

# 이미지 분류 문제를 위한 능동적 학습 관점의 자동 데이터 확장

조윤기<sup>0</sup>, 김재윤, 윤성의

카이스트 전산학부

yoongi@kaist.ac.kr, jaeyoon1603@gmail.com, sungeui@kaist.edu

## 요약

컴퓨터 비전 분야에서의 데이터 확장은 심층 신경망의 성능을 향상시키는 효과적인 기술이다. 최근 데이터 확장을 자동화하기 위하여, 데이터 셋의 특성을 파악한 뒤 어떤 확장 기법을 활용할 것인지 자동으로 선택해주는 연구들이 등장했으며, 이 기법들은 심층 신경망 모델의 성능을 유의미하게 향상시켰다. 그러나 기존의 방법론들은 높은 컴퓨팅 비용이 필요하며, 각 데이터의 특성을 고려하지 않고 모든 데이터에 대하여 동일한 데이터 확장 기법을 적용한다. 본 연구에서는 역전과 기법을 통해 쉽게 학습될 수 있고, 각 이미지의 특성을 고려하여 각기 다른 확장 기법을 적용한 이미지를 생성하는 기법을 제안한다.

## 1. 서론

심층 학습은 다양한 분야에서 성능 향상을 이루어 냈으며, 컴퓨터 비전 관련 작업도 심층 신경망을 통한 많은 발전이 있었다. 심층 신경망을 학습시키기 위해서는 많은 데이터가 필요하며, 능동적 학습(Active Learning), 데이터 확장(Data Augmentation)과 같은 효과적인 데이터 수집 및 활용에 대한 연구가 활발하게 이루어지고 있다. 능동적 학습 [9], [10], [13]은 어떤 데이터를 우선 레이블링하여 학습 데이터로 활용할 것인지, 데이터 선택과 관련된 연구 분야이다. 데이터 확장은 이미지 회전 및 자르기와 같은 기법을 통하여 심층 신경망의 성능을 향상시키는 기법이다. 하지만 데이터 확장은 데이터 셋과 작업의 특성에 따라 적합한 확장 기법을 찾기 위해 전문가의 경험과 지식이 필요하다. 이에 따라, 데이터 셋과 작업에 적합한 데이터 확장 기법을 자동으로 적용하는 연구가 활발하게 진행 중이다. 성능 향상에 적합한 데이터 확장 정책을 검색하여 적용하는 자동화된 기계학습(Automated Machine Learning, AutoML) 기반의 연구 [2], [8]가 있다. 이는 사용자의 재량에 따른 확장 기법 선택 없이 컴퓨터 비전 작업의 성능을 향상시켰으나, 큰 계산 비용을 요구하며, 각 데이터의 특성을 고려하지 못하고 모든 데이터 셋에 대해 동일한 데이터 확장 기법을 적용한다는 한계점이 있다. 생산적 적대 신경망(Generative Adversarial Networks, GAN) 기반의 자동 데이터 확장 연구 [1], [15]도 있으며, 어떠한 데이터 확장 기법을 통해 이미지가 생성되었는지 알 수 없다는 한계점이 존재한다.

본 연구에서는 역전과 기법을 통해 쉽고 빠르게 학습할 수 있는 자동 데이터 확장 기법을 제안한다. 자동화된 기계학습 기반의 자동 데이터 확장과 달리, 제안한 방법은 각 이미지의 특성을 고려하여 각각 적합한 다른 데이터 확장 기법이 적용된 이미지를 생성한다. 이 기법은 회전, 자르기 등 어파인 변환(Affine Transform) 기반의 데이터 확장 기법을 적용할 수 있으며, 이미지별로 어떠한 데이터 확장 기법이 적용되었는지 알 수 있다. 이 기법은 본 연구에서 제안한 새넨 엔트로피(Shannon Entropy) [11] 기반의 손실 함수를 통한 학습 과정을 거친다.

## 2. 관련 연구

심층 학습은 많은 데이터를 요구하며, 데이터 문제와 관련된 많은 연구가 활발히 진행되고 있다. Cutout [3]과 Mixup [14] 등 이미지 인식을 위한 데이터 확장 기법은 심층 신경망의 일반화를 향상시키는 것에 기여하였다.

