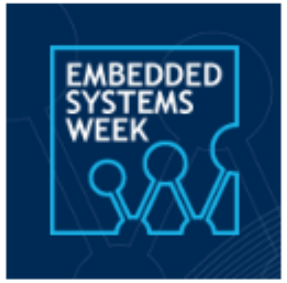


TINY AND FAIR ML DESIGN CONTEST AT ESWEEK 2023

Classification Track: Fair and Intelligent Embedded System Challenge

Rutgers Efficient AI Team

18/09/2023

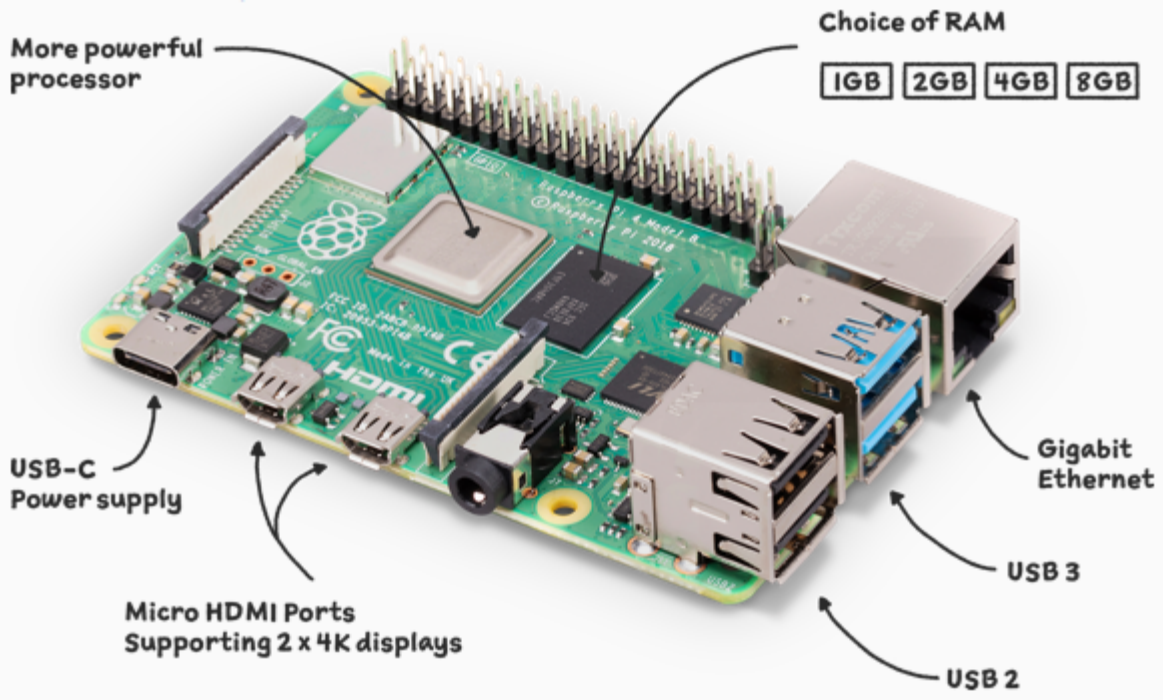
1. Backgroud

2. Our approach
   - 2.1 Self-developed Framework [pycr]
   - 2.2 Training
   - 2.3 Deployment [highport]

3. Ablation Experiments

