Proposal for Label Buddy 2.0: Automated audio-tagging using transfer learning

Ioannis Prokopiou

ΕΛ/ΛΑΚ

Ελεύθερο Λογισμικο /
Λογισμικο Ανοιχτού Κώδικα

Google Summer of Code 2021

Organization: GFOSS - Open Technology Alliance

Google Summer of Code

April 2021

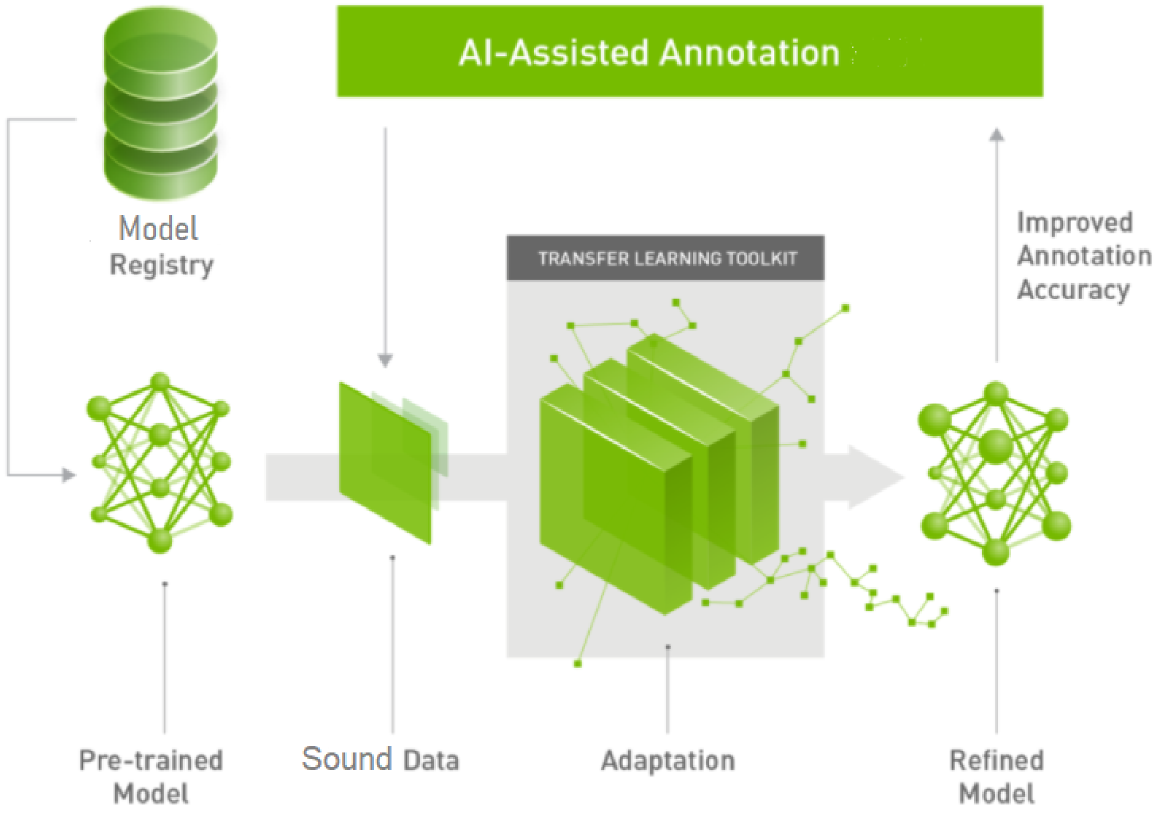
Introduction

The growing volume and variety of sound data slows down identification and analysis of specific features. This reduces the annotation speed at which sound analysers operate.

The demand for artificial intelligence in medical image analysis has drastically grown in the last few years. AI-assisted solutions have emerged to speed up annotation workflows and increase productivity.

A sound annotator takes a sound record, marks annotation boundaries and corrections as needed. This manual process repeats for the next record in which the sound of interest is annotated and corrected again. Because these workflows are completely manual, the annotation process takes time.

Deep learning models can be used for annotation and can kickstart your development effort by enabling faster annotation of datasets for AI algorithms. Deep learning models are sensitive to the data used to train them, this makes it hard to train the deep learning models on a specific dataset and deploy them on a different dataset. As a solution, transfer learning for sound could help adapt pretrained models into various datasets. Deep learning models used for annotation can be tuned and improved by retraining these pretrained models based on new datasets.



In the most general terms, transfer learning (TL) is an approach in machine learning that focuses on storing "knowledge" that a model has learned in order to solve a given problem A and use that knowledge to help with another related problem B. Transfer learning is a popular machine learning technique, in which you train a new model by reusing information learned by a previous model. Most common applications of transfer learning are for the vision domain, to train accurate image classifiers, or object detectors, using a small amount of data -- or for text, where pre-trained text embeddings or language models like BERT are used to improve on natural language understanding tasks like sentiment analysis or question answering. The sound domain is still in its infancy. (YAMNet) Previous work has been done on the AI-assisted annotation of sound (MusicNN).

Project goals

This project is an enhancement to the previous work that has been done previously. Its goal is to make annotation simple and easy while also providing a well-defined manager-annotator-reviewer framework. The goal of this project is to use Transfer Learning (TL) approaches to make the annotation process easier for the user by offering label predictions. This way it will be possible to accomplish more with less data and effort.