데이터 셋과 이를 통해 수행하는 작업의 특성을 고려한 자동 데이터 확장에 관한 연구가 있다. AutoAugment [2]는 저자가 정의한 검색 공간에서 작업의 성능 향상을 위해 데이터 확장 정책을 자동으로 검색하는 기법이다. 이는 검색 공간을 사용자가 정의하여 주기 때문에 완벽한 자동화가 아니며, 최적의 정책을 검색하는데 수천 시간의 그래픽 처리 장치 연산이 요구된다. Fast AutoAugment [8]는 효과적인 검색 기법과 베이지안 최적화를 활용하여 높은 계산 비용 문제를 해결하였다. 그러나 검색된 데이터 확장 정책은 전체 데이터 셋을 통하여 선택되므로, 각 이미지에 적절한 데이터 확장을 적용하기 어

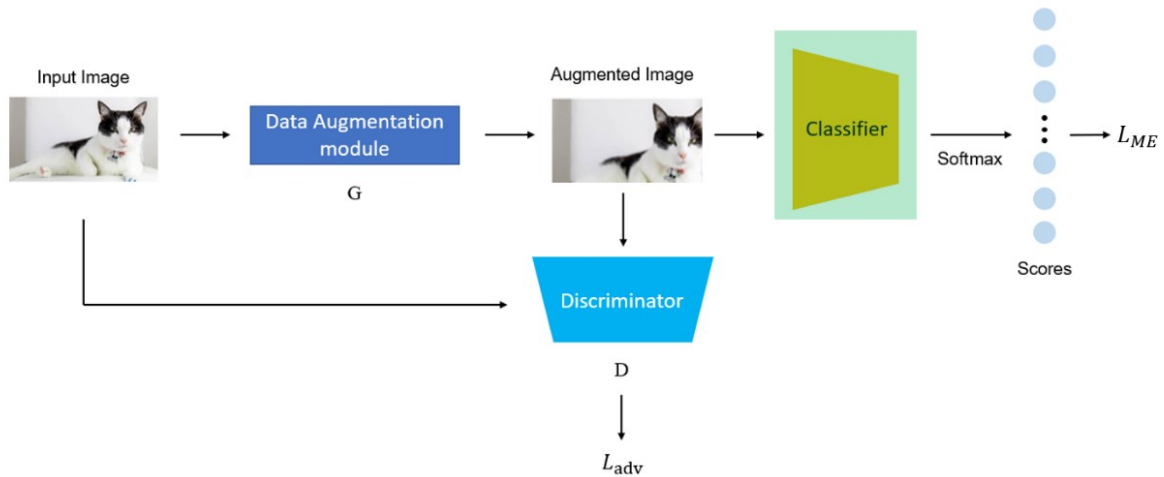


그림 1. 제안 기법의 데이터 확장 모듈 학습 과정 다이어그램. 초록색 사각형은 최대 엔트로피 손실 함수를 통해 학습되지 않고 기존의 가중치를 유지하는 것을 의미한다.

렵다는 한계점이 남아있다. 생산적 적대 신경망 (GAN)을 통한 자동 데이터 확장 관련 연구[1], [15]도 있으며 이는 데이터 확장 정책을 적용하는 형태가 아닌, 직접적으로 데이터 확장이 적용된 이미지를 생성하는 형태이다.

심층 신경망 훈련을 위해 학습에 유효한 데이터 표본을 뽑아 우선으로 학습에 활용하는 데이터 선택과 관련된 연구가 있다. 능동 학습(Active Learning)은 레이블링 되지 않은 데이터 셋에서 어떠한 기준으로 데이터 표본을 뽑아 레이블링하여 학습 데이터로 활용할 것인지에 대한 연구로, [13]은 심층 신경망 모델의 손실 함수 값을 예측하는 모듈을 활용하여 높은 손실 값이 예측되는 데이터를 우선적으로 학습에 활용하였다. 이 방법은 일반적으로 사용하는 랜덤 샘플링뿐만 아니라 엔트로피 샘플링 등 기존에 연구된 기법보다 심층 신경망의 성능 향상에 효과적이었다. 이러한 관점으로 우리는 공간적 변환 네트워크 (Spatial Transformer Network, STN)을 활용해, 학습에 유효한 데이터가 되도록 자동으로 데이터 확장을 적용하는 기법을 제안한다.

