 21
[표 3] 하이퍼 파리미터 에 대한 결과 ····· 21
[표 4] CAM을 얻는 방법에 대한 비교 ····· 21

# Ⅰ. 서 론

이미지 관련 심층신경망 학습으로는 CNN(Convolutional Neural Networks)을 많이 활용한다. ViT(Vision Transformer)[1]의 등장으로 수많은 이미지 분야에서 좋은 성능을 냈지만, 순수 CNN 구조인 ConvNeXt[2]에서 최근 트랜스포머(Transformer)의 기법을 적용하니 트랜스포머없이 좋은 성능을 낼 수 있었다. CNN은 다양한 분야에서 뛰어난 성능을 보이며 중요한 역할을 하고 있다. 특히, 이미지 인식, 음성 인식 및 자연어 처리와 같은 복잡한 문제들을 해결하는 데 있어 CNN의 능력은 강력한 도구로 인정받는다. 그러나 이러한 능력에도 불구하고, CNN은 과적합과 일반화 문제에 취약할 수 있다. 과적합은 모델이 훈련 데이터에만 지나치게 최적화되어 새로운 데이터에 대한 예측 성능이 저하되는 현상을 말한다. 특히 훈련 데이터가 제한적이며 파라미터의 개수가 많을 때 발생할 수 있다.

이러한 문제를 해결하기 위해 데이터 증강법(Data Augmentation)과 규제화(Regularization) 기법이 중요한 역할을 할 수 있다. 데이터 증강법은 기존 훈련 데이터에 다양한 변형을 적용하여 데이터의 양과 다양성을 인위적으로 증가시키는 방법이다. 이는 모델이 더 다양한 패턴과 상황을 학습하게 하여 일반화 능력을 향상시키는 데 도움이 된다. 규제화에는 가중치 감쇠(Weight Decay)와 드롭아웃(Dropout), 배치 정규화(Batch Normalization) 등이 있으며 모델의 복잡성을 제어하고 과적합을 방지할 수 있다. 이러한 기법들은

---

1) Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv, abs/2010.11929.

2) Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11966-11976.

모델이 훈련 데이터에 지나치게 의존하지 않도록 하여 일반화 성능을 향상시키는 데 기여한다.

데이터 증강법 중의 지우기(erasing) 기법은 데이터의 일부를 지우는 것으로 모델의 일반화 성능을 올리는 방법이다. 규제화은 넓은 의미에서 학습 오류 보다는 일반화 오류를 줄이기 위한 목적으로 학습 알고리즘을 수정하는 것이다.[3] 데이터 증강법은 학습 과정 중 입력 데이터 처리를 하는 규제화으로 볼 수 있다. 지우기 기법은 일반화 능력에서도 중요한 영향을 끼치는 폐색(Occlusion)과 관련이 있다. 일반화 성능이 좋은 CNN모델은 객체의 일부가 폐색이 되어도 객체의 나머지 구조를 통해서 인식을 할 수 있어야 한다. 대개의 데이터는 폐색에 대한 데이터의 댜양성이 제한되어 있기에 일부 폐색이 된 테스트 데이터에 대한 성능이 떨어질 수 밖에 없다. CNN과 같은 이미지 데이터를 일부 폐색된 데이터로 증강할 수 있도록 Random Erasing[4]이 제안된 것이다. 이러한 지우기 기법은 이미지의 일부를 무작위로 지워 나머지 부분을 통해서 모델이 추론할 수 있도록 학습을 돕는 방법이다.

기본적인 지우기 기법에서는 이미지를 지우는 것이 무작위이기 때문에 객체를 폐색하지 않고 배경의 일부만 지는 경우가 발생할 수 있다. 이를 CAM(Class Activation Map)을 활용하여, 이미지 내의 객체의 위치를 추정하고 지우기 기법들을 최적화하는 방안을 제안하고자 한다.

본 논문에서는 다음과 같은 구성으로 진행된다. 2장에서는 관련 연구로서 규제화 연구, 데이터 증강법, CAM을 살펴보고, 3장에서는 제안하는 알고리즘과 실험결과를 비교하며, 4장에서는 제안 방법에 대한 결론과 한계 및 향후 연구 방향을 서술한다.

---

3) Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.
4) Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2017). Random Erasing Data Augmentation. ArXiv, abs/1708.04896.

# Ⅱ. 관련 연구

## 1. 규제화 연구

규제화는 CNN 모델에서 훈련 과정 중 과적합을 방지할 수 있는 중요한 구성 요소이다. 기존의 규제화와 관련된 연구로는 드롭아웃, DropConect[5], Adaptive dropout[6], Stochastic Pooling[7], DisturbLabel[8], PatchShuffle[9] 등이 있다. Dropout은 심층 신경망의 각 뉴런(neuron)에 확률을 적용하여, 남아있는 가중치에 대한 기여를 순전파와 역전파에 고려한 방법이다. DropConnect는 뉴런 대신에 임의의 가중치를 선택하여 훈련에 기여할 수 있게 한 드롭아웃의 일반화된 형태이다. Adaptive dropout은 드롭아웃의 뉴런의 확률을  이진 신뢰 네트워크(binary belief network)로 추정할 수 있는 방법을 제안한다. Stochastic pooling은 훈련 과정 중에 다항 분포를 이용하여 활성화될 값을 선택하는 방법으로, 매개변수를 사용하지 않고 다른 규제화 기술과 같이 적용될 수 있다. Distrub Label은 훈련 이터레이션(Iteration)마다 주어진 데이터 샘플의 일부분 레이블을 임의의 잘못된 값으로 변경하여 손실 계층에 노이즈(Noise)를 추가한 방법이다. PatchShuffle은 CNN 모델의 훈련 과정 중 지역 변동을 증가시키기 위하여, 원래의 것과 거

5) Wan, L., Zeiler, M.D., Zhang, S., LeCun, Y., & Fergus, R. (2013). Regularization of Neural Networks using DropConnect. International Conference on Machine Learning.
6) Ba, J., & Frey, B.J. (2013). Adaptive dropout for training deep neural networks. Neural Information Processing Systems.
7) Zeiler, M.D., & Fergus, R. (2013). Stochastic Pooling for Regularization of Deep Convolutional Neural Networks. CoRR, abs/1301.3557.
8) Xie, L., Wang, J., Wei, Z., Wang, M., & Tian, Q. (2016). DisturbLabel: Regularizing CNN on the Loss Layer. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4753-4762.
9) Kang, G., Dong, X., Zheng, L., & Yang, Y. (2017). PatchShuffle Regularization. ArXiv, abs/1707.07103.

