

UE MLPC 2025: Challenge



Team Fumbling:

Abdalaziz Ayoub

Abdulkarim Al Jamal

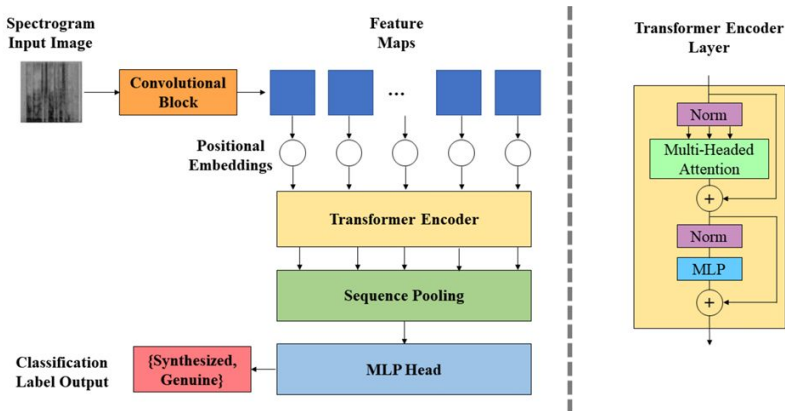
Beibarys Abissatov

Jeronim Bašić

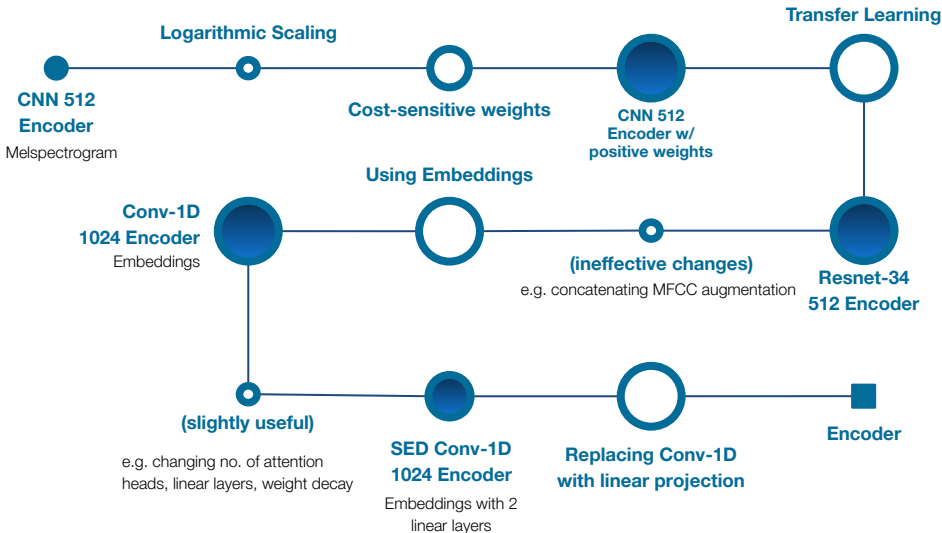


Architecture: Inspiration

- **Most Informative Feature: Mel-spectrograms**
- 2D representations of audio ➡ Convolutional Transformer?
 - Convolutional layers: extract temporal feature maps
 - Transformer: models long-range dependencies

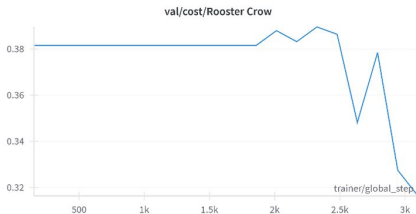
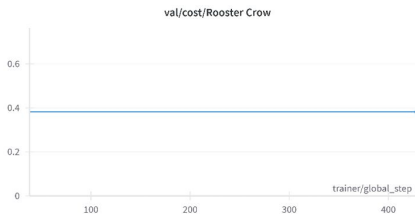


Chronology of the Challenge



Impactful Changes

Cost-Sensitive Tuning of Positive Weights



Adjusted class-specific costs for rare class labels or non-improving ones.

Transfer Learning



Replaced our custom Conv2D stack with a ResNet-34 to extract richer representations.

Impactful Changes

Embeddings



Replaced log-Mel spectrograms with provided audio embeddings (and thus also changed to Conv1D).

Linear projection layer



Used simple linear layer to project frozen precomputed embeddings in-place of Conv1D.

Some Ineffective Changes

Spectrogram Augmentation

Applied time/frequency masking to spectrograms

Outcome: Performance slightly worsened (61 \rightarrow 62.68)

Insight: Augmentation may disrupt critical acoustic features in this context

Concatenating MFCC Coefficients:

Useful MFCCs (e.g., coefficients 0, 1, 22 from previous report) summarize spectral content and can enrich the mel-spectrogram representation.

Outcome: No noticeable improvement in performance.

Insight: concatenating them along the frequency axis without channel separation may have led to suboptimal integration

Adding Additional Linear Layers:

a deeper classification head with two linear layers will improve the transformation of learned embeddings into the class probabilities.

Outcome: Minor performance gain.

Insight: The base model likely already had sufficient capacity.

Thank you!
Any Questions?



JOHANNES KEPLER
UNIVERSITY LINZ

