Student: *Bair Mikhailov*
Research Advisor: *Dmitry Dylov*

Data Science

April, 2025

Development of Machine Learning Methods for Radiology in Brain Evolution Research

Skoltech

General problem / Introduction / Background

One of the challenges in studying brain evolution is the inherent fragility of the brain itself.

To overcome this issue, scientists use endocasts—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.

General problem / Introduction / Background

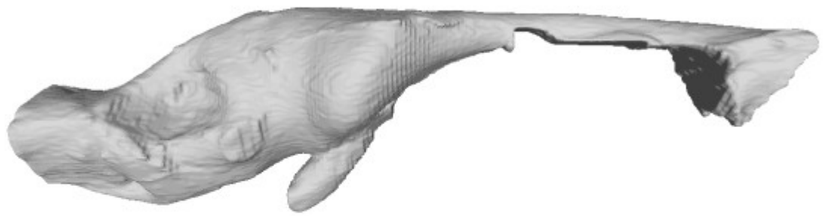


Figure 1: Crocodile's brain endocast.

Figure 2: Alligator's brain endocast.

General problem / Introduction / Background

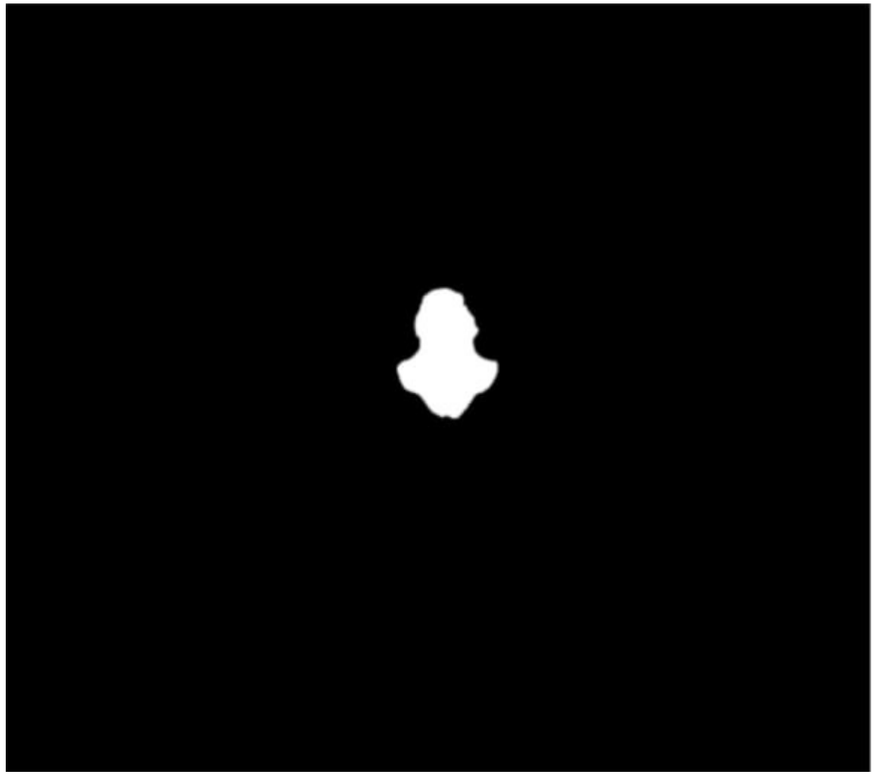


Figure 3: Segmentation of crocodile’s brain endocast.

