

# 그룹 정보를 사용한 Dataset Distillation 의 편향 제거에 관한 연구

김세환, 이정우  
서울대학교

{sehwankim, junglee}@snu.ac.kr

## A Study on the Debiased Dataset Distillation by utilizing group annotations

Sehwan Kim, Jungwoo Lee  
Seoul National Univ.

### 요 약

최근 MNIST, CIFAR10, ImageNet 과 같은 다양한 데이터셋들이 딥러닝의 네트워크 학습에 사용되고 있다. 하지만 네트워크 학습에 필요한 데이터셋의 크기가 늘어남에 따라 증가하는 네트워크 학습 시간, 데이터셋의 저장 공간 등의 문제가 발생하고 있다. 이를 해결하기 위해 제안된 방법이 Dataset Distillation 이다. Dataset Distillation 은 네트워크를 고정된 상태로 유지하고, 대신 큰 데이터셋으로부터 지식을 추출하여 작은 데이터셋을 만드는 방법이다. 만들어진 작은 데이터셋을 통해 네트워크를 학습시켜 큰 데이터셋을 통해 학습시킨 네트워크만큼의 성능을 얻고자 한다. 하지만 이 과정에서 큰 데이터셋에 편향이 있다면 만들어진 작은 데이터셋에도 편향이 있음을 실험을 통해 확인했다. 이러한 편향을 Dataset Distillation 에서 제거하기 위해 그룹 정보를 활용한 학습 방법을 제안한다.

### I. 서론

본 논문에서는 그룹 정보를 활용하여 기존의 Dataset Distillation 과정에서 생기는 편향을 없애, 편향이 제거된 큰 데이터셋의 지식을 물려받은 비교적 크기가 작은 데이터셋을 합성하는 학습 방법을 제안한다. 딥러닝이 발전하면서 다양한 목적의 네트워크들이 학습되고 있고, 이에 따라 용량이 매우 큰 데이터셋들이 네트워크의 학습에 사용되고 있다. 데이터셋의 크기가 늘어남에 따라 네트워크 학습 시간, 데이터셋의 저장 공간 및 메모리 등이 증가하는 문제가 있다. 이를 해결하기 위해 제안된 방법이 Dataset Distillation 으로, 큰 데이터셋의 지식을 추출하여 작은 데이터셋을 만드는 방법이다. 만들어진 작은 데이터셋으로 네트워크를 학습시켰을 때 큰 데이터셋을 통해 학습시킨 네트워크만큼의 성능을 얻도록 하는 것이 Dataset Distillation 의 목적이다. 하지만 큰 데이터셋에 편향이 있다면, 큰 데이터셋의 지식을 물려받은 작은 데이터셋 또한 편향이 존재한다. 이러한 현상을 실험을 통해 확인하고, 본 논문에서 그룹 정보를 활용하여 편향이 제거된 Dataset Distillation 학습방법을 제안한다.

### II. 본론

본 논문에서는 그룹 정보를 활용하여 편향이 제거된 Dataset Distillation 학습 방법을 제안한다.

Dataset Distillation 은 그림 1 에서 나타낸 바와 같이, 사이즈가 큰 데이터셋의 지식을 물려받은 비교적 사이즈가 많이 작은 데이터셋을 만드는 방법이다. 작은 데이터셋은 큰 데이터셋의 지식을 물려받았기에, 작은 데이터셋으로 네트워크를 학습시켰을 때 성능이 큰 데이터셋으로 학습시켰을 때 나오는 성능에 근접하고자 하는 것이 Dataset Distillation 의 목적이다. 하지만 이 과정에서 작은 데이터셋은 큰 데이터셋의 지식을 그대로 물려받기 때문에, 큰 데이터셋에 편향이 있다면 작은 데이터셋에도 편향이 생길 수밖에 없다. 이를 제거하기 위해 그룹 정보를 활용한 학습 방법을 제안한다.

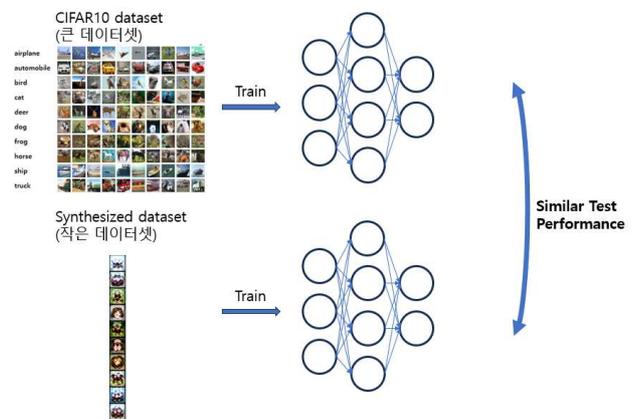


그림 1. Dataset Distillation

## Debiased Dataset Distillation

앞서 언급한 바와 같이, 그룹 정보를 활용하여 Dataset Distillation 과정에서의 편향을 없애고자 한다. 그룹 정보란 데이터셋의 라벨과 편향에 따라 새로 만든 라벨이다. 작은 데이터셋을 학습시킬 때, 기존의 Dataset Distillation 은 배치 내에 있는 같은 클래스의 데이터를 모아서 loss 를 구했다. 하지만 Debiased Dataset Distillation 은 클래스보다 더 자세하게 나누어진 그룹 정보로 각 데이터를 나눠서, 각각 loss 를 구했다. 예를 들어 CMNIST 의 경우, 기존의 Dataset Distillation 은 0~9 까지의 같은 클래스의 데이터끼리 학습시킨다면 Debiased Dataset Distillation 은 (0, red), (0, blue), (0, green), ..., (9, purple)와 같은 그룹 정보로 각 데이터를 나눠서 loss 를 구했다.