During GSoC period development:

- An analysis of pros and cons of existing work such as BAT, YAMNet and MusicNN. This task will help us understand the drawbacks of the existing implementations and try to avoid them while keeping all the already made advantages.
- Enhance the already made database which will be used to store the corrections in order to refine the model.
- Research and design of the transfer learning model and an explanation for the architecture.
- Implementation of the aforementioned model.
- Enhancements of the already made Graphical User Interface providing Mock-up screens. This task includes the design of every addition on every screen in our tool and it should be discussed with the mentors before the implementation.
- Implementation of the changes on User Interface and system for the use of the AI-Assistance.
- Implementation of the whole AI-Assisting model.

Future development:

- Continue working on tasks after official coding period
- Make the annotation tool a mobile app, using React Native and the already built Database (Django)
- Add features such as image and text annotation using the spectrogram etc.
- Implement an import system where the user will be able to import exported annotations and edit them

Implementation

The project consists of **five** components:

1. ***Database** system - Model Registry*

   The Database system will be enhanced using Django. Some of the already implemented tables needed are given below:

   User: will contain all the needed information about the users such as Username,Password, annotations count, permissions (like number of classes that can be added to each file) and property (manager or annotator)

   Annotation: will contain information about an annotation such as the user who annotated, the segment annotated (segment will be another table), the file to which the segment belongs, the date of the annotation, the name, and its status (unfinished or finished)

   File: will contain information about every file (audio for starters) uploaded such as the name, the user who uploaded it, the date of the upload and if the file is shared. In the event that a file is shared, there will be another table that will indicate the users who are able to annotate this file as well. Finally, there will be a state for the file which will indicate if a user is currently annotating it, in order to prevent users (users who have access to the shared file) from annotating it at the same time and thus avoiding conflicts

Example enhancement table:

Prediction: will contain information about the prediction that the implemented model to be compared with the annotator;s decision in order to be refined if needed.

2. ***Pre-Trained mode***l: the implementation of the model that is going to be making the predictions in order to assist the annotators. The model will be chosen after research and discussion with the mentors.
3. *The sound data* ***preprocessing mechanism****: All the necessary implementations of the mechanism that will process the sound data in order to be in a state to pass through the model successfully.*
4. **Adaptation**: adaptation of the transfer learning toolkit in order for the annotator to be able to turn on the assist whenever he wants. Also, by turning it on, we are going to collect the annotation data for the refinement of the model.
5. ***Refine Model****: After the adaptation, we are going to use the data and the “distance” between the prediction and the actual human annotation to refine the model and consistently make it more and more efficient.*

Schedule/Timeline

May 30 – June 13 (Before the official coding time):

- To do self coding with Django to improve my further understanding on these technologies
- Start conducting research for the appropriate model architecture

June 13 - June 30

- Enhance the already existing database
- Mock-up screens

July 1 - July 15

- Conduct research for the appropriate model architecture
- Analyze the architecture

July 16 - July 31

- Implement the pretrained model
- Test the model to check each case

August 1 - August 15

- Implement the preprocessing mechanism
- Enhance for multiple types of inputs

August 16 - August 31

- Adaptation to the existing infrastructure
- Keeping track of all the differences between prediction and annotation

September 1 - September 15

- Mechanism to refine the model
- Test efficiency after multiple corrections

September 15 - End

- Fix possible issues and complete minor unfinished tasks
- Implement and discuss features mentioned in the second section (Future development)

I think that 1-2 meetings per week with my mentors will be enough.

Related Work:

1. Kim, Juae & Kang, Sangwoo & Park, Yongmoon & Seo, Jungyun. (2019). Transfer Learning from Automatically Annotated Data for Recognizing Named Entities in Recent Generated Texts. 1-5. 10.1109/BIGCOMP.2019.8679473.
2. Wang, Y., & Metze, F. (2017). A Transfer Learning Based Feature Extractor for Polyphonic Sound Event Detection Using Connectionist Temporal Classification. INTERSPEECH.
3. Hyejin Won, Baekseung Kim, Il-Youp Kwak , Changwon Lim. (2021) TRANSFER LEARNING FOLLOWED BY TRANSFORMER FOR AUTOMATED AUDIO CAPTIONING

Repositories:

- BAT
- MusicNN
- CVAT

About me

- LinkedIn
- Github

Contact information:

- Name: Ioannis
- Surname: Prokopiou
- Email: giannprokopiou@gmail.com
- Phone number: +30 6976382553
- Country: Greece

My name is Ioannis Prokopiou and I am a final-semester student of the Computer Engineering and Informatics Department at the University of Patras, Greece. I live in Athens, Greece and currently I am doing my internship at Ofium. Orfium, is a technology company providing software, data, and licensing solutions for the entertainment industry's most complex problems around music, content and rights management. I am goal-oriented with a strong commitment to collaboration and solutions-oriented problem-solving. Committed to high standards user experience, usability and speed for multiple types of end-users. My main field of interest is Machine Learning and I have already collaborated for 2 accepted research papers. My hobbies include movies/series, food and gaming and more passionately basketball, music, traveling with friends and making people laugh. I just love to make an impact at everything I do and make my environment happier and better.