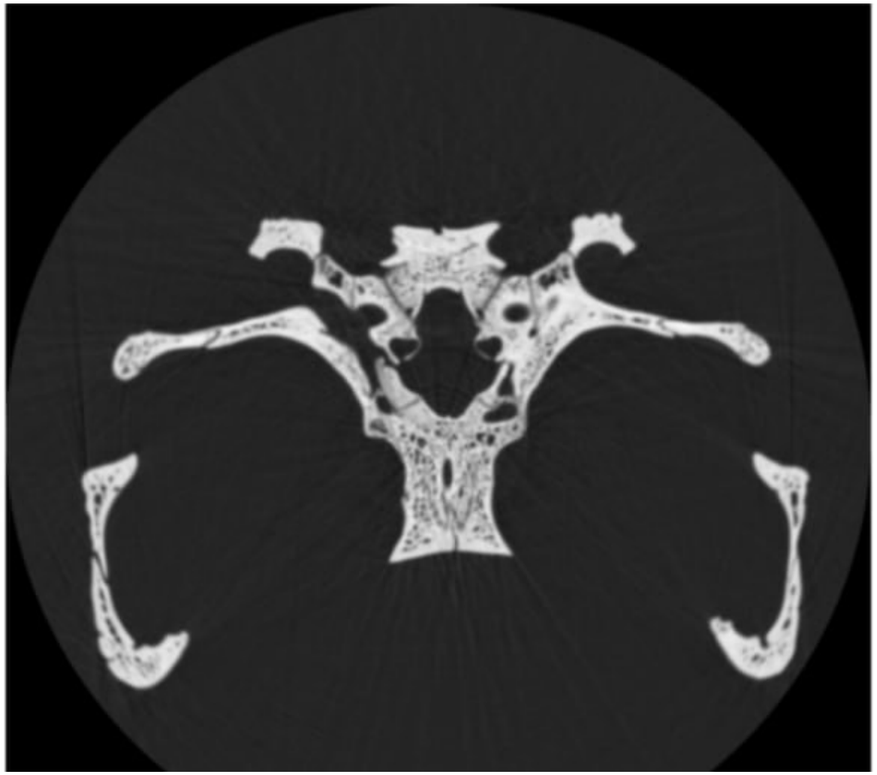


Figure 4: CT of crocodile’s brain

General problem / Introduction / Background

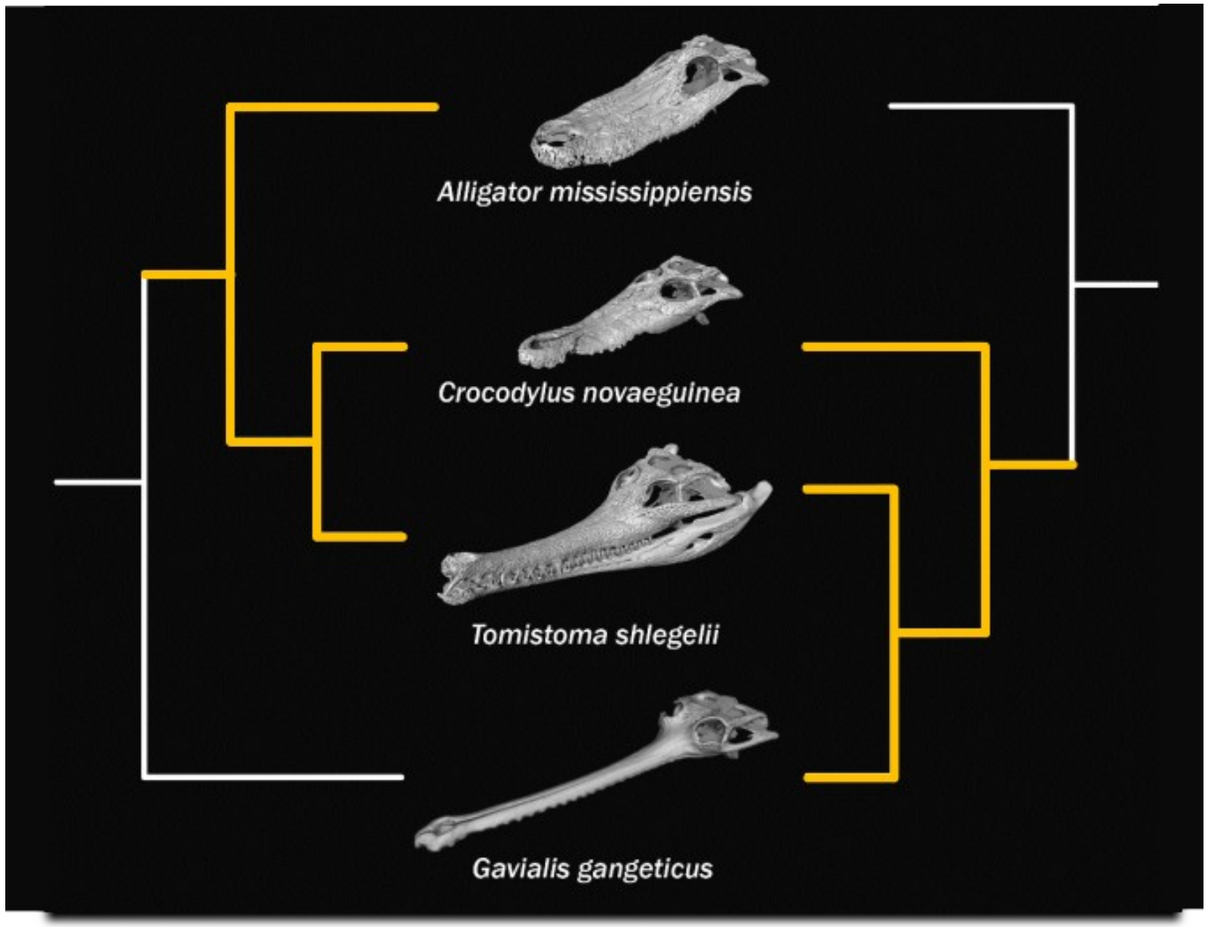


Figure 5: 2 hypothesis of familiar connections of crocodiles.