### 3. 능동적 학습 관점의 자동 데이터 확장

이 장에서는 제안한 능동적 학습 관점의 자동 데이터 확장 기법에 대해 학습 순서대로 소개한다. 우리의 목표는 입력 이미지가 분류기의 성능 향상에 유효한 이미지가 되도록 자동으로 데이터 확장시키는 것이다. 제안한 자동 데이터 확장 기법은 총 3 단계의 학습 과정으로 구성된다. 먼저, 분류기를 교차 엔트로피(Cross Entropy) 손실 함수를 통해 학습시킨다. 다음으로, 본 연구에서 제안한 선택 엔트로피 기반의 최대 엔트로피 손실 함수(Max Entropy Loss)와 적대적 손실 함수(Adversarial Loss)를 통하여 데이터 확장 모듈(Data Augmentation Module)과 선별기(Discriminator)를 학습시킨다. 마지막으로 선별기를 제외한 뒤, 데이터 확장 모듈에서 생성된 이미지를 통해 분류기를 미세조정(Fine-tune)하여 성능을

향상시킨다.

#### 3.1 데이터 확장 모듈 (Data Augmentation Module)

데이터 확장 모듈은 공간적 변환 네트워크 (Spatial Transformer Network, STN)[6]와 동일한 구조를 가진다. 데이터 확장 모듈은 회전, 자르기, 기울임 등의 어파인 변환 기반의 데이터 확장 기법을 적용할 수 있다. 데이터 확장 모듈은 입력 이미지와 지역화 네트워크(Localization Network)를 통해 어파인 변환 매개 변수를 출력한다. 이후 이 매개 변수를 통하여 출력 이미지에 해당하게 되는 격자를 입력 이미지에 투영한 뒤, 입력 이미지의 샘플링 격자 (Sampling Grid)로부터 출력 이미지를 샘플링한다. 데이터 확장 모듈은 STN 과 동일한 구조이나, STN 을 활용하는 다른 네트워크와 다른 학습 과정을 가진다.

#### 3.2 최대 엔트로피 손실 함수 (Max Entropy Loss)

데이터 확장 모듈이 분류기의 성능 향상에 유용한 확장 이미지를 생성하기 위한 손실 함수를 제안한다. 선택 엔트로피(Shannon Entropy)는 능동적 학습 기법 중 하나 인 최대 엔트로피 샘플링 [12]에 사용된다. 엔트로피가 높은 데이터를 순서대로 학습했을 때 무작위 샘플링으로 학습했을 때보다 분류기의 성능이 더 향상되었다. 이러한 연구에서 영감을 받아 분류기의 분류 확률 분포의 엔트로피를 증가시키는 최대 엔트로피 손실 함수(Max Entropy Loss)을 제안한다.

입력 이미지  $I$ 가  $m$ 개의 클래스를 분류하는 분류기를 통과한 뒤, 소프트맥스(Softmax) 함수를 적용한 후의 확률 분포를  $P$ 라고 한다.  $p_i$ 는 각 클래스에 해당하는 확률 값이며, 이 때 엔트로피  $H$ 의 최대값은 아래와 같다.

$$P = \{p_i(I)\}, \quad i \in (\{1, \dots, m\})$$

$$H(I) = -\sum_i^m p_i(I) \log_2 p_i(I) \leq \log_2 m$$

최대 엔트로피 손실 함수와 이의 목적 함수는 아래와 같이 정의할 수 있다. Q는 입력 이미지 데이터 셋의 확률 분포를 의미한다.

$$L_{ME}(I) = -\log([\log_2 m - H(I)])$$

$$J(\theta_G|I') = \mathbb{E}_{I \sim Q}[L_{ME}(I)]$$

데이터 확장 모듈이 최대 엔트로피 손실 함수에 의해서만 학습된다면, 현실적이지 않거나 유효하지 않은 이미지를 생성하는 문제가 발생할 수 있다. 이 문제를 해결 하기 위해 데이터 확장이 적용된 이미지와 기본 이미지를 구분하는 식별기(Discriminator)를 활용한다. 데이터 확장 모듈과 식별기가 적대적 손실 함수(Adversarial Loss)를 통해 학습하면서, 데이터 확장 모듈은 현실적이고 높은 엔트로피를 갖는 데이터 확장을 할 수 있게 된다. 이 확장된 이미지는 분류기에게 유효하고 도전적인 데이터로 작용함으로써 분류기의 일반성을 향상시키는데 사용될 수 있다. 데이터 확장 모듈과 식별기를 학습할 때에는 분류기는 학습되지 않으며 본래 가중치 값을 유지한다. 우리의 목적 함수는 아래와 같으며, G, D는 각각 데이터 확장 모듈과 식별기를 의미하며,  $I'$ 은 데이터 확장 모듈의 출력 이미지를 뜻한다.  $\lambda$ 는 적대적 손실 함수에 대한 가중치이다.

