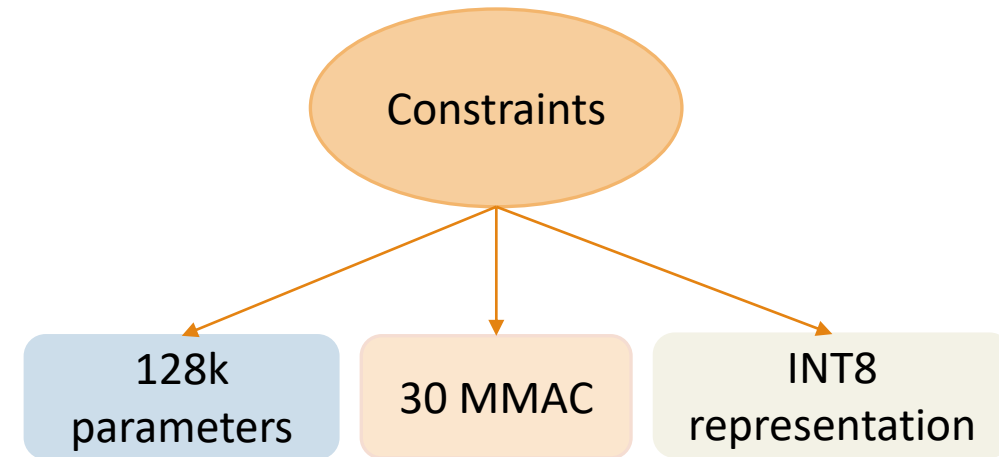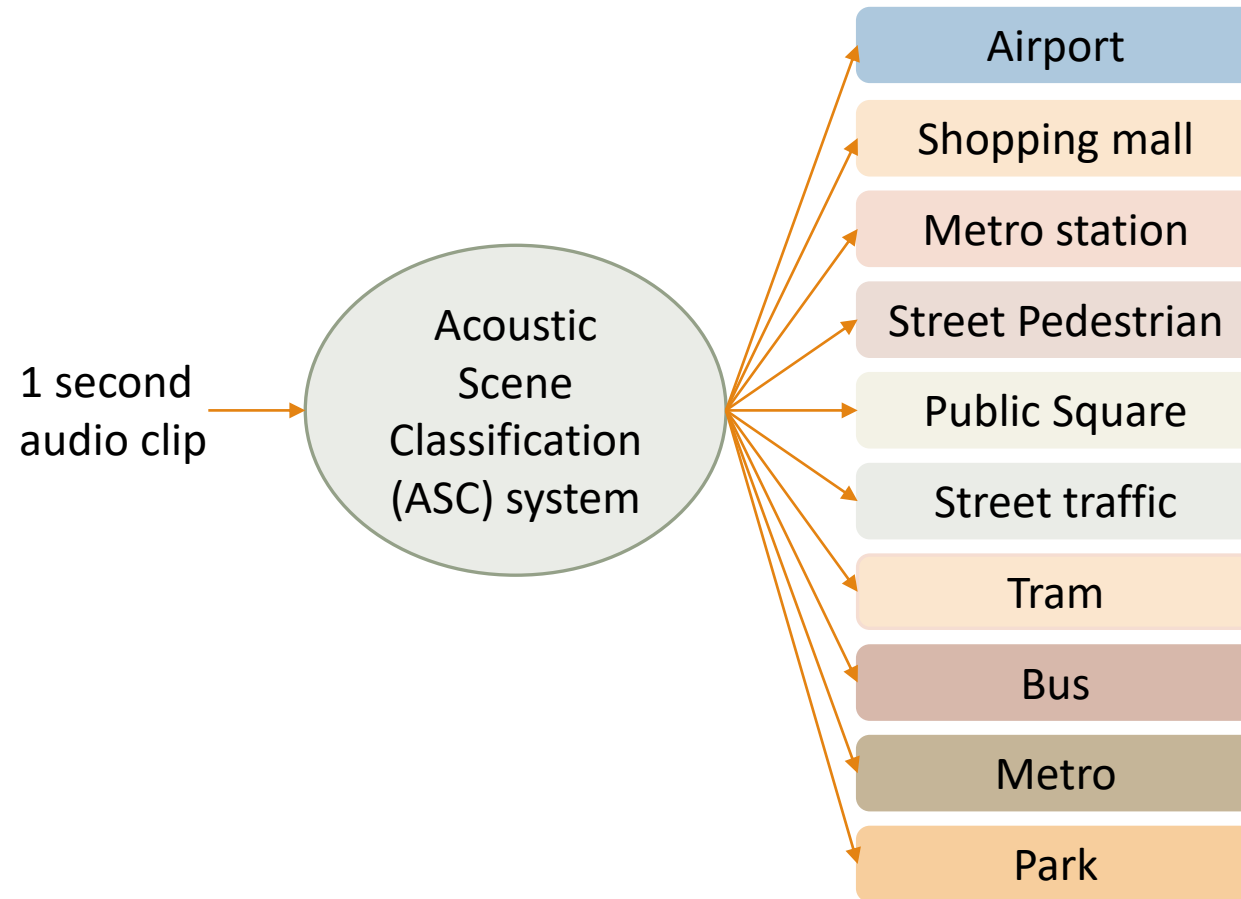# A DEVICE CLASSIFICATION-AIDED MULTI-TASK FRAMEWORK FOR LOW-COMPLEXITY ACOUSTIC SCENE CLASSIFICATION