의 동일하게 전역 구조를 유지하고 무작위로 각 지역 패치(patch) 내의 픽셀을 섞는다. 본 연구에서는 딥러닝 모델에 규제를 가하는 것이 아닌 데이터에 규제를 가하는 데이터 증강법에 대해서 논의하고자 한다.

## 2. 데이터 증강법

데이터 증강법은 기존의 데이터를 활용하여 데이터의 다양성을 확보하는 방법이다. 이미지 데이터 증강법은 크게 기본 접근법과 고급 접근법이 있다.[10] 그림[1]과 같이 세부적으로 분류할 수 있다. 기본 접근법은 이미지 조작, 이미지 지우기, 이미지 혼합이 있다. 이미지 조작은 데이터에 간단한 조작을 가하는 것이다. 대부분 확률적으로 동작하며, 이미지를 수평 및 수직으로 뒤집거나 회전 크기, 비율을 조절하는 방법 등이 있다. 조작은 이미지의 변형을 통해서 객체의 크기 변형, 위치 변형 등을 통해서 다양성을 확보한다. 이미지 지우기는  확률적으로 이미지의 일부를 지우는 것이다. 이미지의 일부를 지워 객체에 대한 다양한 폐색 수준을 제공할 수 있다. 여기에는 Random Erasing[11], Cutout[12], HideAndSeek[13] 등이 있다. 본 연구는 이미지 지우기 증강법에 대한 주제를 다루므로 본 장에서 자세히 서술하겠다. 이미지 혼합은 서로 다른 이미지를 혼합하여 다양한 이미지를 만드는 것이

10) Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.
11) Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020, April). Random erasing data augmentation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13001-13008).
12) DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with Cutout. arXiv preprint arXiv:1708.04552.
13) Singh, K. K., Yu, H., Sarmasi, A., Pradeep, G., & Lee, Y. J. (2018). Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. arXiv preprint arXiv:1811.02545.

다. Mixup[14]은 서로 다른 이미지를 일정한 비율로 섞는 것이다. CutMix[15]는 Cutout과 Mixup을 결합한 것으로 Cutout을 먼저 수행하여, 지워진 자리를 다른 이미지로 채우는 것이다. PuzzleMix[16]는 CutMix가 무작위로 Cutout을 사용한 점을 개선하기 위해서 모델의 기울기를 활용한 객체의 돌출맵(saliency map)으로 각 이미지의 객체의 형상을 유지하면서 혼합하는 방법이다. 고급 접근법으로는 Auto augment, 특징(feature) 증강법, 생성형 모델이 있다. Auto augment[17]는 주로 기본 접근법에 해당하는 여러 가지 증강법을 탐색공간으로 설정하여, 여러 가지 하위 증강 정책을 탐색하면서 검증 데이터에 대한 정확도에 대한 보상으로 RNN 컨트롤러를 학습시키는 방법이다. 데이터에 맞는 최적의 정책을 찾을 수 있지만 학습에 큰 비용이 든다. 특징 증강법은 입력 공간인 이미지 데이터를 변환하는 대신에, 입력 공간(feature space)에서 변환을 수행하는 것이다. FeatMatch[18]는 데이터 셋 내 다른 이미지들의 특징에서 추출된 대표성을 띠는 프로토타입을 활용하여 이미지의 특징을 증강하는 방법이다. MomentExchange[19]는 학습된 특징의 평균, 분산과 같은 통계적 특성을 활용하여 데이터를 증강하는 방법이다.

---

14) Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

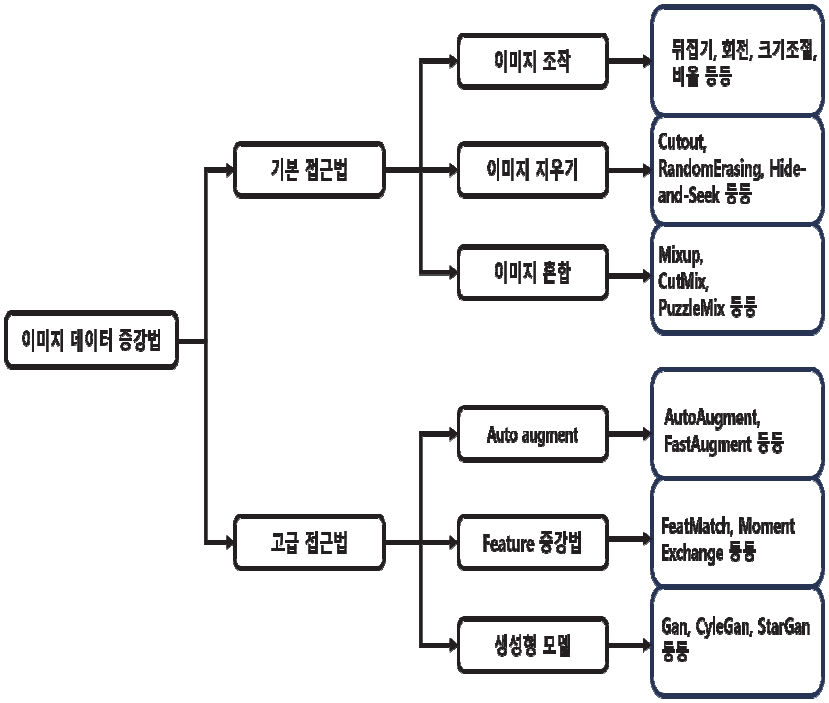
15) Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6023-6032).