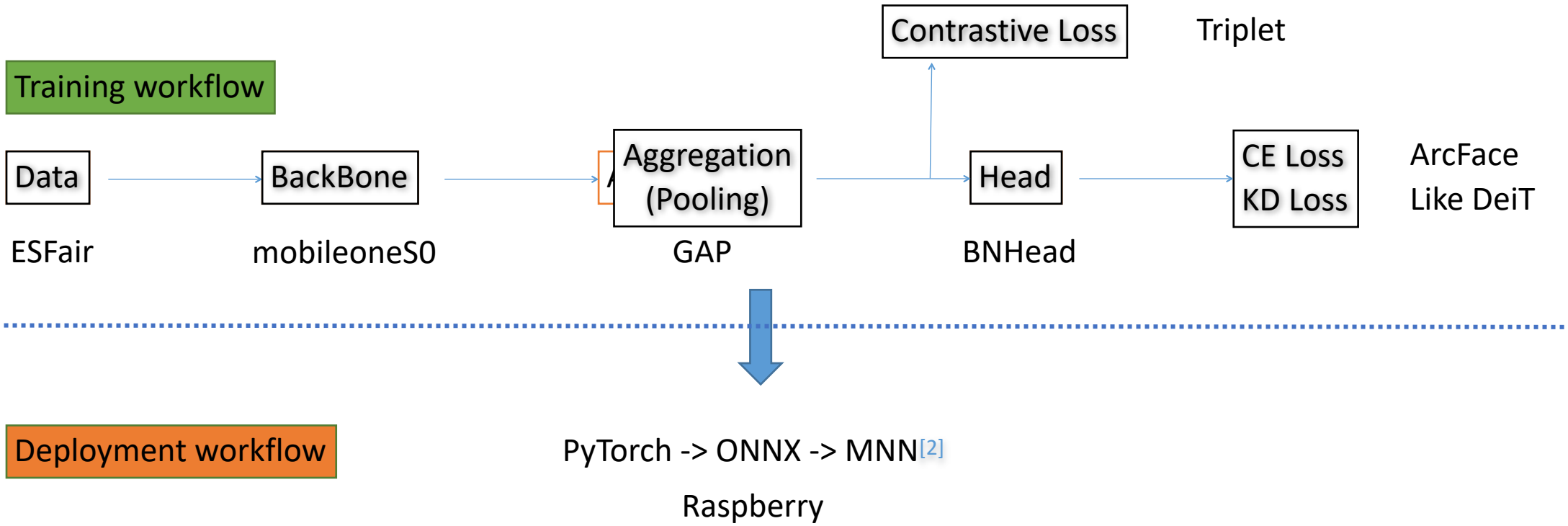
1. Background

To achieve fairness, accuracy, and latency for the Raspberry4B from the dermatology dataset.



2. Our approach

2.1 Self-developed Framework - pycr(PyTorch for Classification and Retrieval)[1]



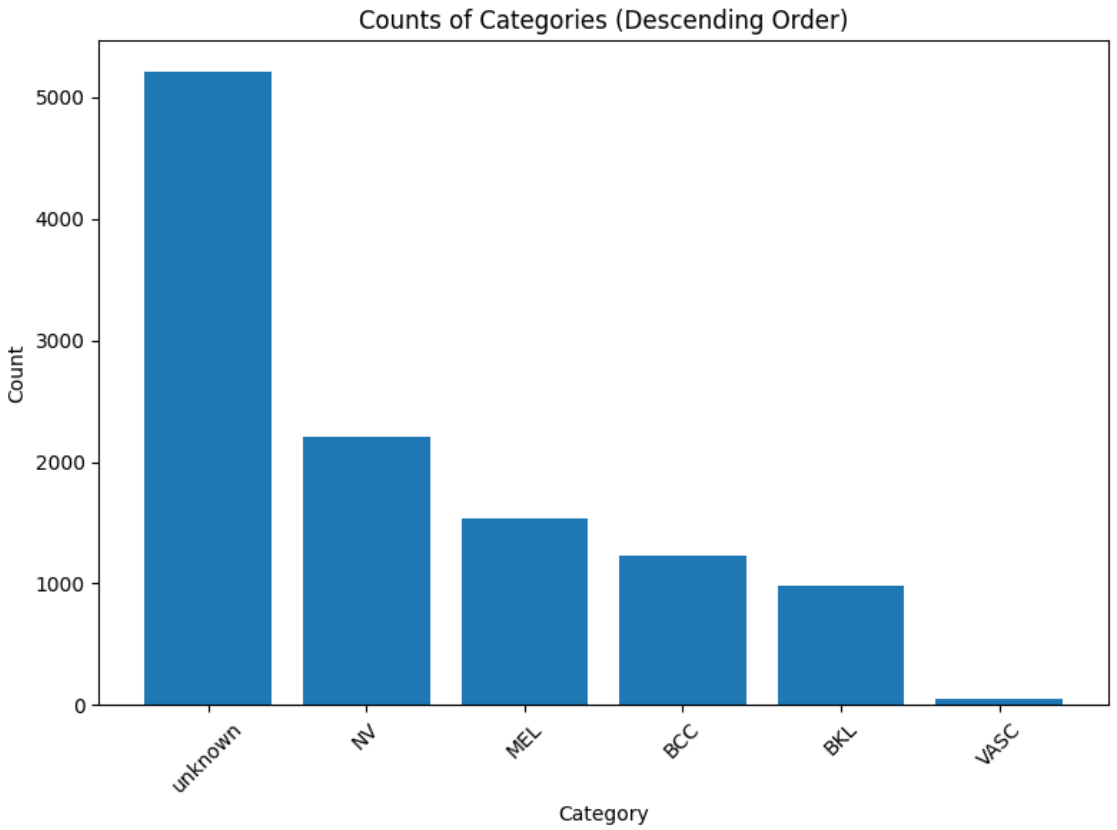
[1] https://github.com/xiaochengsky/pycr.git
[2] https://github.com/dreamflyforever/highport