Aim

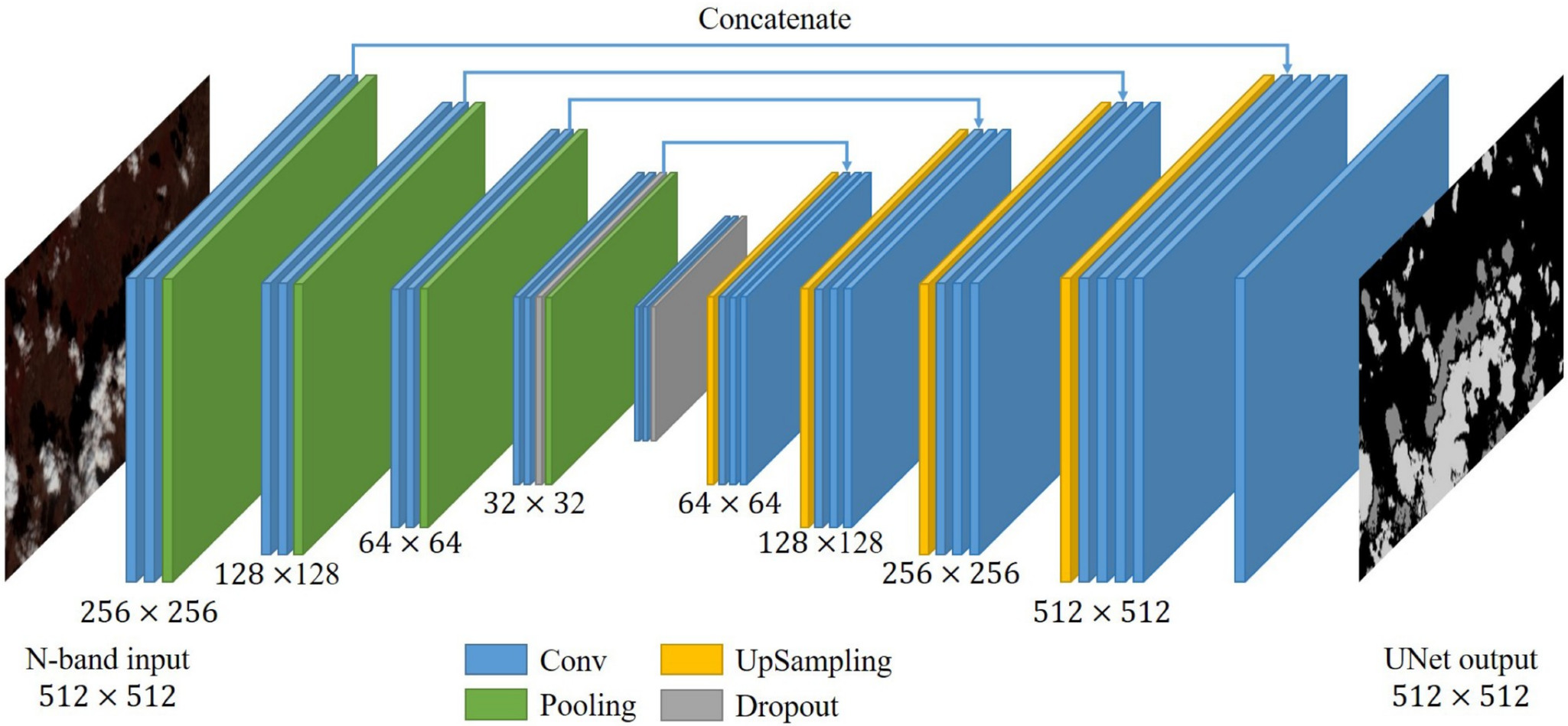
The main goal of the work is to develop a machine learning model for the segmentation of of 3D braincase cast, then its classification, after that application of this model to classify braincase casts of debatable species and interpret its results.

Objectives

The objectives of the present research run as follows:

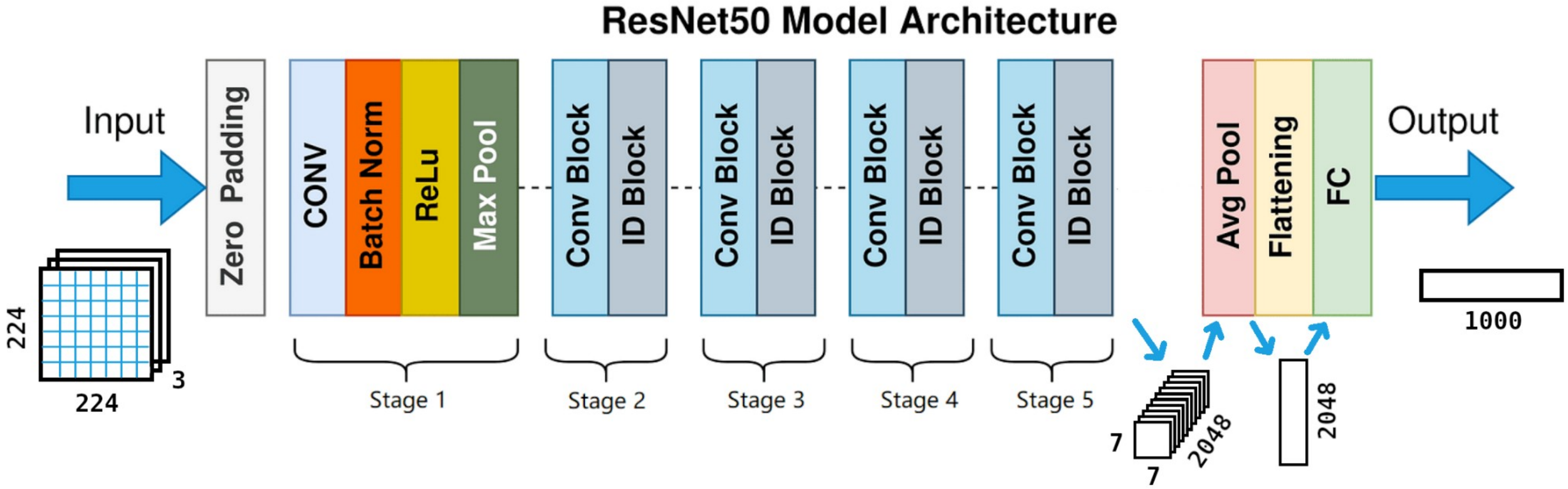
1. To segment the endocasts of all available crocodilians
2. To classify the debatable species of crocodilians using their endocasts
3. To interpret the results of the classification above