16) Kim, J. H., Choo, W., & Song, H. O. (2020, November). Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In International Conference on Machine Learning (pp. 5275-5285). PMLR.

17) Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501.

18) Kuo, C. W., Ma, C. Y., Huang, J. B., & Kira, Z. (2020). Featmatch: Feature-based augmentation for semi-supervised learning. In Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XVIII 16 (pp. 479-495). Springer International Publishing.

19) Li, B., Wu, F., Lim, S. N., Belongie, S., & Weinberger, K. Q. (2021). On feature normalization and data augmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12383-12392).




[그림 1] 데이터 증강법 분류

생성형 모델을 활용하는 방법은 대표적으로 GAN[20](Generative Adversarial Networks) 모델이 있다. Cycle GAN[21]은 쌍을 이루지 않는 데이터 간의 변환을 학습하여 도메인 간 변환을 하는 것이 목표이며, 소스 도메인에서 타겟 도메인으로 변환 후, 다시 소스 도메인으로 복원하는 원본과의 일관성을 유지하는 순환 일관성 손실(Cycle Consistency Loss)을 사용하였다. StarGAN[22]은 하나의 네트워크를 통해서 다양한 도메인을 생성하도록

20) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

21) Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

22) Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In

하였다.

1) Random Erasing

Random Erasing은 2017년에 제안되었으며 무작위로 이미지를 지우는 방법이다. 이미지 데이터 모두를 지우지는 않고, 일정 확률을 학습할 때마다 지울지 말지 결정한다. 과정은 다음과 같다. 확률 $p$라는 하이퍼파라미터를 사용하여 처음 지울지 말지 결정한다. 확률 $p$에 해당하는 이미지는 Random Erasing을 적용하고, 나머지 $1-p$는 원본 데이터를 유지하게 된다. 보통 0.5의 확률을 사용하여, 절반의 데이터의 일부를 지우고 절반의 데이터는 보존한다. 원본 이미지 데이터의 너비와 높이를 $W, H$라 하고 이미지 크기 $S$라 하면 $S = W \times H$이다. 확률 $p$에 해당하는 이미지에서 지울 직사각형 영역을 $R_e$라고 하면 다음과 같이 정의된다.

$$R_e = (x_e, y_e, x_e + W_e, y_e + H_e) \tag{1}$$

이때, $x_e$, $y_e$는 원본 이미지 내에서 지울 영역의 임의 시작 위치이고 $x_e + W_e$와 $y_e + H_e$는 시작 위치와 마주하는 위치로 $W, H$를 넘지 못한다. $W_e$와 $H_e$는 $R_e$의 너비와 높이에 해당에 해당한다. 지우는 영역의 크기 $S_e$는 측면 비율 $r_e$과 함께 다음과 같이 정의된다.

$$W_e = \sqrt{\frac{S_e}{r_e}}, \qquad H_e = \sqrt{S_e \times r_e} \tag{2}$$

위의 공식에서 $S_e$는 $s_l, s_h$ 두 하이퍼파리미터의 범위 내에서 무작위로 결정되며, $r_e$는 $r_1, r_2$ 두 하이퍼파라미터 범위 내로 정해진다. $r_2$는 $r_1$의 역수의 값을 갖는다.

본 논문에서는 Random Erasing의 기존 설정과 같이 CIFAR10과 ResNet

---

Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).

을 활용한 이미지 분류를 통하여 성능 비교 실험을 진행한다. Random Erasing의 하이퍼파라미터 $s_l, s_h, r_1$는 각각 0.02, 0.4, 0.3으로 설정된다.

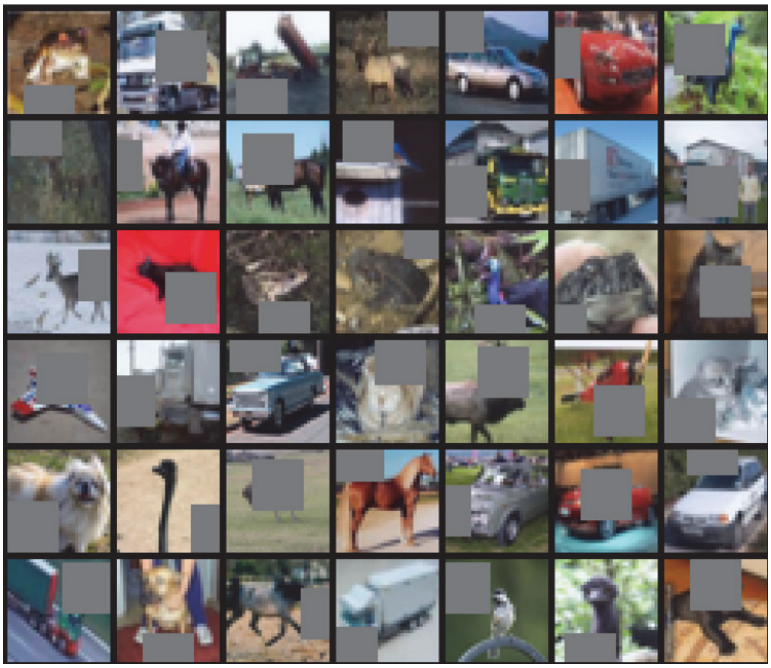
그림[2]는 Random Erasing에 대한 예시이며, 지워지는 영역의 값은 전체 데이터의 평균값, 랜덤값, 0 중 하나로 채워지며 본 논문에서는 지워지는 영영의 값을 0으로 채워서 진행하였다.