$$\min_G \max_D V(D, G)$$

$$= \mathbb{E}_{I \sim P} [\log D(I) + \log(1 - D(G(I')))] + \lambda J(\theta_G|I')$$

$$= \mathbb{E}_{I \sim P} [\log D(I) + \log(1 - D(I')) + \lambda L_{ME}(I')]$$

따라서, 데이터 확장 모듈과 식별기는 각각 아래의 기울기의 역전과를 통해 경사 하강법으로 학습할 수 있다.

$$\nabla_{\theta_g} V(D, G) = \nabla_{\theta_g} \frac{1}{n} \sum_{i=1}^n [\log(1 - D(I'^{(i)})) + L_{ME}(I'^{(i)})]$$

$$\nabla_{\theta_d} V(D, G) = \nabla_{\theta_d} \frac{1}{n} \sum_{i=1}^n -[\log D(I'^{(i)}) + \log(1 - D(I'^{(i)}))]$$

데이터 확장 모듈과 식별기의 학습 과정 다이어그램은 그림 1.을 통해 볼 수 있다.

#### 4. 실험 결과 및 분석

본 연구에서 제안한 자동 데이터 확장 기법의 성능 확인을 위하여 CIFAR-10 데이터 셋에 대하여 실험 및 분석을 진행하였다. 분류기는 ResNet-18[5]을 활용하였으며, 학습을 위해 Adam optimizer[7]를 0.1의 러닝 레이트로 100번 반복 학습하였다.

데이터 확장 모듈 학습을 위하여 Adam optimizer를 0.0001의 러닝 레이트, 식별기 학습을 위하여 SGD optimizer를 0.0001의 러닝 레이트 및  $\lambda=2$ 로 50번 반복 학습하였다. 이후, 데이터 확장 모듈과 함께 분류기를 Adam optimizer와 0.01의 러닝 레이트로 100번 미세 조정 학습을 진행하였다.

표 1.은 CIFAR-10 데이터셋에 대한 제안한 자동 데이터 확장 기법 실험 결과를 나타낸다. 기준치는 ResNet-18 모델을 Adam optimizer로 0.1의 러닝 레이트로 100번, 0.01의 러닝 레이트로 100번 순차적으로 학습한 모델이다.

기준치와 비교하였을 때 성능 향상이 없었음을 확인할 수 있으며, 손실 함수 그래프와 데이터 확장 모듈의 출력 이미지를 통하여 추가적인 분석을 진행하였다.

	정확도
기준치	0.8736
제안 기법	0.8760

표 1. CIFAR-10 데이터 셋에 대한 제안 기법의 실험 결과

#### 4.1 손실 함수 분석

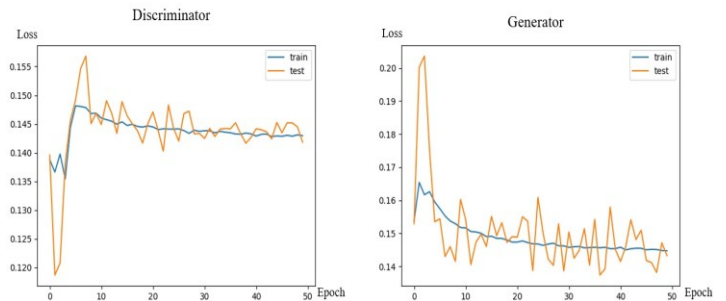


그림 2. 식별기와 데이터 확장 모듈의 학습 과정에 따른 적대적 손실 함수 그래프. 주황색 선은 테스트 데이터 셋에 대한 값을, 파란색 선은 학습 데이터 셋에 대한 값을 나타낸다.

그림 2.는 학습 과정에 따른 데이터 확장 모듈과 식별자의 적대적 손실 함수 값의 그래프를 나타낸다. 그래프를 보면 알 수 있듯, 학습 과정이 진행될수록 식별기의 손실 함수 값은 증가하는 경향성을, 데이터 확장 모듈은 손실 함수 값이 감소하는 경향성을 보이고 있다. 이러한 결과로부터, 적대적으로 경쟁하며 학습하던 과정속에서 데이터 확장 모듈이 식별기를 경쟁에서 이겼으며, 매우 작은 회전이나 항등 변환 등 입력 이미지를 거의 변경하지 않았음을 추측할 수 있다.

