A Semi-Supervised Acoustic Scene Classification Network Based on Multi-Modal Information Fusion

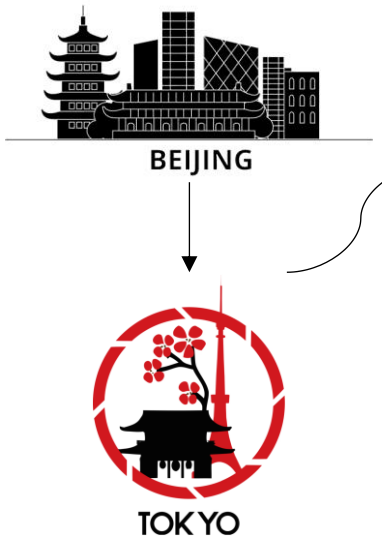
Junkang Yang, Hongqing Liu, Liming Shi, Lu Gan, Hiromitsu Nishizaki and Chee Siang Leow

University of Yamanashi, Japan
Chongqing University of Posts and Telecommunications, China
Brunel University London, United Kingdom

Motivation

Current Limitations

Challenge for Domain Shift



- Decline in performance with acoustic data from different cities.
- Reason: over-reliance on acoustic features.

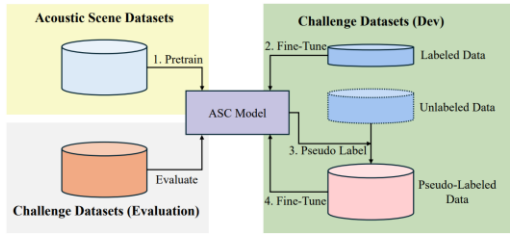
Ineffective Utilization of Contextual Information

- Focus on audio signal only.
- Ignored the associated metadata like geographic location and timestamps.

