BLAST

Burst Learning Audio Spectrogram Transformer

ELECOMP Capstone Design Project 2024-2025

Sponsoring Company:

SEACORP
62 Johnny Cake Hill, Middletown, RI 02842
http://www.seacorp.com

Company Overview:

SEACORP provides engineering solutions to sustain and expand our nation's maritime advantage. We drive innovation from the ocean floor to the expanse of the skies. We are more than just submarine electronics systems; we are the innovators behind the technology that keeps our Navies ahead. Every day, we deliver mission-ready solutions that bolster our nation's defense capabilities.

SEACORP is at the forefront of cutting-edge AI and ML technologies. We specialize in developing advanced solutions for the defense industry, including QA systems with Large Language Models, acoustic machine learning, data annotation, and AI code optimization. Our expertise enables us to deliver critical capabilities for mission-critical applications.

Technical Directors:

Name Bill Matuszak
Title: Director of Advanced Development
Email: wmatuszak@seacorp.com
https://www.linkedin.com/in/bill-matuszak-71528012/

Name: Megan Chiovaro
Title: AI Data Scientist
Email: mchiovaro@seacorp.com
https://www.linkedin.com/in/megan-chiovaro-phd-179861b6/

Project Motivation:

Motivation: The Audio Spectrogram Transformer (AST) has shown remarkable promise in various audio classification tasks, but its performance can be further enhanced for specific domains. By fine-tuning a pre-trained AST model on a targeted dataset, we aim to: 1) tailor the model to the unique characteristics and patterns of the specific audio data, improving its accuracy and efficiency; 2) reduce the need for extensive training from scratch, saving computational resources and time; and 3) explore the potential of transfer learning to leverage the model's existing knowledge to tackle new audio-related challenges, such as anomaly detection, speaker identification, or sound event localization.

Anticipated Best Outcome:

This is an applied research project to determine the feasibility of using a foundation model with fine tuning for a specific audio application. The anticipated best outcome for this project would be a get performance equivalent to “Audio Set”, on a data set of interest to SeaCorp with a fine-tuned version MIT AST model defined in the paper by Yuan Gong, Yu-An Chung, James Glass.

Additionally, a successful fine-tuning process could demonstrate the effectiveness of transfer learning in the audio domain, opening up new possibilities for applying AST models to a wider range of audio-related applications.

Project Details:

This project is for engineers who are interested in applied research in the areas of Artificial Intelligence, machine learning and/or audio. We will apply machine learning to fine tune a model created by MIT to detect, identify and track sounds of interest.

Hardware:

- **Central Processing Unit (CPU):** Handles general-purpose computations and tasks.
- **Graphics Processing Unit (GPU):** Accelerates the computationally intensive operations involved in training and inference of deep learning models.
- **Memory:** Stores data, instructions, and intermediate results.
- **Storage:** Provides long-term storage for datasets, models, and results.
- **Receive array –** Directional microphone that receives analog audio and outputs digital

Software:

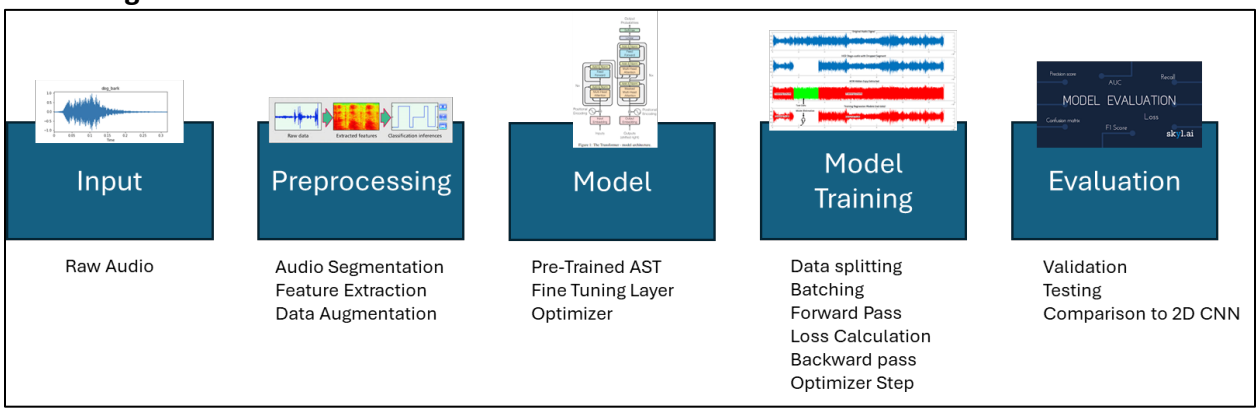
- **Operating System:** Provides a platform for running applications and managing system resources.
- **Python:** The programming language used for most machine learning tasks, including AST fine-tuning.
- **Deep Learning Framework:** TensorFlow or PyTorch, which provide tools and libraries for building and training neural networks.
- **Audio Processing Libraries:** Librosa or Soundfile for loading, processing, and manipulating audio data.
- **Data Manipulation and Analysis Tools:** NumPy, Pandas, and Matplotlib for numerical operations, data analysis, and visualization.
- **Version Control System:** Git or similar tools for tracking changes to code and data.

Overall system concept:
The Blast system aims to enhance the performance of an Audio Spectrogram Transformer (AST) model for a specific audio classification task. BLAST will be a research prototype to determine the feasibility of applying transfer learning in the audio domain. The system will involve:

1. **Data Preparation:** Collecting and preprocessing audio data, including segmentation, feature extraction, and data augmentation.
2. **Model Selection and Loading:** Choosing a suitable pre-trained AST model and loading its weights.
3. **Fine-Tuning:** Training the AST model on the target dataset to tailor its performance to the specific task.
4. **Evaluation:** Assessing the model's performance using appropriate metrics.
5. **Deployment:** Integrating the fine-tuned model into a production environment for real-world applications.

