YOHO model for Audio Segmentation and Sound Event Detection

Davide Capone [SM3500601] Enrico Stefanel [SM3500554]
{davide.capone, enrico.stefanel}@studenti.units.it

Data Science and Scientific Computing Master's Course
Department of Mathematics and Geosciences
University of Trieste

A.Y. 2023–2024

Contents

Audio Segmentation and Sound Event Detection

...

Datasets

...

Metrics

...

YOHO model

Presented in 2021[**Venkatesh_2022**]...

Output shape

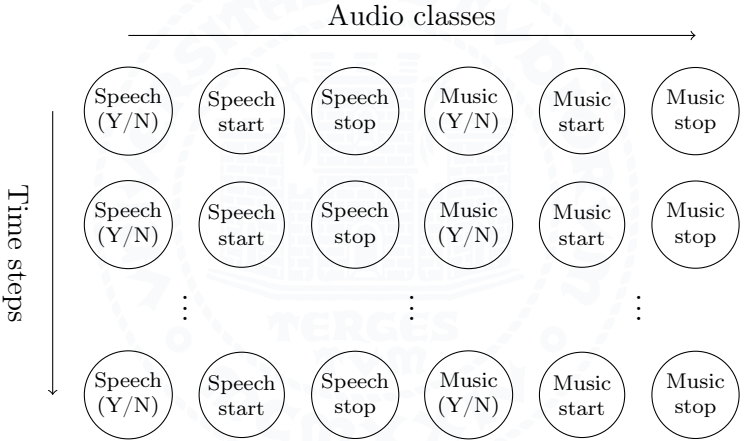

Figure: The YOHO output shape.

Network Architecture

...

Loss Function

$$\mathcal{L}_c(\hat{y}, y) = \begin{cases} (\hat{y}_1 - y_1)^2 + \\ (\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2 & \text{if } y_1 = 1 \\ (\hat{y}_1 - y_1)^2, & \text{if } y_1 = 0 \end{cases}$$

where y and $\hat{y}$ are the ground-truth and predictions respectively. $y_1 = 1$ if the acoustic class is present and $y_1 = 0$ if the class is absent. y_2 and y_3, which are the start and endpoints for each acoustic class are considered only if $y = 1$. In other words, $(\hat{y}_1 - y_1)^2$ corresponds to **the classification loss** and $(\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2$ corresponds to **the regression loss**.

Other Details

...

Proposed improvements

...

New backbone

...

Conclusions

Questions?

References