NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

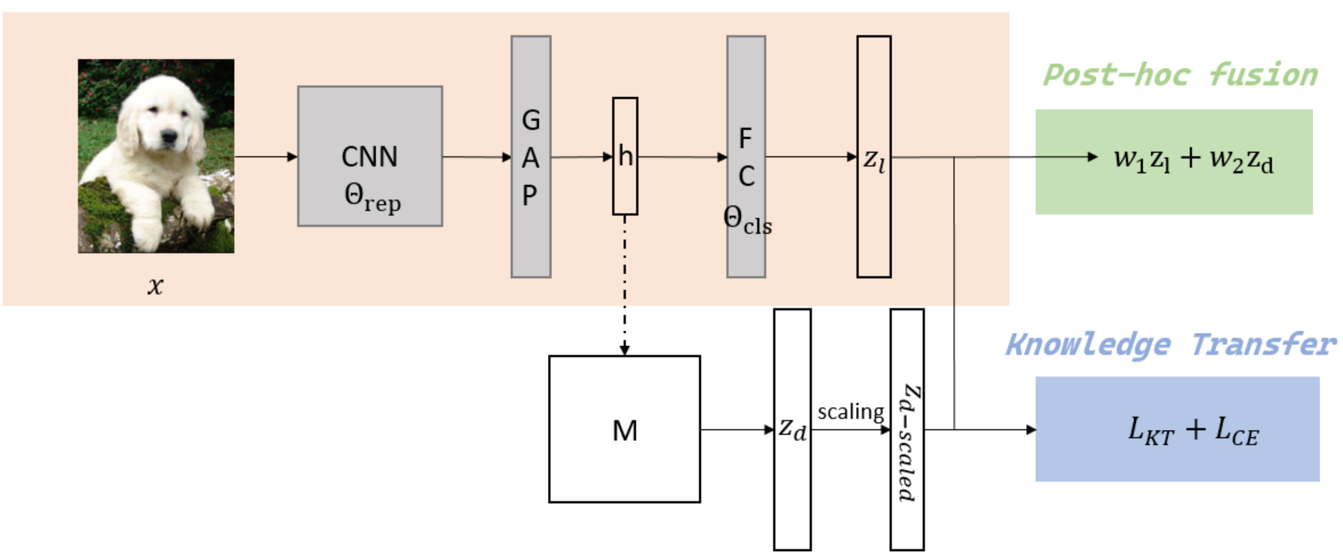
Long-Tailed Image Recognition

Student: Li Zhaochen Supervisor: Assoc Prof Loy Chen Change

Project Description:

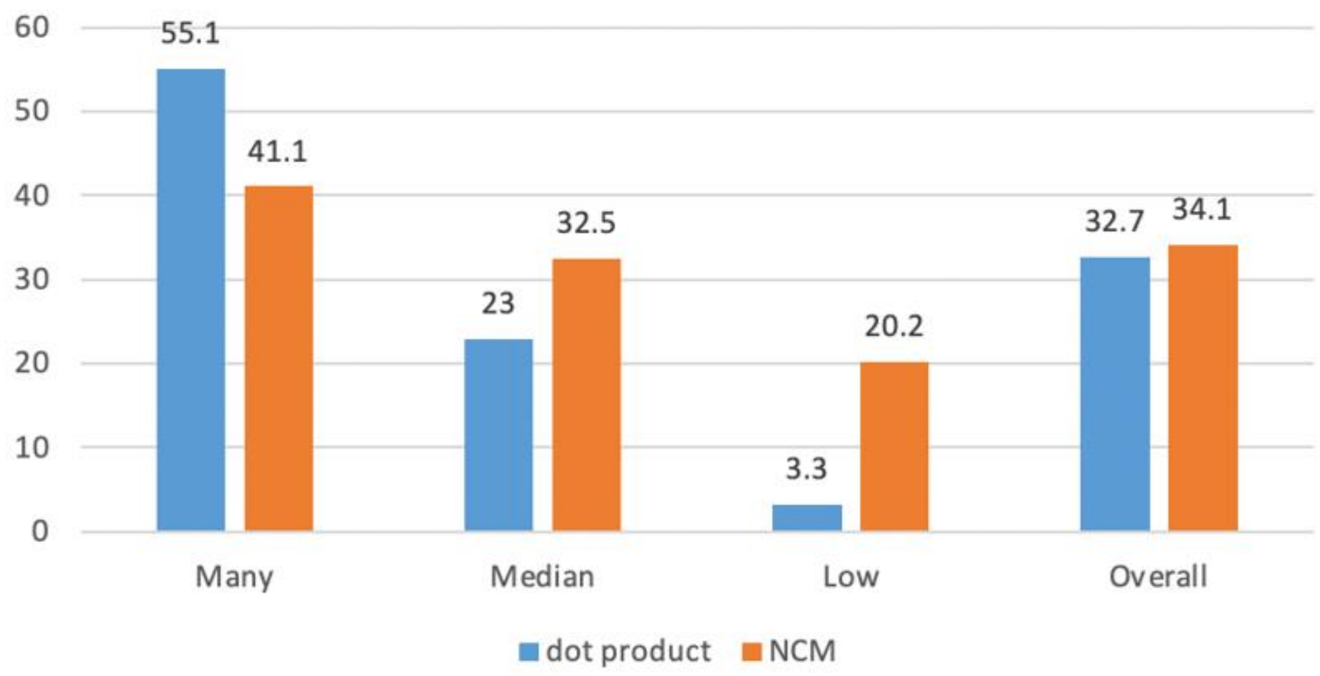
The long-tailed distribution problem poses great challenges to deep learning based computer vision tasks, making the models perform poorly on the balanced test set, particularly for the less frequent classes. In this project, we set out to explore the long-tailed problem via two approaches: 1) How Mixup, a commonly used data augmentation technique, could affect the model's performance. 2) We design a post-hoc fusion and knowledge transfer mechanisms with a memory module to ameliorate the model's performance on tail classes. This poster presents the second approach.

Standard Training Architecture



Memory Module:

Let θ be the optimized model parameters. We decouple θ to θ_{rep} and θ_{cls}, where θ_{rep} are the parameters of all layers before the final fully connected layer, represented by h. The memory module M is calculated by averaging feature h for all inputs x from class i. Using Nearest Class Mean classifier (NCM), there is significant gains over Median and Low classes.



*Code is available at https://github.com/Sonseca97/fyp-long-tail-recognition

Post-hoc fusion:

We merge the two outputs from the respective classifier after backbone training is complete with Instance-Balanced Sampling. We propose four straightforward fusion baselines and two finetuning mechanisms: **Learnable Logits Weight (LLW)** and **Assignment Module (ASM).**

fusion baselines

1. $argmax\{z_l + z_{d-scaled}\}$
2. $argmax\{\sigma(z_l) + \sigma(z_{d-scaled})\}$
3. $argmax\{maxConcat\{z_l, z_{d-scaled}\}\}$
4. $argmax\{w \times z_l + (1-w) \times z_{d-scaled}\}$

LLW

$argmax\{w_1 \times z_l + w_2 \times z_{d-scaled}\}$

ASM

Multi-Layer Perceptron with BCE.

Knowledge Transfer:

M is initialized after some k epochs and is updated per minibatch using a moving average estimate for effective computation. We treat NCM classifier output as "teacher" and add a Kullback-Leibler Divergence loss on a second dot product classifier for knowledge distillation. We create a dynamic mask to make sure only the correct prediction is used for distillation. The final loss for second branch is

$$L = \varphi L_{KT} + (1-\varphi) L_{CE}$$

Results:

Following previous works' experiment settings, here we report accuracy on ImageNet-LT dataset with ResNet-10 backbone. We compare our approaches with baseline model and some of the previous works. The result manifests the effectiveness of our proposed methods.

	Many	Median	Low	Overall
CE Baseline	**56.9**	25.6	3.7	34.6
Previous works				
OLTR	43.2	35.1	18.5	35.6
LFME	47.0	37.9	19.2	38.8
Ours				
ASM	51.6	33.3	16.3	38.0
LLW	50.2	**39.1**	24.0	**41.3**
KT	53.9	31.9	13.5	37.8