[그림 2] Random Erasing 예시

2) Cutout

Cutout은 Random Erasing과 비슷한 시기에 나왔으며, 방식 또한 많이 닮았다. Random Erasing과 마찬가지로 무작위 영역을 지운다. 다만 차이점이라면 그림[3]과 같이 지우는 영역은 일정한 정사각형 영역이다. 즉, 정사각형의 크기를 결정하고, 이미지 영역 내에서 위치를 무작위로 선택하는 것

이다. 다만 영역 내에서 선택하는 위치는 정사각형의 중심이다. 이미지를 벗어나서 자를 수 있기에 항상 일정한 면적을 자르지는 않는다. Cutout의 실험결과 CIFAR10 데이터셋에 대해서는 16x16 크기가 정확도가 좋았다.



[그림 3] Cutout

3) Hide-and-Seek

Hide-and-Seek은 그림 [4]와 같이 이미지를 일정한 간격의 그리드와 같이 패치를 나누고 각 패치를 확률에 의해 지운다. 지우는 과정은 그림[5]의 의사 코드와 같은 절차로 진행된다. 하이퍼파라미터 패치 크기(patch-size)와 숨길 확률(hide-prob)는 지워질 패치 크기와 패치가 지워질 확률을 나타낸다. 이미지 너비와 높이가 W,H라고 하면, W와 H에 대해서 패치 크기 만큼 탐색을 하고 이미지의 각각의 패치에 대한 숨길 확률에 의해서 결정된

다. CIFAR10 데이터셋에 대해서는 패치 크기와 숨길 확률은 각각 8과 0.5 이다.

**Algorithm 1** Hide-and-Seek Augmentation

```
procedure HideAndSeekAugmentation(image, patch_size, hide_prob)
    H, W ← GetHeight(image), GetWidth(image)
    for i ← 0, H, patch_size do
        for j ← 0, W, patch_size do
            patch ← Crop(image, i, j, patch_size)
            random_num ← Random
            if random_num < hide_prob then
                HidePatch(patch)                    ▷ Hide the patch
            end if
        end for
    end for
    return image
end procedure
```

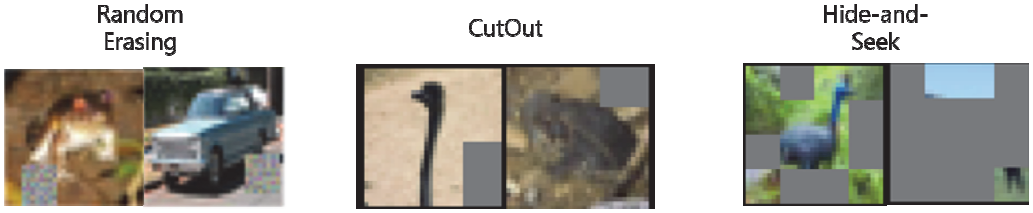
[그림 4 ] Hide-and-Seek 의사 코드

[그림 5] Hide-and-Seek 예시

# Ⅲ. 제안 방법

## 1. 지우기 증강법 문제점

기존의 지우기 증강법은 지우는 것을 확률적으로 하기에 객체와 관련없는 부위를 지우거나 폐색의 확률을 높일 경우 객체 대부분이 폐색이 되는 경우가 있다. 그림 [6]은 기존 지우기 증강법 중 실패 사례에 대한 예시이다. Random Erasing과 Cutout은 비슷하게 패치 한 개를 사용해서 지우므로 객체와 상관없는 배경만 지울 가능성이 있다. 패치의 개수를 늘려 지우는 영역을 늘릴 경우, 객체의 대부분이 폐색될 확률이 높아진다. Hide-and-Seek의 경우 확률상 이미지의 절반을 패치가 자르지만, 객체와 배경과 가까운 부분만 지울 때도 있고, 반대로 객체 전체를 폐색할 수도 있다.




[그림 6] 기존 지우기 증강법

이를 해결하기 위해서 객체 탐지 모델을 활용하는 방법[23]도 있지만 이는 추가적인 학습과 비용이 필요하다. CAM(Class Activation Mapping)을 활용

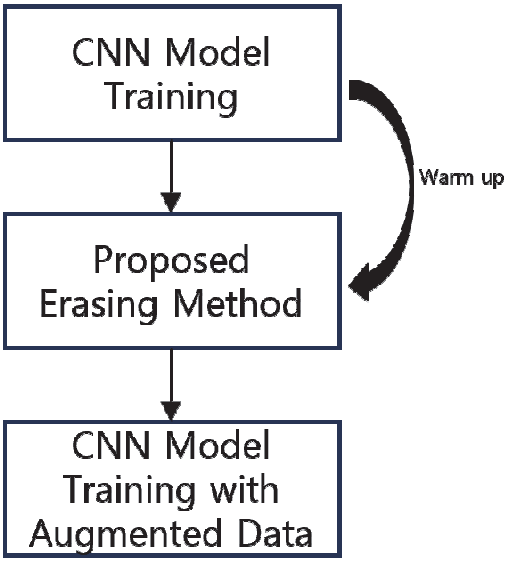
23) Wang, X., Shrivastava, A., & Gupta, A. (2017). A-fast-rcnn: Hard positive generation via adversary for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2606-2615).