Theory and Algorithms

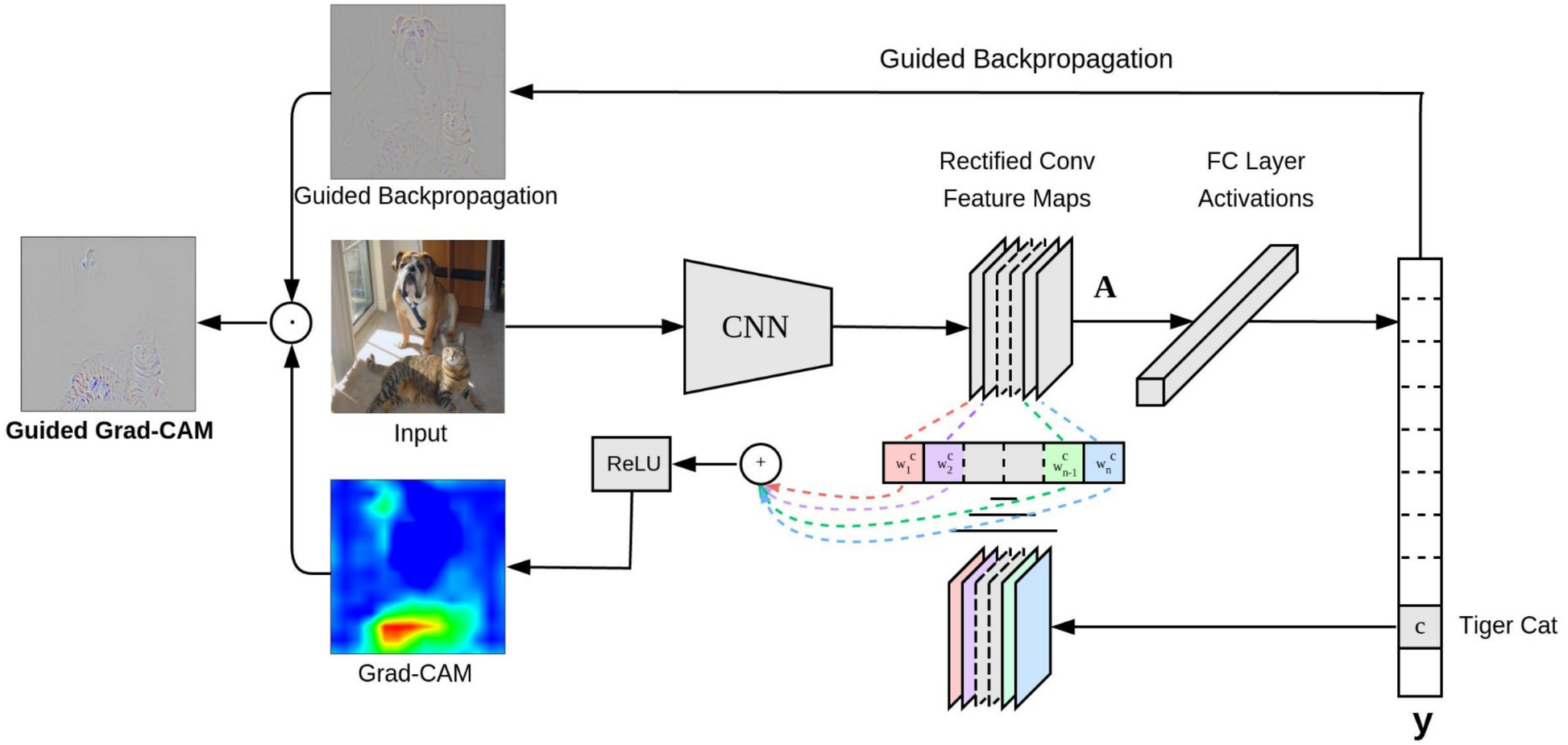


Jiao, L.; Huo, L.; Hu, C.; Tang, P. Refined UNet: UNet-Based Refinement Network for Cloud and Shadow Precise Segmentation. *Remote Sens.* **2020**, *12*, 2001. https://doi.org/10.3390/rs12122001

Theory and Algorithms



Theory and Algorithms



Selvaraju, Ramprasaath R., et al. "Grad-CAM: visual explanations from deep networks via gradient-based localization." *International journal of computer vision* 128 (2020): 336-359.