2. Our approach

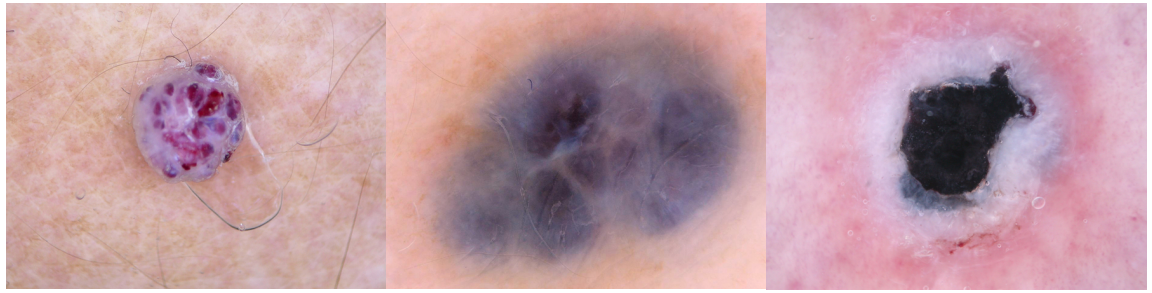
2.2 Training

a) data analysis & processing

long tail distribution



VASC



The features are very obvious(acc 98+%)

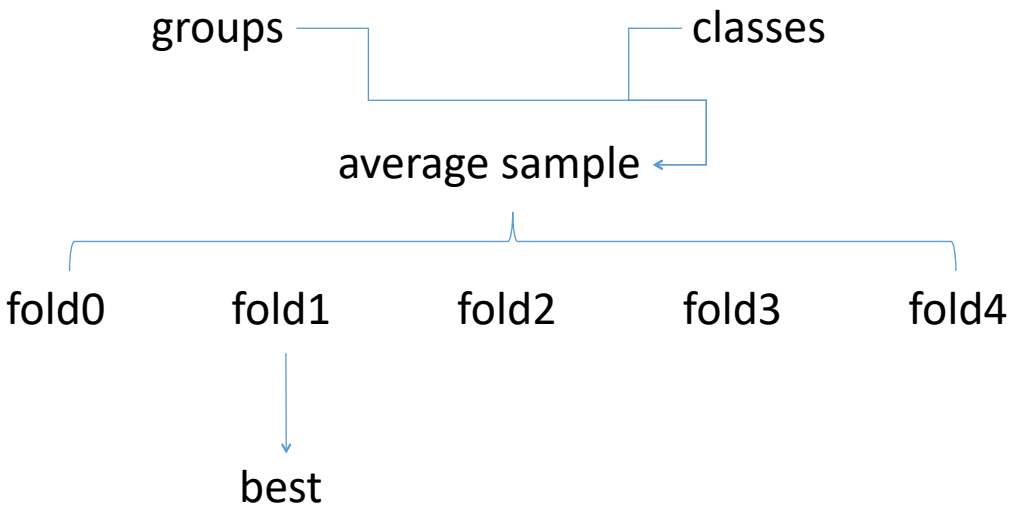
So, we don’t need to worry about the long tail.