ROHITH MARS & ROHAN KUMAR DAS

FORTEMEDIA, SINGAPORE
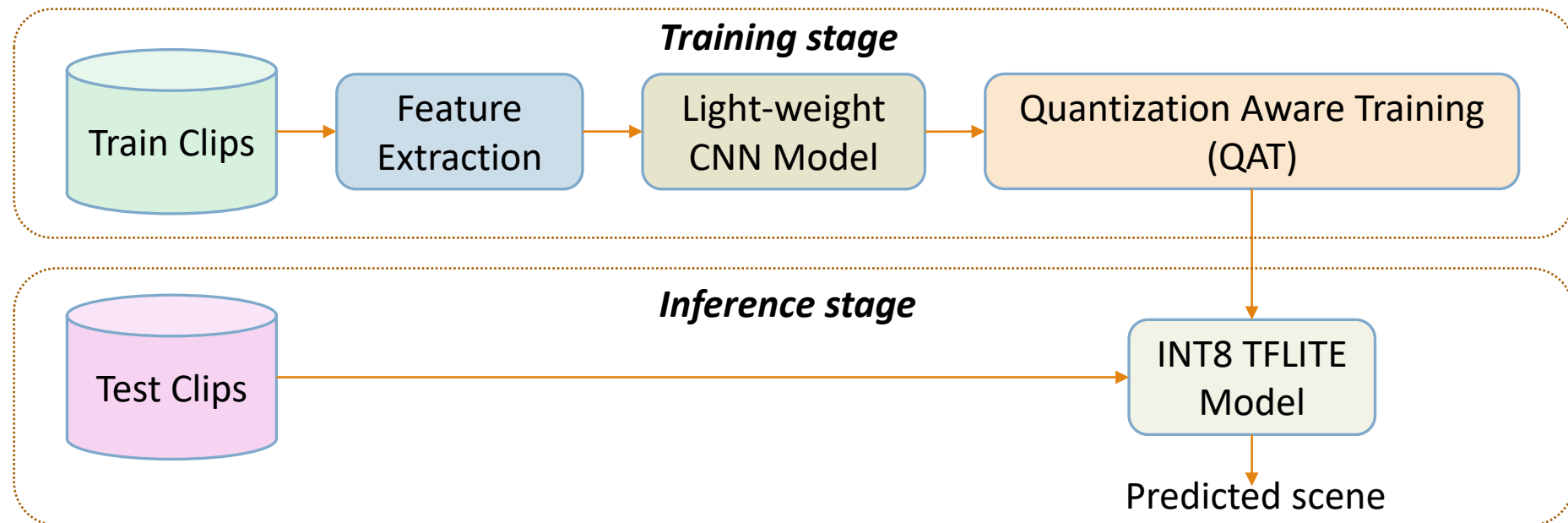
Objective & Constraints