By fine-tuning the AST model, the system aims to improve its accuracy, efficiency, and applicability to specific audio-related tasks.

Block Diagram



Hardware/Electrical Tasks:

Select components and design system for training and inference

- Select microphone for direction audio recording
- Make/buy (e.g. rent from AWS) decision for GPU / CPU
- Design model training system / Design inference system
- Dataset definition and collection (open sources, new recordings)

Fine Tuning Setup mathematics and statistic

- Define Hyperparameters: Configure parameters such as learning rate, batch size …
- Loss Function: Select an appropriate loss function for the task
- Evaluation Metrics: Determine metrics to assess model

Evaluation

- Testing: Assess the final model's performance on the unseen test set using the chosen evaluation metrics.
- Comparison: Compare the results to baseline models or previous state-of-the-art methods (e.g. CNN, audio Signal Processing)

Firmware/Software/Computer Tasks:

Data Preprocessing:

- Audio Segmentation: Divide long audio files into shorter segments for efficient training.
- Audio Feature Extraction: Compute spectrogram representations for each audio segment.
- Data Augmentation: Apply techniques like time shifting, pitch shifting, or adding noise to increase dataset diversity and prevent overfitting.
- Labeling: Assign appropriate labels to each audio segment based on its content.

Model Selection and Loading:

- Choose a Pre-trained AST Model: Select a suitable AST architecture from existing research or libraries (e.g., Hugging Face Transformers).
- Load the Model: Load the pre-trained weights of the AST model.

Training Process:

- Data Splitting**:** Divide the dataset into training, validation, and testing sets.
- Model Training**:** Iterate through the training data, feeding input spectrograms and corresponding labels to the model.
- Backpropagation: Update model parameters based on the calculated loss.
- Validation: Evaluate the model's performance on the validation set to monitor progress and prevent overfitting.

6. Deployment and Usage:

- **Save the Model:** Save the fine-tuned model for future use.
- **Integration:** Integrate the model into applications or systems that require audio analysis.

Composition of Team:

1 Electrical Engineer & 1 Computer Engineer ***(preference will be given to those engineers who will be taking Dr. Chiovaro's course on Thursday evenings!)***

Skills Required:

Electrical Engineering Skills Required:

Machine learning: A strong understanding of machine learning concepts, including supervised learning, neural networks, and deep learning architectures.
Audio signal processing: Knowledge of audio signal processing techniques, such as Fourier transforms, spectrograms, and time-frequency analysis.
Audio domain knowledge: Understanding of audio-related concepts, such as audio features, signal-to-noise ratio, and audio quality metrics.

Computer Engineering Skills Required:

- **Python programming**: Proficiency in Python. Understanding of basic Python constructs like variables, data types (integers, floats, strings, lists, tuples, dictionaries), control flow statements (if-else, loops), and functions.
- **Deep learning frameworks**: Familiarity with popular deep learning frameworks like TensorFlow or PyTorch for building and training AST models.

Anticipated Best Outcome's Impact on Company's Business, and Economic Impact

The best economic outcome for a company that implemented this project would be a portable prototype to demonstrate the concept at conferences and business development opportunities resulting in revenue growth.

- **Improved product or service performance:** A fine-tuned AST model will enhance the accuracy and efficiency of our products or services that rely on audio/acoustic analysis, leading to increased customer satisfaction and market share.
- **New product or service development:** A successful AST model will enable the development of innovative products or services that leverage advanced audio analysis capabilities, opening new revenue streams.
- **Competitive advantage**: By possessing a superior audio analysis technology, SEACORP can differentiate ourselves from competitors and gain a competitive edge in their market.

Broader Implications of the Best Outcome on the Company's Industry:

Broader Implications of a Rapidly Trainable Audio Spectrogram Transformer System
The development of a rapidly trainable Audio Spectrogram Transformer (AST) system has significant implications across various fields, including:

1. Audio Processing and Analysis

- **Real-time audio applications:** ASTs can enable real-time audio processing tasks such as speech recognition, music transcription, and sound classification.
- **Audio event detection:** ASTs can be applied to detect specific audio events, such as gunfire, car crashes, or animal sounds.

2. Machine Learning and AI

- **Transfer learning:** Pre-trained AST models can be used as a starting point for transfer learning tasks in other audio-related domains.
- **New applications:** ASTs can enable new applications in machine learning and AI, such as audio-based anomaly detection, sentiment analysis, and personalized audio recommendations.

3. Industry and Commerce

- **Improved customer experience:** ASTs can be used to enhance customer experiences in various industries, such as call centers, entertainment, and healthcare.
- **Increased efficiency:** ASTs can improve efficiency in tasks such as quality control, surveillance, and content moderation.
- **New products and services:** The development of ASTs can lead to the creation of new products and services, such as intelligent audio assistants, audio-based search engines, and personalized audio experiences.

4. Research and Development

- **Scientific discoveries:** ASTs can be used to analyze and understand complex audio signals, leading to new scientific discoveries in fields such as acoustics, neuroscience, and linguistics.
- **Medical applications:** ASTs can be applied to medical research, such as diagnosing diseases based on audio signals from the heart, lungs, or other organs.
- **Educational tools:** ASTs can be used to develop educational tools for teaching music theory, speech therapy, and other audio-related subjects.

In conclusion, the development of a rapidly trainable Audio Spectrogram Transformer system has the potential to revolutionize the field of audio processing and analysis, with far-reaching implications across various industries and research areas