하면 추가적인 객체 탐지 모델 없이 한 네트워크에서 약한 지도 객체 탐지(Weakly supervised object detection)가 가능하다. CAM은 분류 모델만으로도 그림[7]과 같이 대략적인 객체 탐지가 가능하다. 첫 번째 행의 이미지는 원본 이미지이고, 두 번째 행의 이미지는 원본 이미지와 CAM 히트맵(Heat map)을 겹친 이미지이다. 세 번째 행은 CAM 히트맵에 일정값(0.4)로 쓰레쉬홀딩(thresholding)하여 객체의 박스를 그린 것이다.

[그림 7 ] CAM 이미지 예시

## 2. 제안 알고리즘

제안 알고리즘은 그림[8]과 같은 흐름으로 진행한다. 처음은 지우기 기법 없이 일반적으로 훈련을 한다. CAM을 통한 객체 탐지를 위해서는 어느 정도의 훈련이 필요하다. 일정 웜업(warm up) 훈련을 한 후 제안한 지우기 방법을 사용한다. Resnet20 기준으로 훈련 정확도가 절반 에포크 이후 90% 이상이 되므로 웜업은 전체 훈련 에포크(epoch)의 절반 정도로 하였다. 웜업 이후 증강된 데이터와 원본 데이터와 함께 학습을 하도록 하였다.



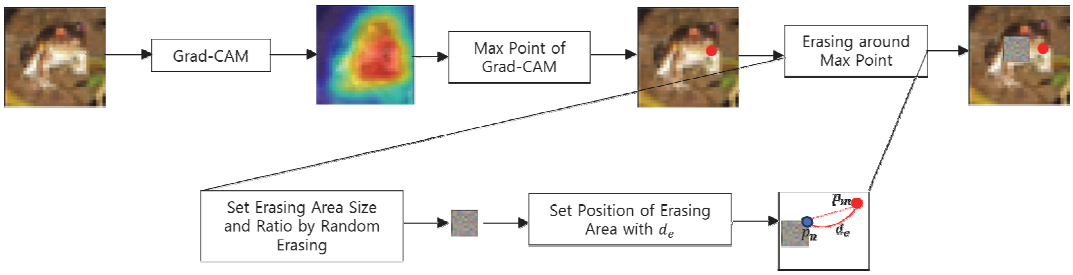

[그림 8] 알고리즘 흐름도

제안한 지우기 방법은 그림[9]와 같다. 먼저 이미지에서 Grad-CAM (Gradient-weighted Class Activation Mapping)을 통해 객체에 대한 위치 정보를 얻는다. 그 중 최댓값 점 $p_m$을 객체의 중심으로 여겨 지우기를 진행한다. 지우는 면적은 Random Erasing 방식을 따라하여, 객체에 대한 다양한 폐색을 주도록 한다. 점 $p_n$은 지우는 패치의 꼭짓점 중 하나이다. 점$p_m$과 $p_n$의 거리인 $d_e$는 수식(3)으로 결정되며, $b_h$는 하이퍼 파라미터로 객체로부터 최대 멀어질 수 있는 값이다.

$$d_e = (rand(0,b_h), rand(0,b_h)) \tag{3}$$

표[1]은 Renet20 및 CIFAR10에 대한 지우기 기법들에 대한 성능을 나타낸 것이다. 성능은 error rate로 모델 훈련을 5번 실행 후에 오답률에 대한 평균을 나타낸 것이다. Cutout의 성능이 가장 좋았으나 Random Erasing과

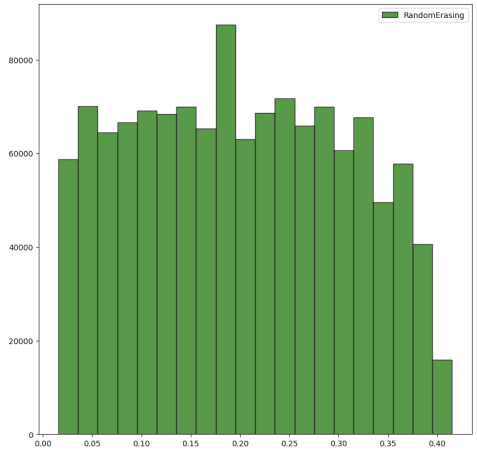
성능이 비슷했고, Hide-and-Seek의 성능이 가장 좋지 않았다. 그림[10]~그림[12]는 2,500,000 횟수에 대한 지우는 비율을 그래프로 나타낸 것이다. Random Erasing 방식을 개선한 이유는 Cutout과 마찬가지로 성능이 좋았으며, 그림[10]을 통해 알 수 있듯이 Random Erasing이 상대적으로 지우는 비율의 분포가 균등하여 객체의 다양한 폐색 수준을 만들 수 있다는 장점이 있다.



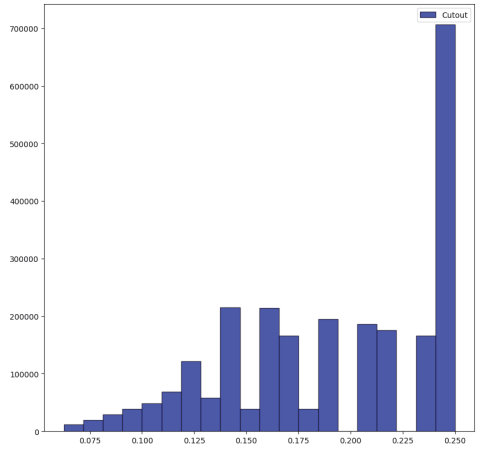

[그림 9] 제안한 지우기 방법

| 지우기 기법 | Random Erasing | Cutout | Hide-and-Seek |
|---|---|---|---|
| error rate | 6.75 | 6.68 | 6.95 |

