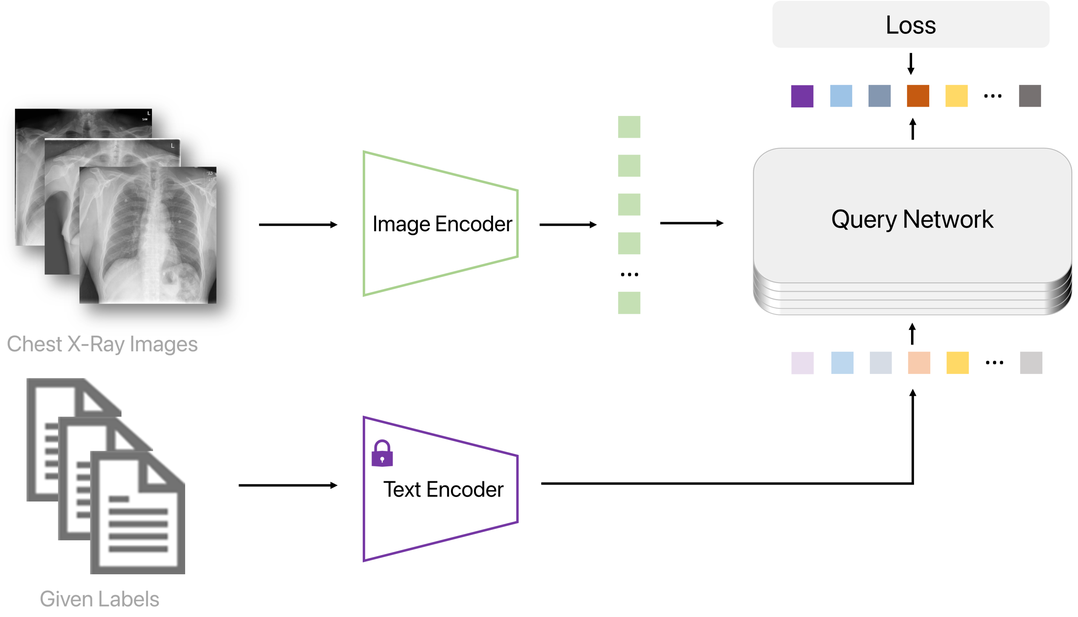
Bag of Tricks for Long-Tailed Multi-Label Classification on Chest X-Rays

Feng Hong*, Tianjie Dai*, Jiangchao Yao, Ya Zhang, Yanfeng Wang

Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, Shanghai AI Laboratory

Framework



- Image encoder, pretrained text encoder and query network.
  - The image encoder can be almost any well-known vision models. We preserve the visual features before pooling that can provide richer information in the subsequent query network.
  - During training, we freeze the text encoder to avoid overfitting.
  - After extracting the features of the two modalities, we use multiple transformer decoder layers for disease diagnosis, with label embedding as Query, and CXR image feature as Key and Value. The transformer decoder aims to combine textual information into corresponding image features, further improving predictive performance.

- Problem Formulation

$$S(x) = \Phi_{\text{query}}(\Phi_{\text{image}}(x), \Phi_{\text{text}}(\mathcal{Y})),$$ where x denotes image, $\mathcal{Y}$ denotes label set and $S(x) \in [0,1]^{|\mathcal{Y}|}$ denotes the prediction of x.

LT-specific Designs

- Design 1: Textual Feature Extractor

  Two pretrained models: finetuned PubMedBERT or clinical-T5

- Design 2: Loss Function Reweighting

  Upweight loss of classes performed poorly on development set

- Design 3: Separate Classifier

  For the last linear layer of the query network, we assign class-wise classifiers for each given class to preserve the independence of predictions between classes.

- Design 4: Data Augmentation

  MixUp is a technique that randomly mixes up different images and labels.

- Design 5: Test-time Augmentation (TTA)

  Multiple random transformations are applied on the test image and we test the model on all transformed images, and take the average as the output.

- Design 6: Class-Wise Ensemble

  For each class, we select best models based on development set and average their predictions.

- Design 7: Exogenous Data Replenishment

Training Dataset	Amount	Label Space
MIMIC-CXR	264849	26
ChestXRay14	98637	14
CheXpert	223414	14
Sum	586900	26

Results

- On development set

Table 1. Results of some different combinations of designs on the development set.

Image Encoder	Text Encoder	Separate Classifier	Reweighting	MixUp	TTA	mAP
ResNet-50	PubMedBERT					0.3187
ResNet-50	PubMedBERT	✓				0.3193
ResNet-50	PubMedBERT	✓			✓	0.3220
ResNet-50	PubMedBERT	✓	✓, 2		✓	0.3273
DenseNet-121	PubMedBERT	✓	✓, 2		✓	0.3236
ResNet-50	Clinical-T5	✓	✓, 3	✓	✓	0.3273
ResNet-50	PubMedBERT	✓	✓, 3	✓	✓	0.3247
ResNet-50	Clinical-T5	✓	✓, 2	✓	✓	0.3244
ResNet-50	PubMedBERT	✓	✓, 2	✓	✓	**0.3280**

- When separate classifier, reweighting(upweighting ratio=2), MixUp and TTA are simultaneously combined together with ResNet-50 and PubMedBERT, the best mAP score is achieved.
- The incorporation of a series of LT-specific designs help boost the chest X-rays disease diagnosis ability of the base general architecture from 0.3187 to 0.3280.

- On testing set

Table 2. Results of model ensemble and data replenishment on the test set.

Ensemble Method	Data Replenishment	mAP
Model-wise Ensemble		0.347
Class-wise Ensemble		0.348
Class-wise Ensemble	✓	**0.349**

- Class-wise ensemble outperforms the model-wise ensemble, which directly averages the predictions of distinct models.
- External data replenishment further improves the performance, achieving the highest score among all attempts at a remarkable 0.349 mAP and ranking in the top five of ICCV CVAMD 2023 CXR-LT Competition.

SHANGHAI JIAO TONG UNIVERSITY

上海人工智能实验室 Shanghai Artificial Intelligence Laboratory