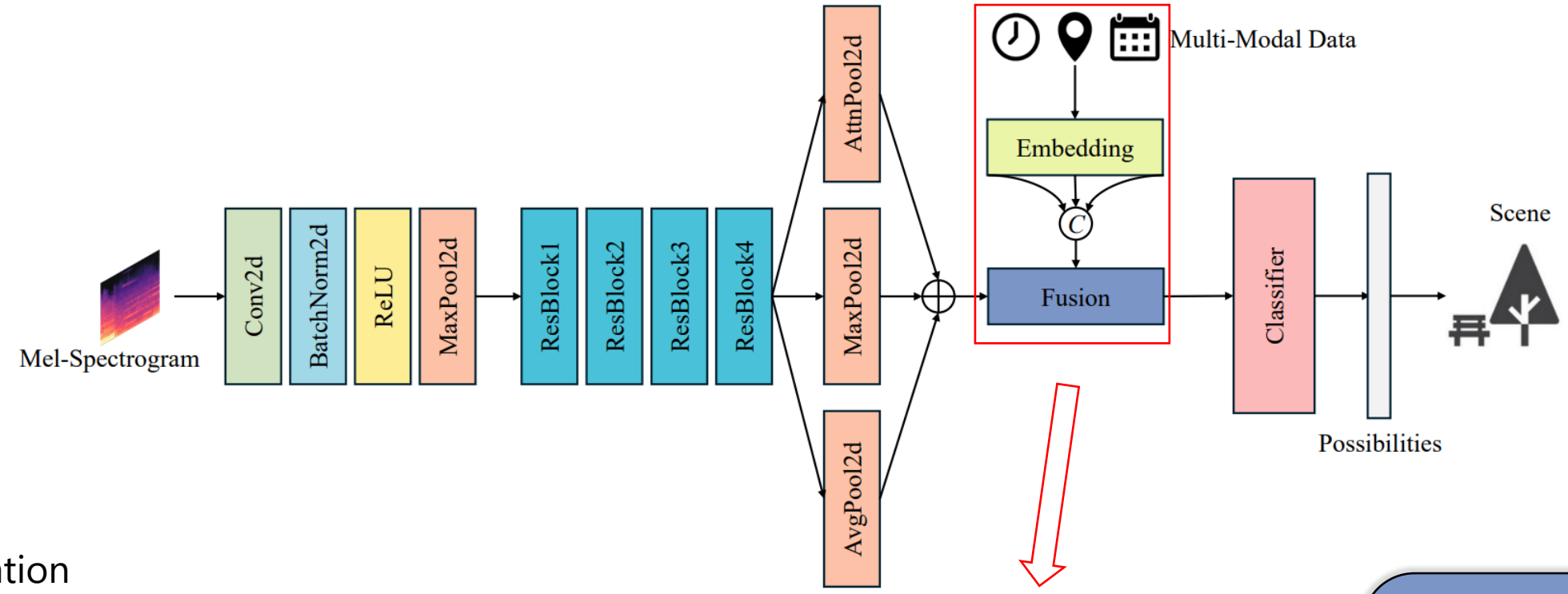
Targets

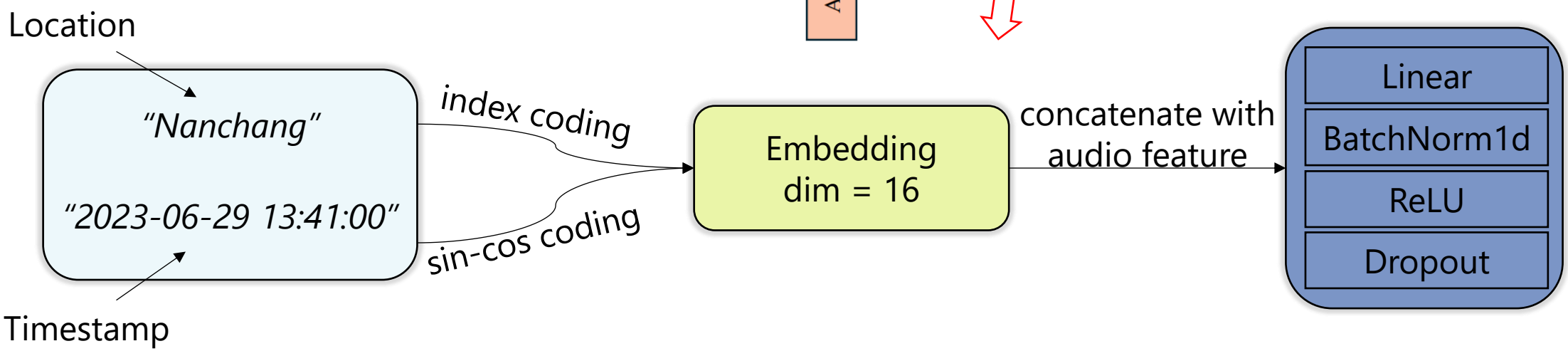
Semi-Supervised Multimodal Classification Network

- Improving cross-domain robustness by multi-stage training.
- Making the model city- and time-aware by embedding contextual into it.



Proposed Network

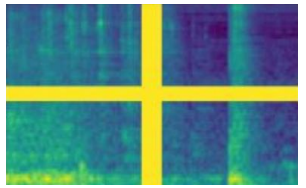
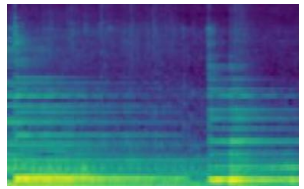




Data & Training Details

■ Data

Pre-train: TAU Urban Acoustic Scenes 2020 Mobile and CochlScene (re-labeled and normalized).

Augmentation Methods: SpecAugment  and Mixup  .

Finetune: data provided by challenge.

■ Semi-Supervised Learing

(1) Pre-training on labeled data;

(2) Supervised fine-tuning on labeled challenge data;

(3) Pseudo-labeling, predicting labels for unlabeled challenge data;

(4) Pseudo-label finetuning, with labeled and pseudo-labeled challenge data.

Results

TABLE II
TRAINING ACCURACY ON VALIDATION DATA OF DIFFERENT STAGES.

Stage	Accuracy (Average)
Pre-Training	93.70%
First Round Fine-Tuning	87.00%
Second Round Fine-Tuning	87.60%

TABLE IV
COMPAXITY ANALYSIS OF PROPOSED MODEL.

Item	Value
#Params	21.65M
MACs	2.34G
CPU Inference Time	40ms

TABLE III
FINAL RESULTS ON EVALUATION DATA.

Item	Accuracy
Bus	0.440
Airport	0.693
Metro	0.920
Restaurant	0.750
Shoppingmall	0.580
Public square	0.040
Urban park	0.700
Traffic street	0.650
Construction site	0.510
Bar	0.850
Macro-accuracy	0.613

Based on macro-accuracy on evaluation data, our method got the **3rd** place in the challenge this time.

Thank you for listening.

UNIVERSITY OF YAMANASHI

Regional Core & Global Professionals