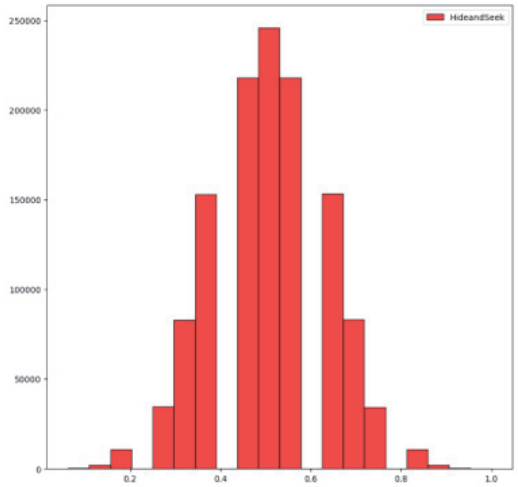
[표 1] Resnet20 및 cifar10에 대한 성능 비교, error rate(5번 평균)




[그림 10] Random Erasing 지우기 비율 분포




[그림 11] Cutout 지우기 비율 분포




[그림 12] Hide-and-Seek 지우기 비율 분포

## 3. Class Activation Mapping(CAM)

### 1) CAM

CAM(Class Activation Mapping)[24]은 딥 러닝 모델, 특히 컨볼루션 신경망(CNN)의 피쳐맵(Feature Map)의 GAP(Global Average Pooling)을 적용하여 특정 클래스 이미지의 히트맵(Heat Map)을 생성한다. 피쳐맵이 GAP를 통해서 각 채널이 하나의 대푯값으로 계산되고, 대푯값이 FCL(Fully Connected Layer)를 통해서 특정 클래스에 대한 가중치를 곱하여 학습을 하게 된다. 이렇게 학습된 특정 클래스에 대한 가중치를 각 피쳐맵 채널에 곱하고 채널을 합산하면, 특정 클래스에 대한 중요도를 나타내는 히트맵이 된다.

CAM을 통해서 CNN모델이 각 클래스에 대해서 이미지의 어디 부분을 중점적으로 살펴보는지 알 수 있다. 하지만 CAM은 구조적으로 문제가 있다. GAP와 FCL가 필수적이므로 GMP(Global Max Pooling), strided CNN등을 사용하는 경우, 적용이 불가능하다.

### 2) Gradient weighted-CAM

기존 CAM의 GAP에 종속적인 문제를 해결하기 위해 나온 방법이 Grad-CAM(Gradient weighted-CAM)[25]이다. Grad-CAM은 CAM의 일반화

---

24) Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

25) Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626)

된 버전이다. CAM은 GAP을 FCL을 통하여 각 클래스에 대한 가중치를 얻는 반면에 Grad-CAM은 convolution layer에서 만들어진 특징 맵(Feature Map)과 클래스에 대한 기울기를 가중치로 활용한다. 이렇게 산출된 가중치를 CAM과 마찬가지로 피쳐맵과 곱하고 합산하여 얻는다.

$$w_k^c = \frac{1}{i+j}\sum_i\sum_j \frac{\partial s^c}{\partial F_{ij}^k} \quad (3)$$

$$M^c = ReLU(\sum_k w_k^c F^k) \quad (4)$$

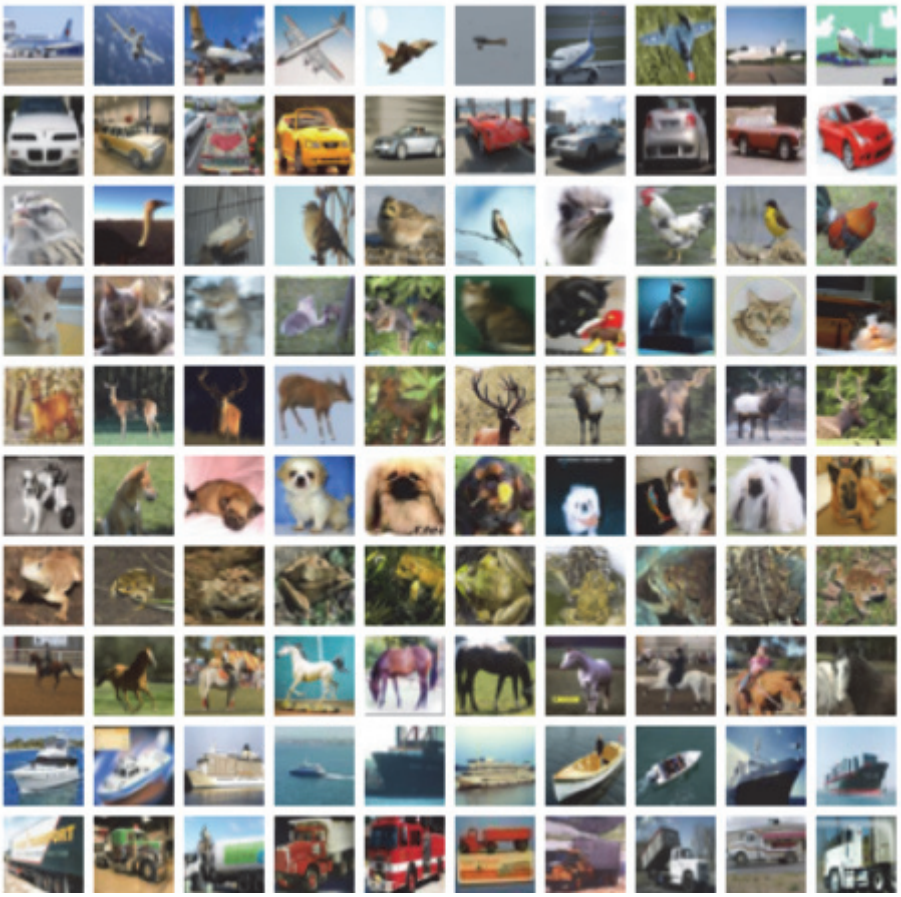
식(3)과 식(4)는 Grad-CAM의 계산식이다. 식(3)에서 스코어 $s^c$는 특정 클래스의 값을 의미 하고, $F_{ij}^k$는 우리가 관찰하고 싶은 특정 CNN 레이어이다. 식(3)을 통해서 스코어 $s^c$와 특징 맵 $F_{ij}^k$의 기울기를 구하고, 이를 평균을 내면 CAM에서의 GAP에 대한 FCL의 가중치를 구하는 것과 같은 효과를 낸다. 특징맵 $F_{ij}^k$에 대해서 설명하자면, $k$는 채널이고 $i,j$는 너비와 높이를 나타낸다. 이 과정을 통해서 얻어진 채널별 가중치는 $w_k^c$이다. 식(4)는 앞서 얻은 가중치를 이용하여 클래스에 대한 활성화맵 $M^c$를 얻는 것이다. 각 채널별로 가중치 $w_k^c$과 특징맵 $F^k$을 곱하고 합산하다. 이를 통해서 특징맵의 채널은 1로 압축되어 클래스에 대한 공간적 중요도를 나타내는 히트맵이 된다. 다만, 히트맵의 네거티브 영역은 다른 클래스나 배경을 나타내는 정보로 간주하고 ReLU를 사용한다. 본 논문에서 제안하는 지우기 기법에서는 특정 구조가 필요없이 모든 CNN구조에서 적용 가능한 Grad-CAM을 사용했다.