Methodology / Experimental setup

Data description:

36 3D CT images of alligators (.dcm), **24** images of crocodiles, **7** images of tomistomas and gavials.

Among them:
29 3D CT images of crocodilians with brain endocast segmentation (.tif)

Methodology / Experimental setup

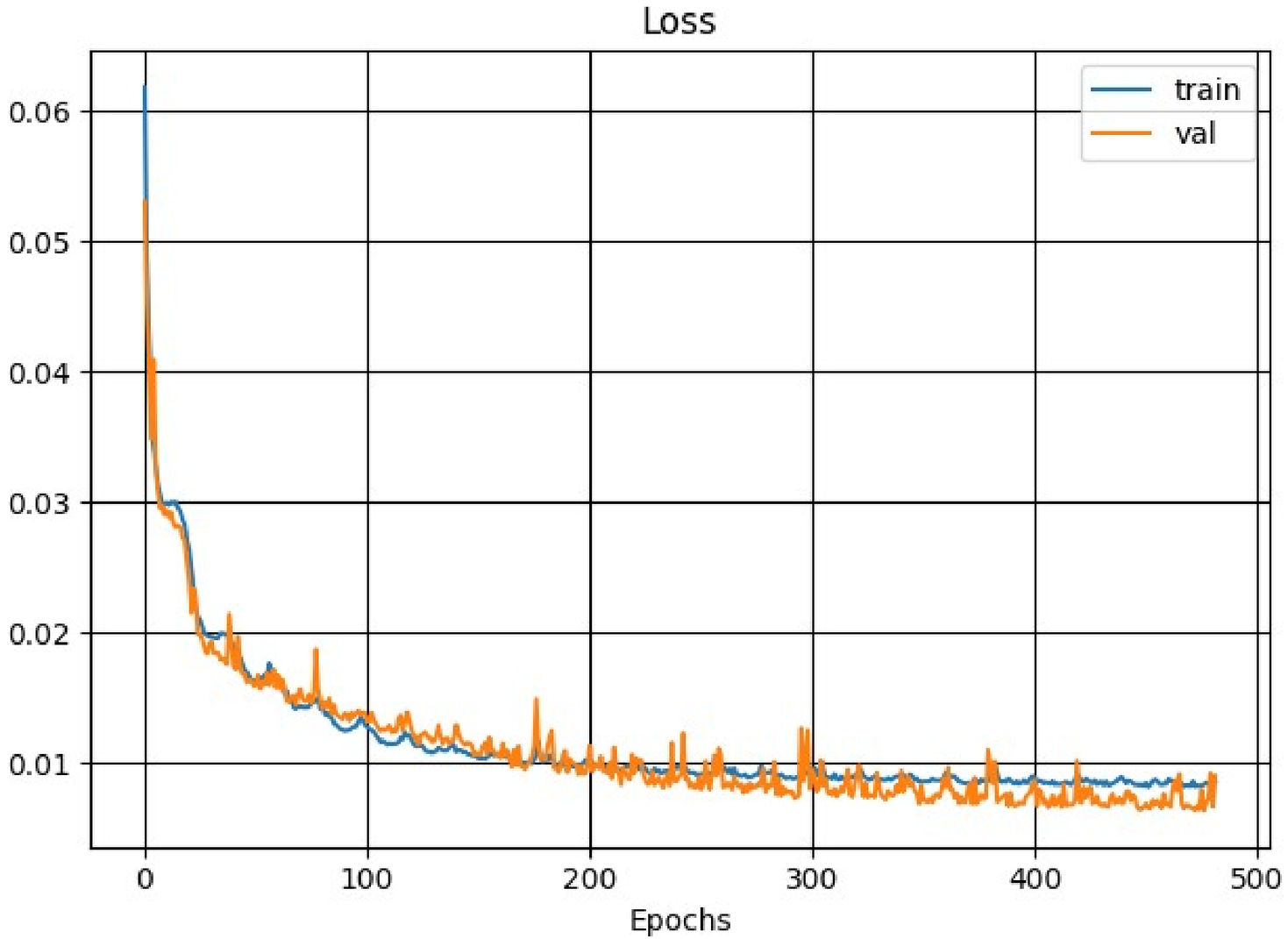
1. **Data Preparation**
- **Dataset**: 29 3D CT scans (24 train / 5 validation)
- **Preprocessing**:
  - Resampled to 256×256 2D slices
  - Augmentations: Random flips, rotations (±60°), intensity scaling
2. **Segmentation Model**
- **Architecture**: 2D U-Net (PyTorch)
  - Encoder-Decoder with skip connections
  - Output: Binary mask (sigmoid activation)
- **Training**:
  - Loss: DiceLoss + BCELoss
  - Optimizer: AdamW (lr=1e-5)
  - Scheduler: CosineAnnealingLR (T_max=10)
  - Epochs: 480