Proposed Solution

- **Key Components**
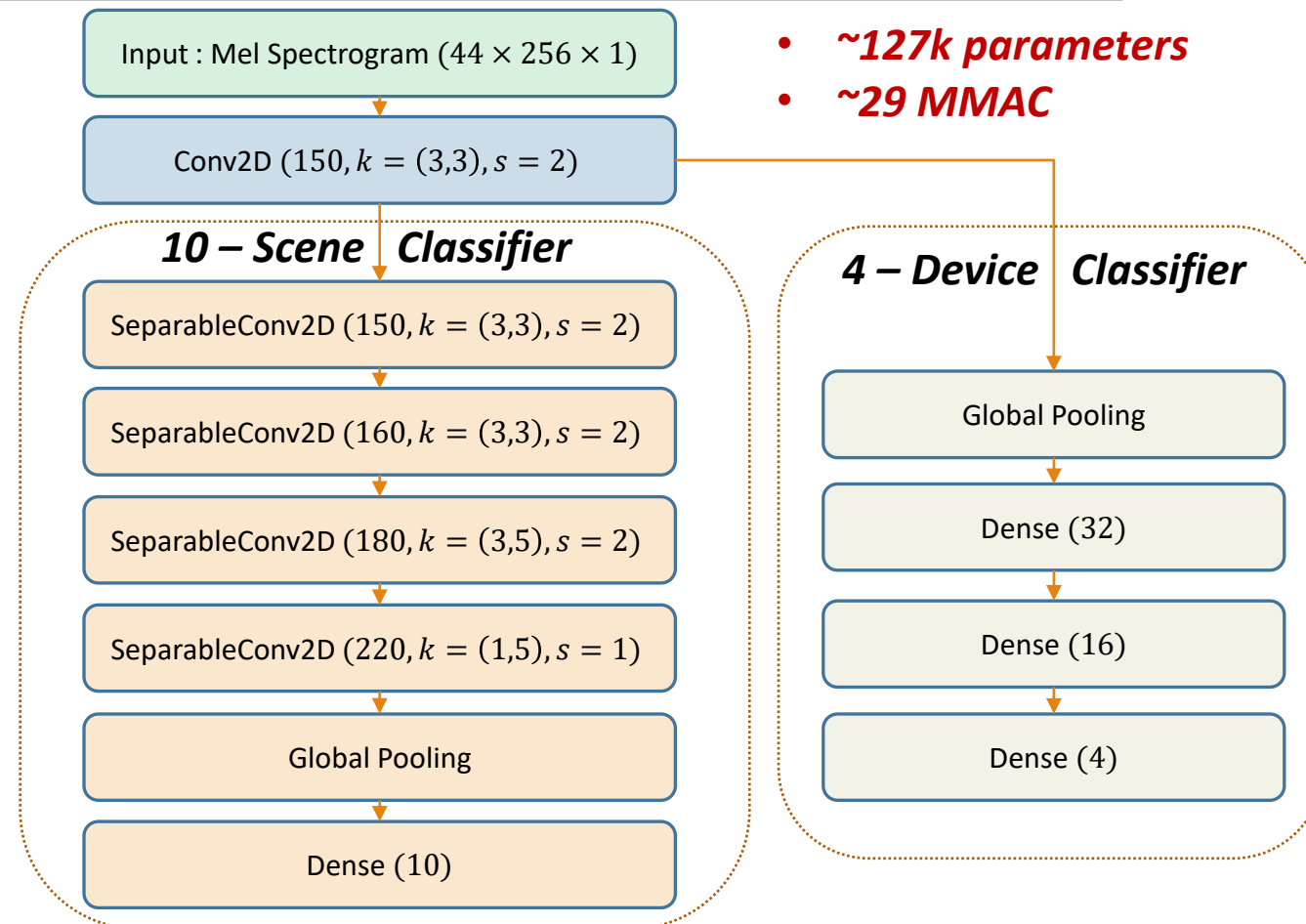 - Multi-task learning framework (MTL) with device classification as auxiliary task
  - Separable and strided convolutions
  - Data augmentation to improve generalization
  - Quantization-aware training (QAT)

Training stage

Train Clips → Feature Extraction → Light-weight CNN Model → Quantization Aware Training (QAT)