2. Our approach

2.2 Training

a) data analysis & processing

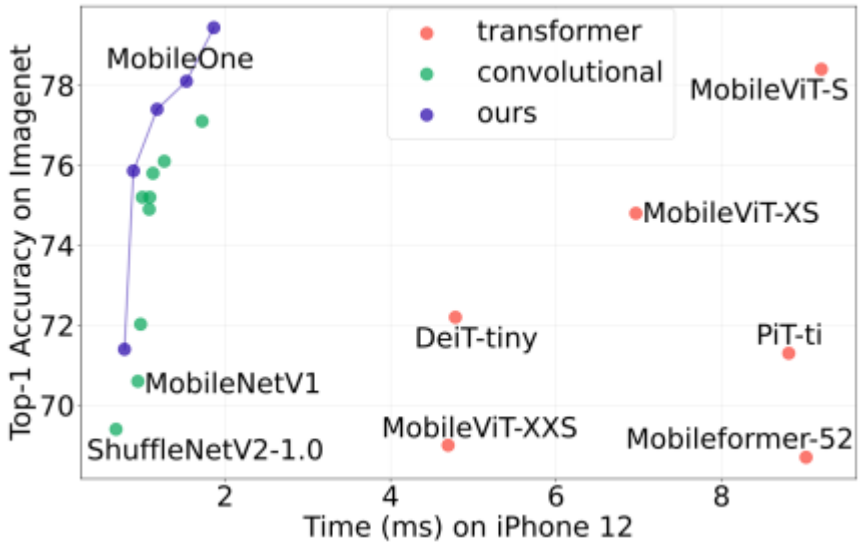
5-fold cross-validation: to find the best fold