Methodology / Experimental setup

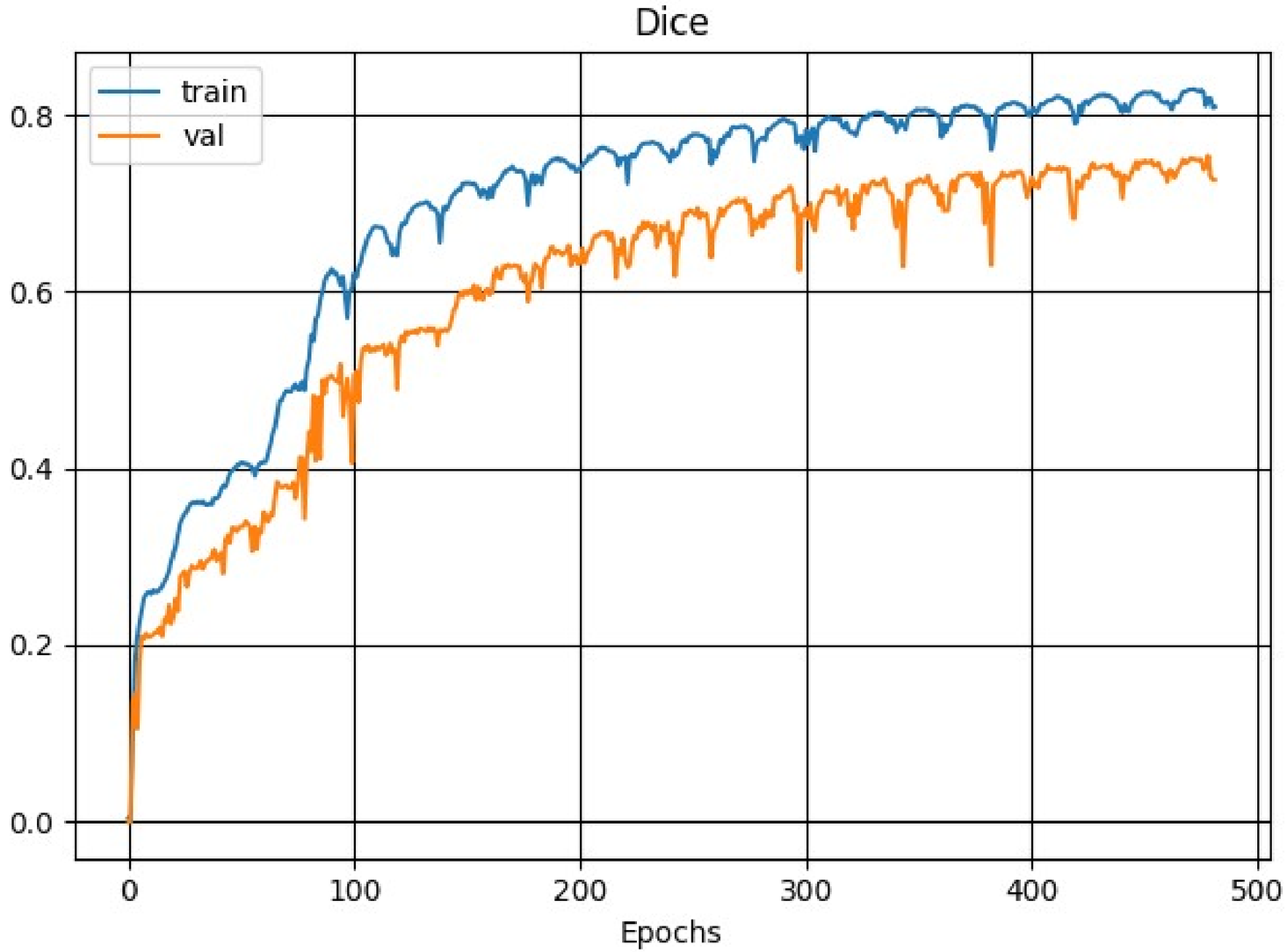
3. Classification & Interpretation

- **Model**: ResNet-50 (pretrained on ImageNet)
  - Input: 224×224 endocast slices
  - Output: 2-class probabilities
- **Explainability**: Grad-CAM heatmaps
  - Highlights neuroanatomical discriminators