Inference stage

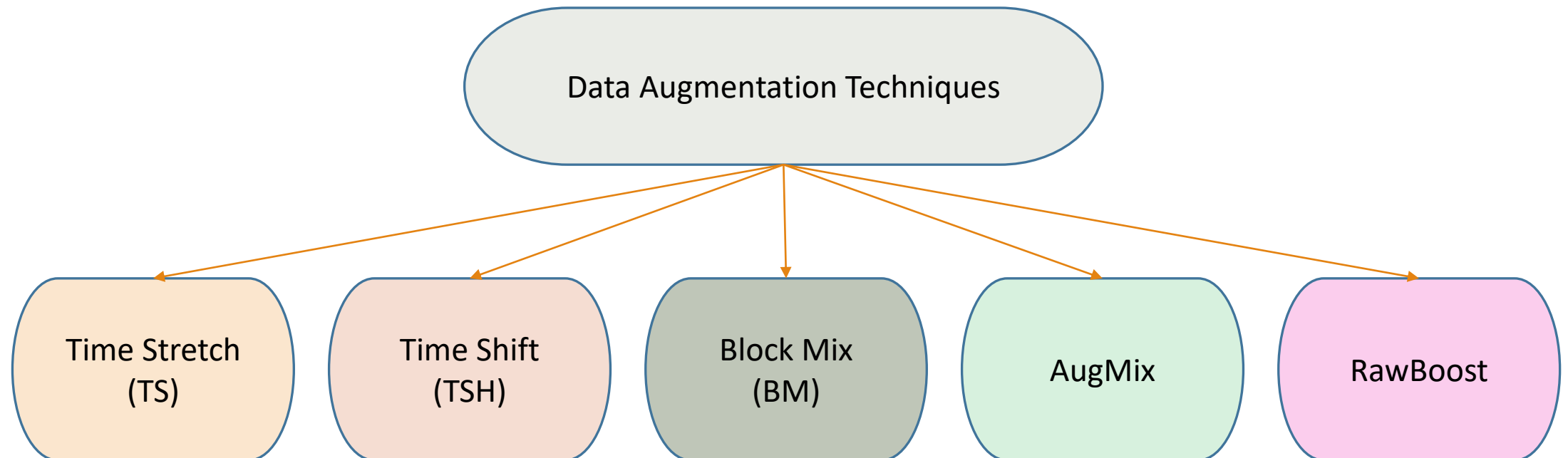Test Clips → INT8 TFLITE Model → Predicted scene

Device-Classification Aided MTL

- Introduce device classification as an auxiliary task

- From development set
  - 3 classes – real devices (A, B, C)
  - 1 class – grouped (simulated/unseen)

- Use a combined loss to train
  - $L_{MTL} = \beta \times L_{device} + (1 - \beta) \times L_{ASC}$

- Remove device classification branch for inference stage

Input : Mel Spectrogram ($44 \times 256 \times 1$)

Conv2D ($150, k = (3,3), s = 2$)

- **~127k parameters**
- **~29 MMAC**

$10 - Scene\ Classifier$

SeparableConv2D ($150, k = (3,3), s = 2$)

SeparableConv2D ($160, k = (3,3), s = 2$)

SeparableConv2D ($180, k = (3,5), s = 2$)

SeparableConv2D ($220, k = (1,5), s = 1$)

Global Pooling

Dense (10)

$4 - Device\ Classifier$

Global Pooling

Dense (32)

Dense (16)

Dense (4)

Data Augmentations

Results (1/2): Ablation Studies

Framework	Avg. Log Loss	Avg. Acc. (%)
STL	1.333 ± 0.012	51.25 ± 0.33
MTL, $\beta = 0.1$	1.327 ± 0.009	51.65 ± 0.31
MTL, $\beta = 0.2$	$\mathbf{1.319 \pm 0.017}$	$\mathbf{51.98 \pm 0.55}$
MTL, $\beta = 0.3$	1.323 ± 0.014	51.31 ± 0.55
MTL, $\beta = 0.4$	1.347 ± 0.014	50.53 ± 0.54
MTL, $\beta = 0.5$	1.335 ± 0.010	51.31 ± 0.40

Method	Avg. Log Loss	Avg. Acc. (%)
No Data Augmentation	1.319 ± 0.017	51.98 ± 0.55
TS + TSH + BM	1.314 ± 0.007	52.08 ± 0.33
AugMix	1.309 ± 0.014	52.45 ± 0.56
RawBoost	1.298 ± 0.010	52.22 ± 0.71
Combined (1) + (2) + (3)	$\mathbf{1.265 \pm 0.009}$	$\mathbf{53.59 \pm 0.43}$

- Best log loss is achieved for MTL framework with $\beta = 0.2$.

- **RawBoost** technique help to improve generalization to unseen devices.

Results (2/2) : Comparison to Other Systems

System	Avg. Log Loss	Avg. Acc. (%)
Lee et al.	0.835	70.1
Anastacio et al.	1.103	60.5
Schmid et al.	1.139	58.0
Sugahara et al.	1.182	56.5
Kim et al.	1.259	54.0
Proposed MTL (Ours)	1.273	53.5
Morocutti et al.	1.288	52.7
Xin et al.	1.295	60.3
Yu et al.	1.305	51.7
Shao et al.	1.360	54.1
Baseline	1.575	42.9

THANK YOU!