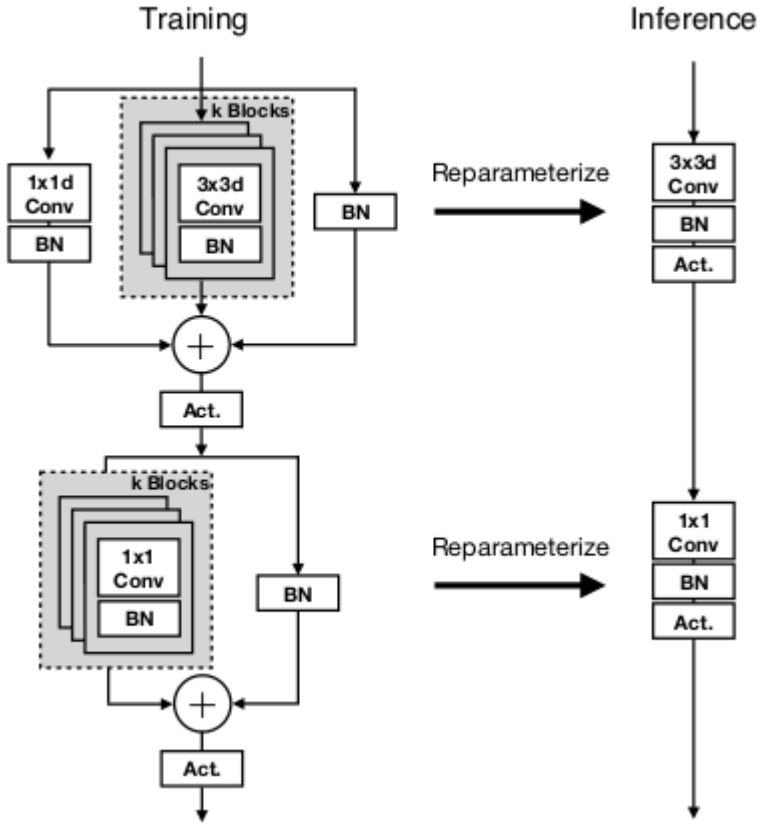


2. Our approach

2.2 Training

b) backbone - MobileOne





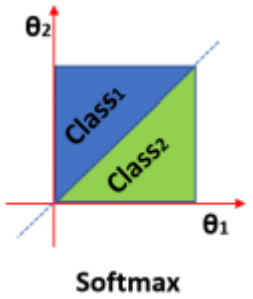
Model Parameters: 21M → 4.3M

2. Our approach

2.2 Training

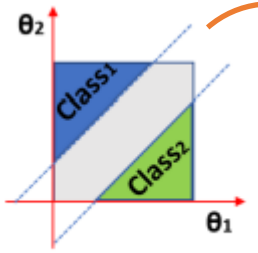
c) Classification Loss - ArcFace loss

Softmax loss $$-\log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{N} e^{W_j^T x_i + b_j}}$$



Softmax

ArcFace loss $$-\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{N} e^{s\cos\theta_j}}$$



ArcFace

margin with interclass
in cosine distance space

2. Our approach

2.2 Training

c) Contrastive Loss - Triplet loss

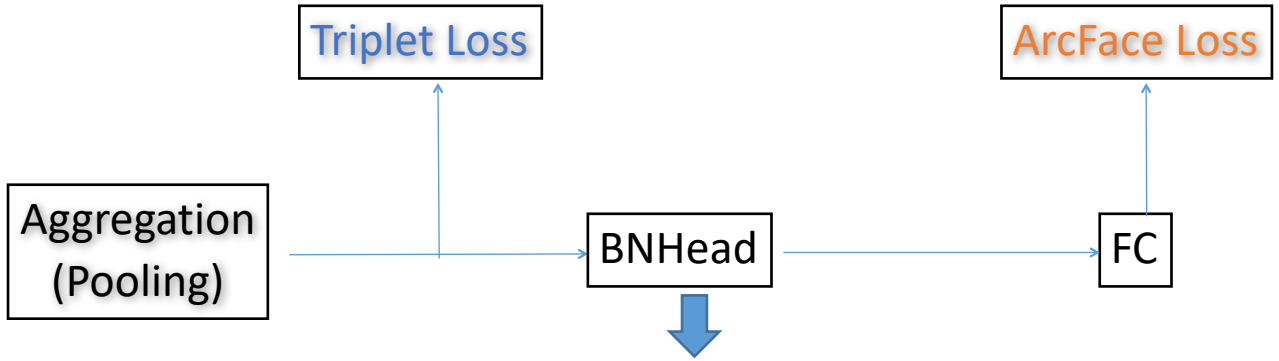
enlarge margins with interclass, and
enclose margins with intraclass
in Euclidean distance space,