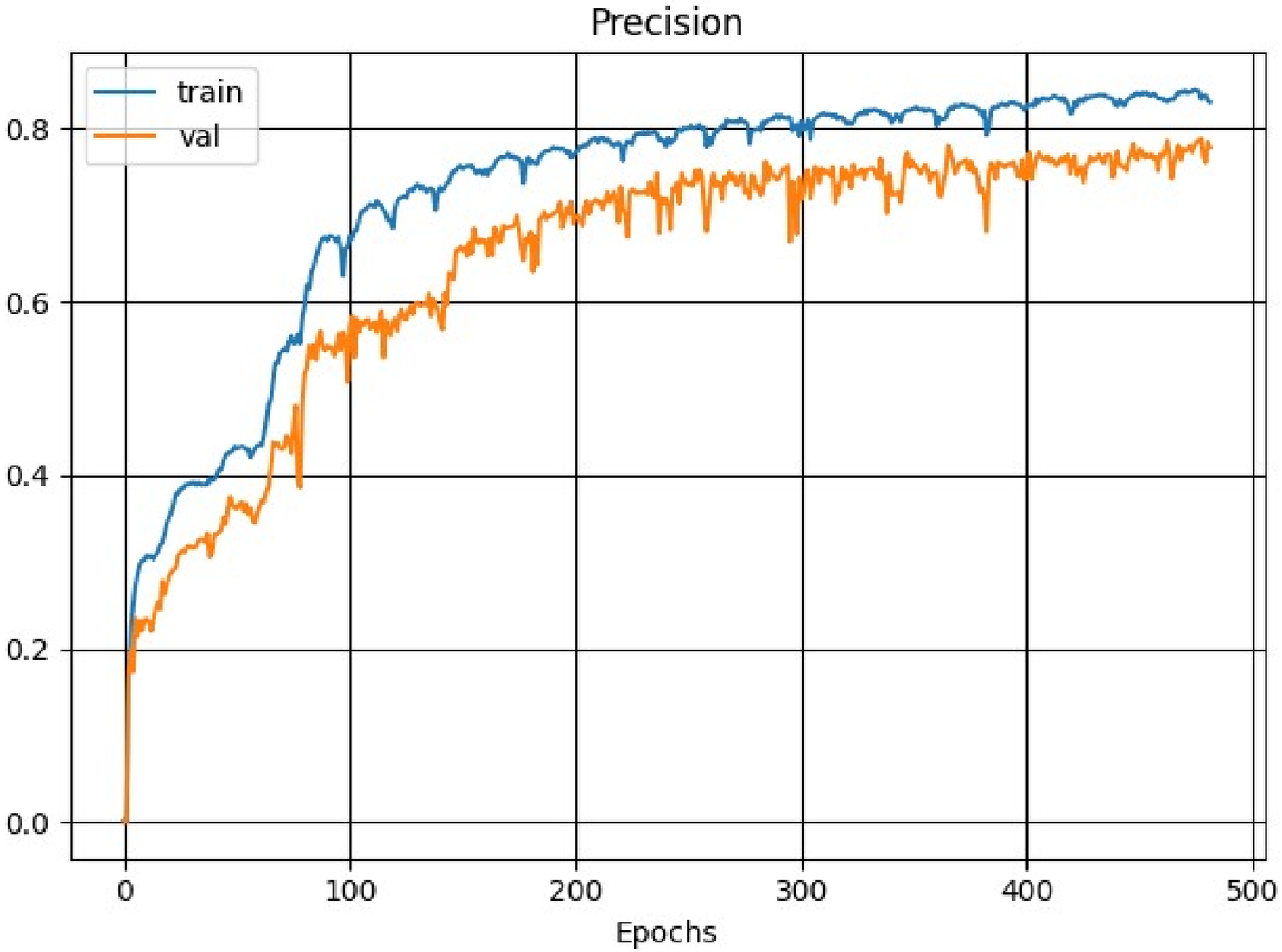
Results of the Experiment



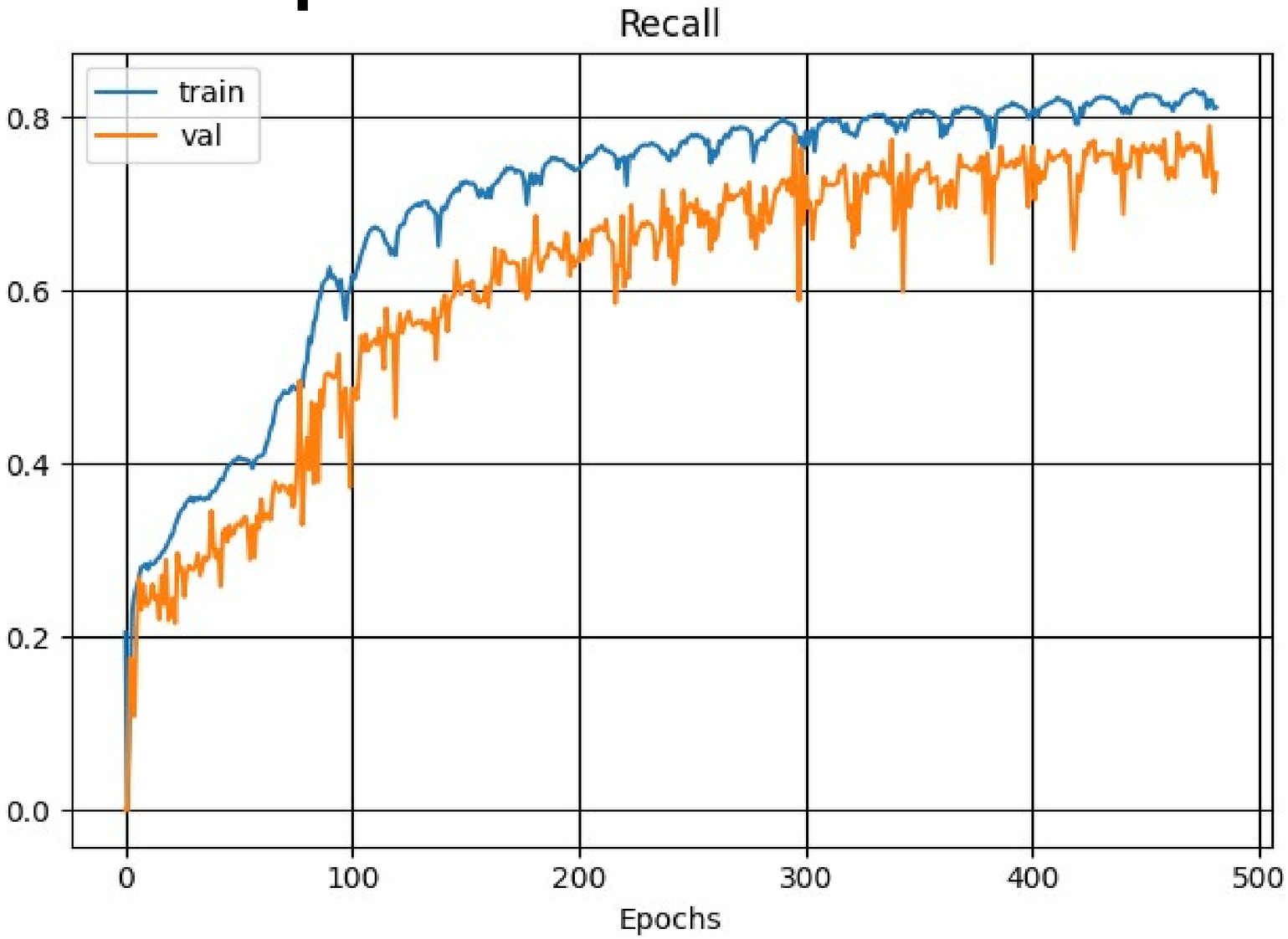
Results of the Experiment



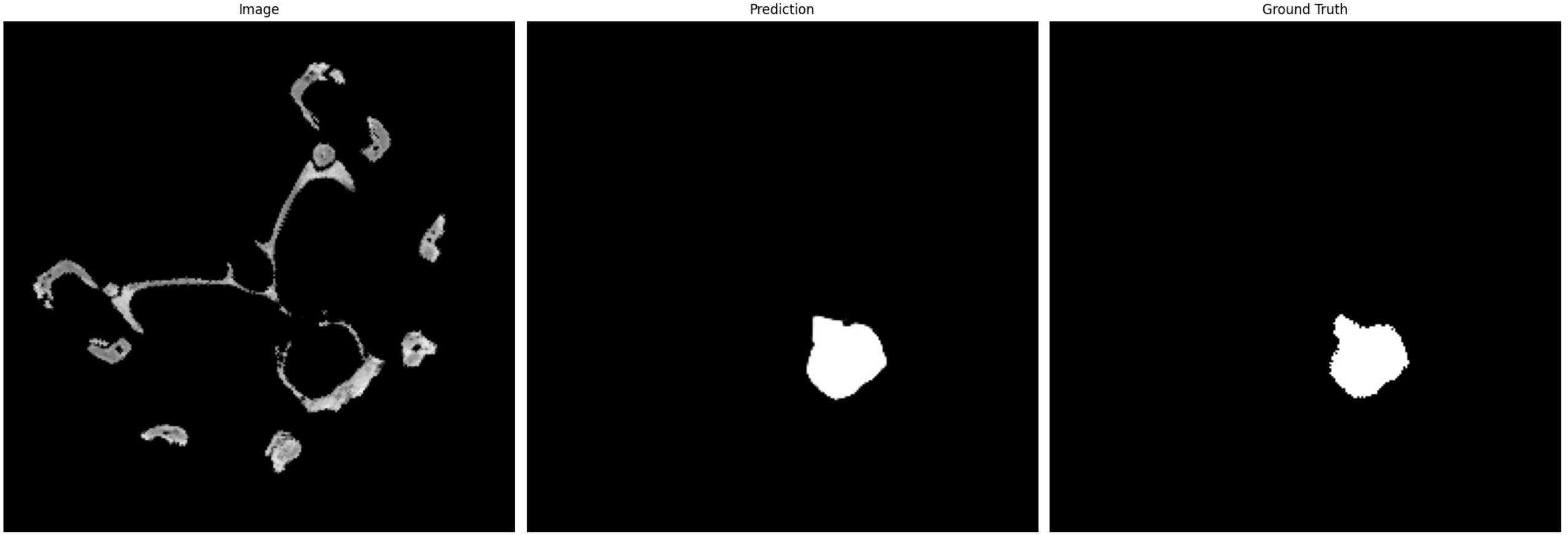
Results of the Experiment



Results of the Experiment



Results of the Experiment



Scientific novelty

The State-of-the-art component of my research project is the application of Machine Learning for classification (and its reasoning) of the debatable species.

Conclusions

1. 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.

2. This project stands out as one of the pioneering applications of machine learning in the field of paleoneurobiology

Current Status

1. Finished with endocast segmentation
2. Almost finished with classification
3. Started working on interpretation
4. Started working on 3D implementation

Outlook

The achieved results indicate that the work along the same line should proceed. Thus, the long-term results of the project are to:

1. Apply the pipeline to the fossils

2. Test the pipeline on other species of animals

Acknowledgements

I would like to express my sincere gratitude to researchers of SPbSU Department of Vertebrate zoology Ivan Kuzmin and Evgenia Mazur for suggesting the initial project and providing the required data