NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

# Long-tailed Classification

## With a Weight-Normalization Based Causal Classifier

Student: An Zheyuan

Supervisor: A/P Zhang Hanwang

## Objective

Datasets tend to become long-tailed as they expand. Therefore, long-tailed classification is crucial to large scale computer vision tasks that rely on large datasets. This project aims to propose a weight-normalization based causal classifier to handle long-tailed classification under a causal framework. Through weight-normalization, the proposed approach is able to disentangle the magnitude and direction of the weight vectors to allow for causal intervention on each of their affects.

## Approach

The proposed training pipeline is end-to-end with a single stage, where the causal classifier takes in the feature vector and feature moving average to calculate an output logit.

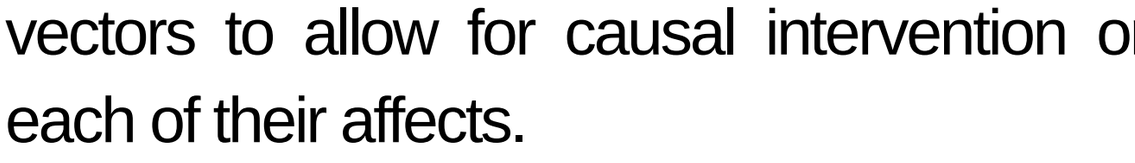


Figure 1: The proposed training pipeline.

The weight-normalization based causal classifier is implemented as a linear classifier with causal norm:

$$Y = \tau \frac{g_i(\boldsymbol{v}_i)^\top \boldsymbol{x}}{\|\boldsymbol{v}_i\|(g_i + \gamma)\|\boldsymbol{x}\|}$$

Equation 1: The proposed classifier. The weight vector has been replaced through weight-normalization.

## Results

The model was trained and tested on long-tailed CIFAR-10 and CIFAR-100 datasets with different class imbalance ratios. The performance was compared with previous methods under the same settings.

Table 1: Top-1 accuracy on long-tailed CIFAR-10/100 datasets

| Dataset | Long-tailed CIFAR-100 | | | Long-tailed CIFAR-10 | | |
|---|---|---|---|---|---|---|
| Imbalance ratio | 100 | 50 | 10 | 100 | 50 | 10 |
| Focal Loss | 38.4 | 44.3 | 55.8 | 70.4 | 76.7 | 86.7 |
| Mixup | 39.5 | 45.0 | 58.0 | 73.1 | 77.8 | 87.1 |
| LDAM | 42.0 | 46.6 | 58.7 | 77.0 | 81.0 | 88.2 |
| BBN | 42.6 | 47.0 | **59.1** | 79.8 | 82.2 | **88.3** |
| (Proposed) Weight-norm | **44.8** | **49.7** | 58.4 | **81.5** | **85.8** | 87.0 |

The proposed model outperformed the previous methods on datasets with high imbalance ratios 50 and 100, suggesting that the model was much more able to handle extremely biased datasets. However, the model did not manage to outperform the previous methods on datasets with smaller imbalance ratio, suggesting that the causal operations might have been too strong and hurt the performance.