#### 4.2 생성기와 식별기 사이의 어려움 비대칭 문제

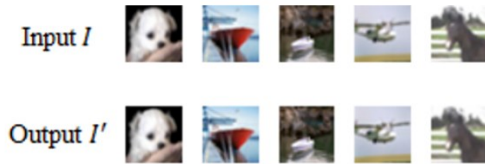


그림 3. 데이터 확장 모듈의 입력 이미지(위)와 출력 이미지(아래)

데이터 확장 모듈의 출력 이미지가 기본 이미지에서 거의 변경이 없을 것이라는 추론을 확인하기 위해, 실제 데이터 확장 모듈을 통해 확장된 이미지를 확인해보았다(그림 3.). 거의 모든 출력 이미지가 입력 이미지와 거의 비슷한 항등 변환의 결과를 보여주고 있다.

이는 생성기(데이터 확장 모듈)과 식별기 사이의 최대-최소 시나리오를 설정하였으나 생성기의 상황이 훨씬 쉬었음으로 해석할 수 있다. 이로 인해 데이터 확장 모듈이 입력 이미지와 동일한 출력 이미지를 생성함으로써 식별기를 경쟁에서 쉽게 이기는 결과를 보여주었다. 이러한 한계점을 극복하기 위해 좀 더 자유도가 낮은 손실 함수를 제시하거나, 데이터 확장 모듈의 지역화 네트워크의 출력 매개 변수를 정규화하고 식별기를 제외하는 환경 구축 등의 해결 방안이 존재한다.

## 5. 결론

본 연구에서는 각 이미지의 특성을 고려하여 각 데이터에 적합한 확장 기법을 적용할 수 있는 최대 엔트로피 손실 함수와 학습 기법을 제시하였다. 최대 엔트로피 손실 함수는 새넨 엔트로피에 의해 정의되며 데이터 확장 모듈이 분류기에기 유효하고 도전적인 데이터를 생성할 수 있게 한다.

하지만 데이터 확장 모듈과 식별기 사이의 어려움 비대칭 문제가 존재하며, 이로 인해 데이터 확장 모듈이 입력 이미지와 동일한 출력 이미지를 생성하는 한계점을 찾을 수 있었다. 최대 엔트로피 손실 함수보다 자유도가 낮은 손실 함수를 제시하고, 데이터 확장 모듈의 지역화 네트워크의 출력 매개 변수를 정규화하는 방법으로, 이 문제를 해결하기 위한 계획에 있다.

## 감사의 글

이 논문은 SW 컴퓨팅산업원천기술개발사업(SW 스타랩, IITP-2015-0-00199)의 지원으로 수행되었음.

## 참고문헌

[1] A. Antoniou, A. Storkey, and H. Edwards. Data

augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340, 2017.

- [2] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. Autoaugment: Learning augmentation strategies from data. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
- [3] T. DeVries and G. W. Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552, 2017.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.
- [5] 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, pages 770–778, 2016.
- [6] M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In Advances in neural information processing systems, pages 2017–2025, 2015.
- [7] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412. 69 80, 2014.
- [8] S. Lim, I. Kim, T. Kim, C. Kim, and S. Kim. Fast autoaugment. arXiv preprint arXiv:1905.00397, 2019.
- [9] W. Luo, A. Schwing, and R. Urtasun. Latent structured active learning. In Advances in Neural Information Processing Systems, pages 728–736, 2013.
- [10] O. Sener and S. Savarese. Active learning for convolutional neural networks: A core-set approach. arXiv preprint arXiv:1708.00489, 2017.
- [11] C. E. Shannon. A mathematical theory of communication. Bell system technical journal, 27(3):379–423, 1948.
- [12] M. C. Shewry and H. P. Wynn. Maximum entropy sampling. Journal of applied statistics, 14(2):165–170, 1987.
- [13] D. Yoo and I. S. Kweon. Learning loss for active learning. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
- [14] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017.
- [15] X. Zhu, Y. Liu, Z. Qin, and J. Li. Data augmentation in emotion classification using generative adversarial networks. arXiv preprint arXiv:1711.00648, 2017.