## 4. 실험 세팅 및 결과 비교

### 1) 실험 세팅

이미지 분류 작업을 통해서 실험을 세팅 하였다. 데이터셋은 CIFAR10데이터셋이고 모델은 Resnet20, Resnet32, Resnet44이다. CIFAR !0은 32x32 컬러이미지 60,000개, 클래스는 10개로 구성되어 있다. 각 클래스마다 6,000개가 있으며, 50,000개는 훈련 데이터 세트이고, 10,000개는 테스트 데이터셋이다. 훈련 데이터는 기본적으로 무작위로 수평 뒤집기, 4로 패딩후 32x32 무작위 자르기(Crop)를 하였다. 실험의 성능 검증은, Top-1 error rate로 오답률을 나태낸다. error rate는 5번 실험의 평균값이고 낮을수록 좋은 성능이다.

서버의 사양은 다음과 같다. 운영체제는 ubuntu 22.04 LTS, CPU는 Intel(R) Core (TM) i9-10900X CPU @ 3.70GHz, 메모리는 128GB DDR4, GPU는 NVIDIA GeForce RTX 3090 24GB이다. 딥러닝 관련 소프트웨어는 Pytorch, Python을 사용했다. 최적화 관련 세팅으로는 SGD(Stochastic Gradient Descent)을 사용했고, 학습률 스케쥴러는 MultiStepLR을 사용하여 150, 225 에포크마다 0.1 비율로 학습률이 감소한다. 추가적인 손실함수 관련 세팅으로는 가중치 감쇠을 0.0005로 설정했다. 훈련 배치 크기는 128이지만, 150epoch부터는 증강된 데이터를 포함하여 256으로 훈련된다.



[그림 13] CIFAR10 데이터 예시

2) 결과 비교

표[2]는 다양한 Resnet 모델과 지우기 방법들에 대한 성능 비교이다. 모델들은 Resnet20, Resnset32, Resnet44로 진행되며, 모두 동일하게 CIFAR10 데이터셋으로 4.1 실험 세팅과 똑같이 진행된다. 제안된 기법들은 처음부터 적용되어 128배치 크기로 진행된다. Random Erasing의 지울 확률은 0.5, Cutut의 지울 확률은 1.0, Hide-and-Seek은 지울 확률 0.5와 각 패치에 대

한 확률 0.5이다. 대체적으로 Hide-and-Seek에 비하여 Random Erasing과 Cutout의 성능이 좋다. Resnet20에서는 기존 방법중 Cutout의 error rate가 6.68로 가장 좋고, 제안한 방법은 6.40으로 0.28로 일반화 성능이 더 증가한다. Resnet32 ,44에서도 기존 방법 중 Cutout의 성능이 가장 좋고, 각각 5.89, 5.33이다. 제안한 방법의 성능은 5.56, 5.15로 Cutout에 비교하면, 0.33, 0.17으로 성능이 향상된다. 결과적으로 제안한 방법을 사용하면 기존의 지우기 기법에 비하면 효과적으로 일반화 성능을 높일 수 있다.

표[3]는 Resnet20에 대한 하이퍼파리미터 $b_h$에 대한 결과이다. $b_h$가 작을수록 가장 활성화 된 부분 근처를 폐색하게 된다. 반면 값이 클수록 객체 근처를 폐색할 확률이 떨어진다. $b_h$가 값이 2일 때 가장 성능이 좋았다. 값이 커질수록 일반 Random Erasing에 가까워지기에 성능이 하락한다고 볼 수 있다.