$$\|f(x_i^a) - f(x_i^P)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2$$

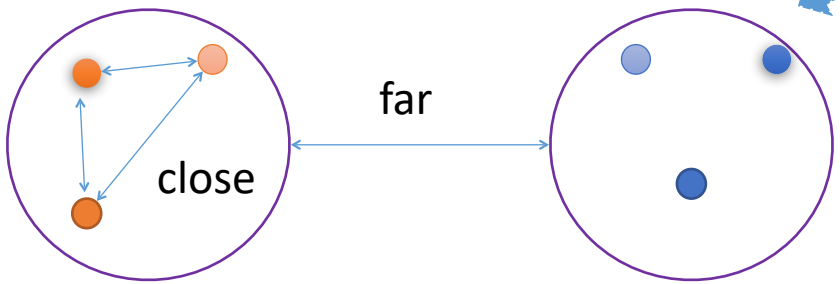
2. Our approach

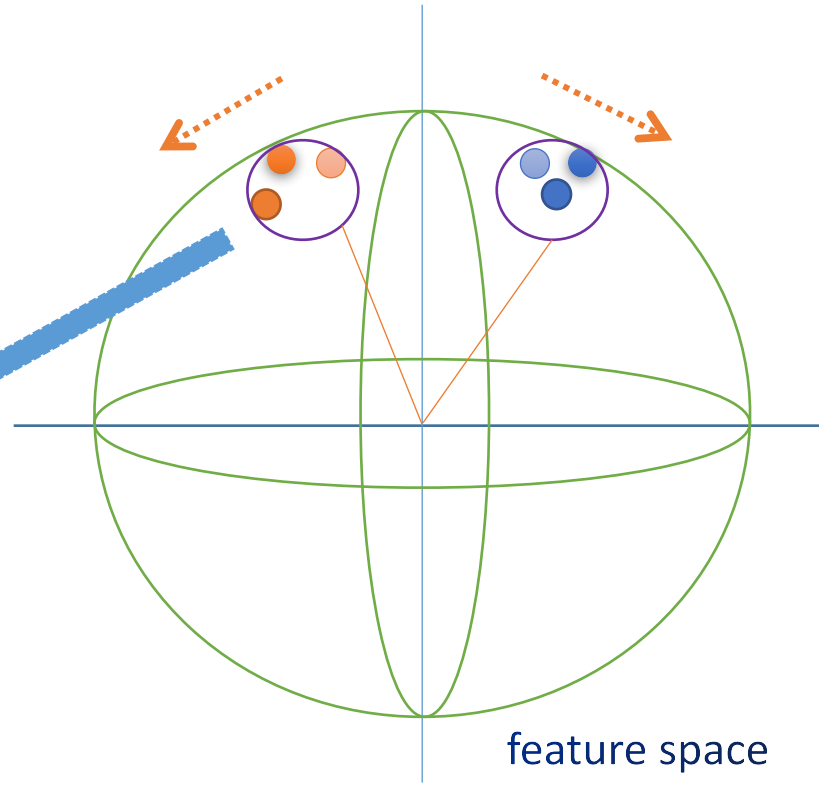
2.2 Training

d) Triplet Loss + ArcFace Loss



for decoupling Euclidean and cosine feature

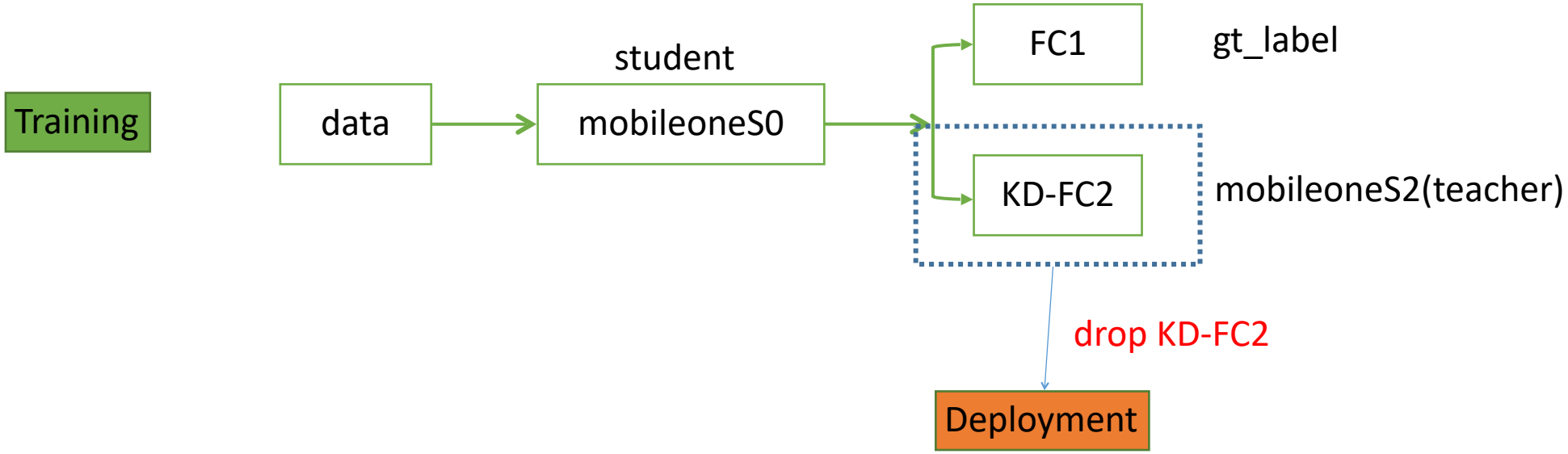




2. Our approach

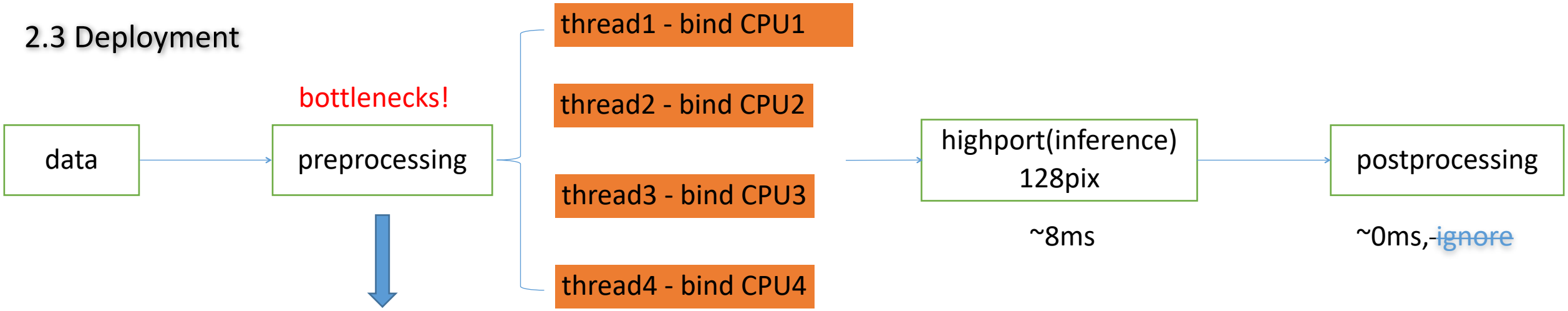
2.2 Training

c) Knowledge Distillation Loss - Kullback-Leibler divergence



2. Our approach

2.3 Deployment



ESFair Dataset

image size	time cost
450 x 600	~ 25ms
1936 x 2592	~300ms
5184 x 3456	~800ms
6000 x 4000	~950ms

image size	num.
small(<1000 x 1000)	3375
large(>3000 x 3000)	3560
medium	4277

Four threads are 3.4 times faster than a single thread!

3. Ablation Experiments(for fairness and accuracy)

Model	pixel	fairness / 3	accuracy / 3	total
MobileNetV3	224	0.277	0.288	0.565
MobileOneS0	224	0.292	0.293	0.585
MobileOneS0	128	0.285	0.287	**0.572**(baseline)
+Augmentation(A)	128	0.290	0.295	0.585(+1.3)
+ArcFace(B)	128	0.295	0.295	0.590(+1.8)
+BNHead(C)	128	0.295	0.298	0.593(+2.1)
+Triplet(D)	128	0.299	0.299	0.598(+2.6)
+EMA(E)	128	0.298	0.303	0.601(+2.9)
+KD(MobileOneS2T)	128	0.302	0.302	**0.604**(+3.2)
MobileOneS2+A+B+C+D+E[MobileOneS2T]	128	0.305	0.303	0.608

Augmentation = RandomHorizontalFlip + RandomVerticalFlip + ShiftScaleRotate +
IAAPerspective + RescalePad + Cutout ~~+ ColorJitter~~

Thanks!