## 실험 환경 및 결과

우선 데이터셋으로 CMNIST 를 사용하였다. CMNIST 란 기존의 MNIST 데이터셋에 색을 입힌 데이터셋을 의미한다. 색을 입힐 때, 보통 숫자마다 다른 색을 입히는데 예를 들어 0 은 빨간색, 1 은 초록색으로 이루어져 있다. 이처럼 CMNIST 는 숫자에 색깔이 편향으로 끼어 있는 데이터셋이다. CMNIST 는 학습 데이터로 45000 장, 검증 데이터로 5000 장, 테스트 데이터로 10000 장으로 이루어져 있다. 그리고 Dataset Distillation 에 사용되는 인공신경망으로는 ConvNet 을 사용했다. 그리고 한 클래스 별로 10 개의 데이터를 생성했다. 즉, CMNIST 의 경우, 총 100 개의 데이터를 생성하여 작은 데이터셋을 이루었다. 학습 배치 크기는 256, training iteration 은 1000 으로 고정하고 실험을 진행했다. 또한, 다양한 Dataset Distillation 방법이 존재하는데, 그중 가장 대표적인 gradient matching 을 통한 Dataset Distillation 을 framework 로 사용했다.

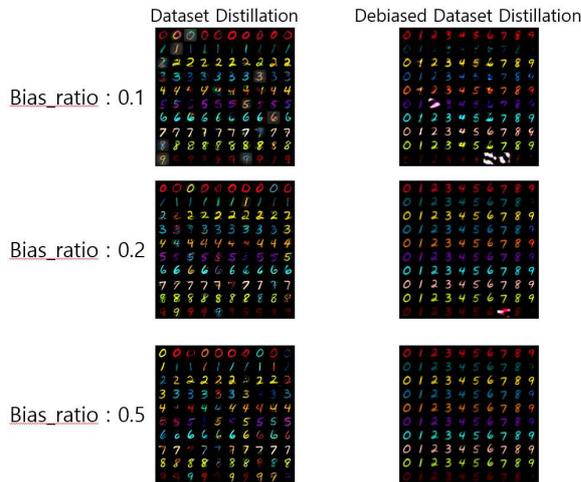


그림 2. Bias\_ratio 에 따른 실험 결과

Bias\_ratio 는 전체 학습 데이터셋중에 색에 대한 편향이 긴 데이터의 비율이라고 볼 수 있는데, 예를 들어 Bias\_ratio 가 0.1 이면 숫자 0 에 대한 데이터의 90%는 빨간색, 10%는 빨간색을 제외한 나머지로 이루어져 있다는 의미이다. 따라서 Bias\_ratio 의 값이 작아질수록 색에 대한 편향이 강력하게 존재한다. 그림 2 에 실험 결과가 나와있는데, 기존의 Dataset Distillation 방법으로 만든 데이터셋의 경우 색에 대한 편향이 강하게 존재함을 확인할 수 있다. 반면에 본 논문에서 제안한 Debiased Dataset Distillation 의 경우, 각 그룹별로

공평하게 데이터를 합성하기에 눈에 보기에도 깔끔하게 합성되었음을 확인할 수 있다. 다만 Bias\_ratio 가 작아질수록 각 그룹에 속하는 학습 데이터의 수가 적어 몇몇 그룹에 대한 합성 데이터가 불완전함을 확인할 수 있었다.

## III. 결론

본 논문에서는 기존의 Dataset Distillation 에 편향이 존재함을 확인하고, 이를 해결하기 위해 그룹 정보를 사용한 Debiased Dataset Distillation 을 제안한다. 실험 결과, 그룹 정보를 활용하여 Dataset Distillation 을 진행하면 더 좋은 질의 합성 데이터를 얻을 수 있음을 확인했다. 다만, bias\_ratio 가 작아질수록 각 그룹에 속하는 학습 데이터의 수가 적어 몇몇 그룹에 대한 합성 데이터가 불완전함을 확인했다.

## ACKNOWLEDGMENT

This work is in part supported by National Research Foundation of Korea (NRF, 2021R1A2C2014504 (30%)), Institute of Information & communications Technology Planning & Evaluation (IITP, 2021-0-00106 (20%), 2021-0-00180 (30%), 2021-0-02068 (20%)) grant funded by the Ministry of Science and ICT (MSIT), INMAC, and BK21-plus.

## 참고 문헌

- [1] Wang, Tongzhou et al. "Dataset distillation." *arXiv preprint arXiv:1811.10959* (2018).
- [2] Krizhevsky Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).
- [3] Sagawa, Shiori, et al. "Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization." *arXiv preprint arXiv:1911.08731* (2019).
- [4] Zhao Bo, Konda Reddy Mopuri, and Hakan Bilen. "Dataset condensation with gradient matching." *arXiv preprint arXiv:2006.05929* (2020).
- [5] Sachdeva Noveen, and Julian McAuley. "Data distillation: A survey." *arXiv preprint arXiv:2301.04272* (2023).