표[4]는 CAM을 얻는 방법에 대한 비교이다. '학습 중 CAM'은 본 논문에서 제안하는 방법으로 학습 중인 모델을 이용해서 CAM을 얻는 방법이다. '학습 후 CAM'은 다른 지우기 증강법 없이 학습이 완료된 기본 모델을 통해서 먼저 CAM을 얻고 해당 CAM을 통해서 객체 근처를 지우는 방법이다. Resnet20으로 실험을 진행하였고, 성능상 크게 차이가 없었다. 하지만 학습 후 방법을 사용하기 위해서는 먼저 기본 모델을 학습하고, 다른 모델을 학습해야 하는 단점이 있다.

| | 제안 기법 $b_h$=2 | Rasndom Erasing | Cutout | Hide-and-Seek |
|---|---|---|---|---|
| Resnet20 | 6.40 | 6.75 | 6.68 | 6.95 |
| Resnet32 | 5.56 | 6.00 | 5.89 | 5.99 |
| Resnet44 | 5.15 | 5.34 | 5.33 | 5.47 |

[표 2] 지우는 방법에 대한 각 모델별 성능 비교

| $b_h$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Error Rate | 6.43 | 6.43 | 6.40 | 6.54 | 6.51 | 6.52 |

[표 3] 하이퍼 파리미터 $b_h$에 대한 결과

| 방법 | 학습 중 CAM | 학습 후 CAM |
|---|---|---|
| error rate | 6.40 | 6.44 |

[표 4] CAM을 얻는 방법에 대한 비교

# Ⅳ. 결론

## 1. 요약

본 논문은 데이터 기본 증강 법 중, 지우기 기법에 대한 새로운 방법을 제시하였다. 기존의 지우기 방법은 무작위로 지우기에 객체에 대한 다양한 수준의 폐색을 제공하지 못한 단점이 있다. Random Erasing과 Cutout의 경우에는 객체와 상관없는 배경을 지우는 경우가 발생하고, Hide-And-Seek의 경우에도 객체를 제외하고 배경만 지울 때가 있다. Grad-CAM을 활용하면, 추가적인 모델없이 이미지 속 관심있는 객체를 효과적으로 지울 수 있다. 이를 활용하여, 비교적 다양한 수준의 폐색을 고르게 지우는 Random Erasing에 적용하여 모델의 일반화 성능을 높였다. 지우기 증강법이 적용하는 시점은 학습 중 중간 에포크로 하여, 모델이 데이터에 대한 특징을 충분히 학습하도록 했다. 그 후 증강된 데이터를 활용한 학습을 통해서 기존의 지우기 기법의 성능을 능가했다.

## 2. 한계 및 향후 연구

본 논문은 ResNet20, ResNet32, ResNet44에 대한 실험을 진행하여 보다 복잡한 모델에 대한 검증을 하지 못했다. 이미지넷에 대한 ResNet18의 경우 파라미터 수가 11.7M이다. ResNet20, 32, 44는 각각 0.27M. 0.46M. 0.66M이다. 본 논문에서는 자세히 다루지 않았지만, ResNet18을 사용할 시, Hide-and-Seek의 성능이 Random Erasing보다 좋았다. CAM을 활용하려고

했을 때, 마지막 피쳐맵 사이즈의 크기가 4x4로 작아서 CAM을 활용한 실험을 진행하지 못했다. 모델의 크기와 이미지 데이터 크기에 대한 검증이 더 필요하다.

# 참 고 문 헌


1) Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. ArXiv, abs/2010.11929.

2) Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11966-11976.

3) Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. The MIT Press.

4) Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2017). Random Erasing Data Augmentation. ArXiv, abs/1708.04896.

5) Wan, L., Zeiler, M.D., Zhang, S., LeCun, Y., & Fergus, R. (2013). Regularization of Neural Networks using DropConnect. International Conference on Machine Learning.

6) Ba, J., & Frey, B.J. (2013). Adaptive dropout for training deep neural networks. Neural Information Processing Systems.

7) Zeiler, M.D., & Fergus, R. (2013). Stochastic Pooling for Regularization of Deep Convolutional Neural Networks. CoRR, abs/1301.3557.

8) Xie, L., Wang, J., Wei, Z., Wang, M., & Tian, Q. (2016). DisturbLabel: Regularizing CNN on the Loss Layer. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4753-4762.

9) Kang, G., Dong, X., Zheng, L., & Yang, Y. (2017). PatchShuffle Regularization. ArXiv, abs/1707.07103.

10) Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. arXiv preprint arXiv:2204.08610.

11) Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020, April). Random erasing data augmentation. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 13001-13008).

12) DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural

networks with Cutout. arXiv preprint arXiv:1708.04552.

13) Singh, K. K., Yu, H., Sarmasi, A., Pradeep, G., & Lee, Y. J. (2018). Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. arXiv preprint arXiv:1811.02545.

14) Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

15) Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6023-6032).

16) Kim, J. H., Choo, W., & Song, H. O. (2020, November). Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In International Conference on Machine Learning (pp. 5275-5285). PMLR.

17) Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501.

18) Kuo, C. W., Ma, C. Y., Huang, J. B., & Kira, Z. (2020). Featmatch: Feature-based augmentation for semi-supervised learning. In Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XVIII 16 (pp. 479-495). Springer International Publishing.

19) Li, B., Wu, F., Lim, S. N., Belongie, S., & Weinberger, K. Q. (2021). On feature normalization and data augmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12383-12392).

20) Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

21) Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

22) Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp.

8789-8797).

23) Wang, X., Shrivastava, A., & Gupta, A. (2017). A-fast-rcnn: Hard positive generation via adversary for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2606-2615).

24) Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

25) Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626)

# ABSTRACT

## Image processing based data augmentation for deep learning

Juhyeon Oh
Department of Future Convergence Technology Engineering
Graduate School of
Sungshin University

This paper presents a novel approach within the domain of basic data augmentation techniques, specifically focusing on erasing methods. Traditional erasing methods suffer from a drawback of randomly erasing areas, failing to provide various levels of occlusion for objects. Techniques like Random Erasing and Cutout sometimes erase irrelevant background instead of the object, and Hide-And-Seek may also only erase the background, excluding the object. Utilizing Grad-CAM enables the effective erasure of objects of interest within images without requiring additional models. By applying this to Random Erasing, we achieved a more uniform level of various occlusions, thereby enhancing the model's generalization performance. The point of applying the erasing augmentation method was set during the middle epochs of training,

allowing the model to sufficiently learn the features of the data. Subsequently, training with augmented data surpassed the performance of the traditional erasing methods.