Development of self-supervised methods for radiology in brain evolution research

In collaboration with AIRI

Student(s)
Bair Mikhailov, Bair.Mikhailov@skoltech.ru, +79148894878

Program
Data Science

Skoltech supervisor
Dmitry Dylov

Introduction

One of the challenges in studying brain evolution is the inherent fragility of the brain itself. To overcome this issue, scientists use endocasts (Figure 1 and 2)—natural molds of the brain's shape that form in the space within the skull. By examining these endocasts, researchers can study the brains of extinct animals and compare them with those of modern species. Consequently, applying machine learning techniques to analyze endocasts can be a highly effective approach for investigating brain evolution.

Objectives

The main goal of this project is to develop a machine learning model for the classification of 3D braincase casts and then apply this model to classify braincase casts of debatable species.

Process & Methods

First, I prepared the dataset by converting the scans into a manageable format, reducing their “resolution” to make them more computationally efficient, and splitting the data into training and validation subsets. Additionally, I removed the endocasts of juvenile specimens, as they are particularly difficult to classify.

Next, I utilized this refined dataset to fine-tune a pretrained PointNet model, which is specifically designed for processing 3D point cloud data. During this fine-tuning phase, I addressed the issue of class imbalance by modifying the cross-entropy loss function. This adjustment ensured that the model did not become biased towards the more prevalent class in the dataset.

After successfully training the model, I applied it to a set of debatable species whose classification between crocodiles and alligators has been uncertain.

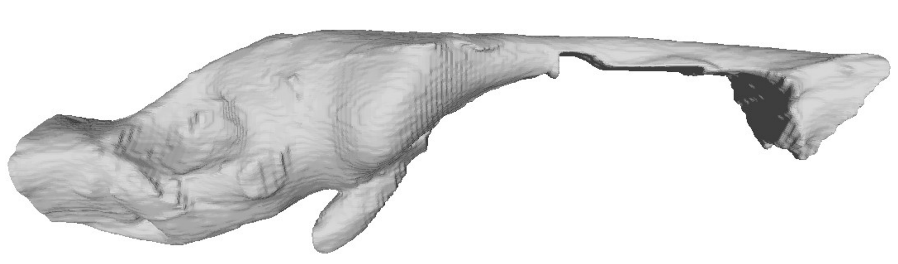

Figure 1: Crocodile’s brain endocast.

Figure 2: Alligator’s brain endocast.

Results

On the training data (Figure 3), the model achieved an accuracy of 92%, primarily misclassifying alligators as crocodiles.

On the validation data (Figure 3), the model demonstrated an accuracy of 80%, again with a tendency to incorrectly classify alligators as crocodiles.

One of the main reasons for these issues is the overall lack of data and the skewed distribution within the dataset.

For the debatable species (Figure 4), the model classified more than 86% of the dataset as crocodiles. This result aligns with certain hypotheses about species classification. However, the model showed relatively low confidence in its predictions for each individual case.

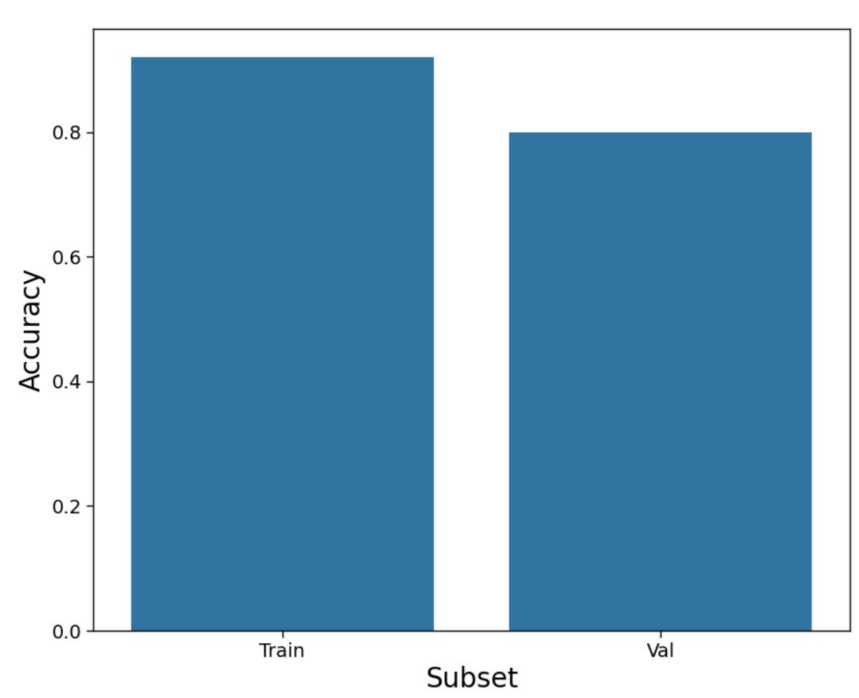

Figure 3: Accuracy on test and validation subsets.

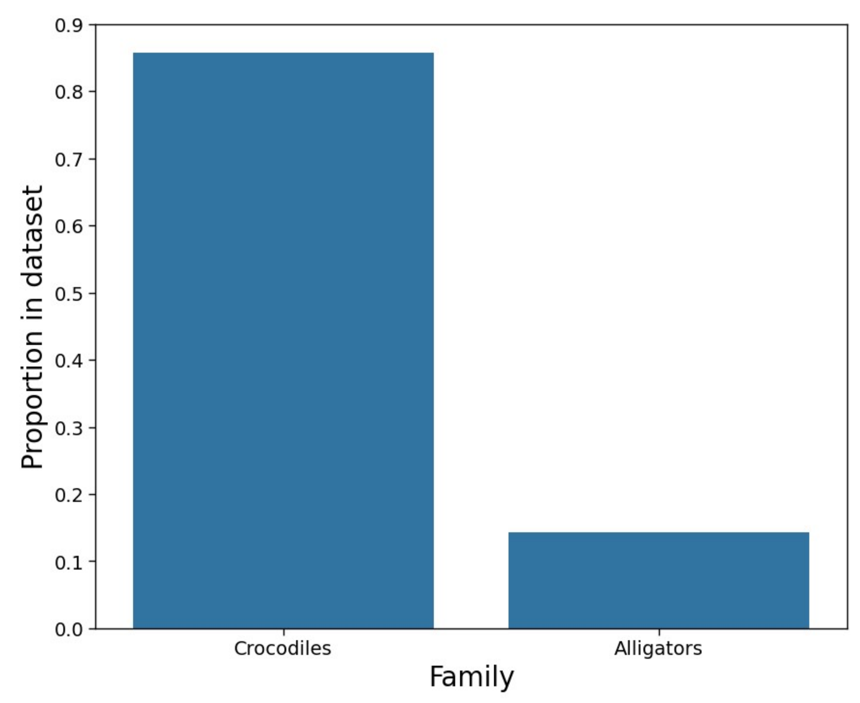

Figure 4: Debatable species classification.

Conclusions

- The results of this project can serve as valuable evidence in the determination and classification of species whose categorization has been subject to debate among experts. This adds a new dimension to the ongoing discussions and provides a scientifically grounded perspective.
- It was particularly exciting to utilize my knowledge and skills in an interdisciplinary field, where the integration of various scientific domains allowed for a more comprehensive approach to solving complex problems. This experience has broadened my understanding and application of machine learning.
- This project stands out as one of the pioneering applications of machine learning in the field of paleoneurobiology.
- I would like to thank Victor Gombolevsky, Ivan Kuzmin and Evgenia Mazur for their help and supervision during the project.