Sample DATA 586 Project Proposal

Background and Motivations

Marine mammals, including dolphins and whales, confront significant risks from intense maritime activities such as military sonar operations and heavy ship traffic. Protecting these animals necessitates swift detection and identification techniques. Traditional methods rely on direct human observation through passive acoustic monitoring or telescopic visuals, while newer strategies involve analyzing spectrograms and employing machine learning for automatic classification. Despite advancements, these approaches face limitations: they often depend on specialist involvement, are affected by adverse weather conditions, and generally exhibit slow response times.

Objectives

Hence, leveraging the progress in deep learning, I aim to develop a deep-learning model that integrates a Residual Network (ResNet) [1] for capturing spatial details and a lightweight Vision Transformer (ViT) [2] for understanding temporal dynamics.

Detailed Activities

Initially, I will conduct a thorough review of current deep learning methodologies for classifying marine mammals. Following this, I will compile all accessible public databases containing marine mammal sounds. With these resources, I'll construct and train my network on the collected datasets. The final step involves refining the model through fine-tuning for optimal performance.

Deliverables

The project will yield a meticulously trained and finely-tuned sound classifier capable of accurately identifying at least ten different marine mammal species. Furthermore, the model will be designed for compactness, ensuring its adaptability and deployment across a variety of platforms. Finally, a project report and a presentation will be submitted.

Foreseen Challenges

I anticipate encountering several potential obstacles. Firstly, the scarcity of training data could hinder the development of a highly accurate model. Secondly, limitations in computational resources might pose a challenge; either due to the constraints of my current GPU's capabilities or by exceeding the free usage limits on platforms like Colab.

References

1. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
2. S. Mehta and M. Rastegari, "MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer," *2022 The International Conference on Learning Representations (ICLR).*