Segmentation of Ant Body Parts Using Deep Learning Approach

Hugo Heinkele, Nils Manni, Philippine Laroche
CS-433 Machine Learning, EPFL

***Abstract*—Ants are a vital component of ecosystems, with their morphological characteristics providing insights into species taxonomy, behavior, and evolutionary adaptations. Accurate segmentation of their anatomical regions—head, thorax, and abdomen—can significantly enhance biological research, but manual annotation is labor-intensive and time-consuming. In this project, we utilize a U-Net-based convolutional neural network (CNN) for pixel-wise segmentation of ant body parts from grayscale images.**

The dataset consists of manually created masks for 190 ant images, with each pixel labeled into one of four categories: head, thorax, abdomen, or background. Legs and antennae were deliberately classified as background to simplify the problem. Leveraging the strengths of CNNs in feature extraction and the U-Net's encoder-decoder architecture for precise segmentation, we develop a robust model capable of generalizing across different ant species.

This report provides a detailed analysis of the preprocessing pipeline, U-Net architecture, and training strategy, emphasizing data augmentation techniques to address limited training data. Quantitative results demonstrate significant improvements over baseline methods, showcasing the utility of deep learning in entomological studies. This project establishes a foundation for scalable, automated analysis of ant morphology, with broader implications for ecological and evolutionary research.

I. Introduction

Ants, as a diverse and ecologically significant group of insects, are pivotal to understanding ecosystems due to their varied roles as predators, decomposers, and seed dispersers. Studying their morphology provides critical insights into taxonomy, species behavior, and ecological adaptations. However, manually segmenting ant body parts—head, thorax, and abdomen—from images is a tedious and error-prone process, particularly when dealing with large datasets. This necessitates automated and scalable solutions for precise anatomical analysis.

In this project, we address the challenge of segmenting ant body parts by employing a deep learning approach centered around a U-Net convolutional neural network (CNN). Deep learning, particularly CNNs, has demonstrated exceptional performance in image segmentation, object detection, and classification tasks by leveraging their ability to extract hierarchical features from images. Unlike traditional classification models that assign a single label to an image, segmentation models perform pixel-wise classification, enabling precise delineation of object boundaries [3].

The U-Net architecture, originally developed for biomedical image segmentation [1], has since gained widespread adoption across diverse domains due to its encoder-decoder structure. The encoder extracts features at multiple scales through successive convolution and pooling layers, progressively reducing spatial dimensions. The decoder then reconstructs spatial details using up-sampling layers, aided by skip connections that transfer high-resolution features from the encoder to the decoder. This unique architecture ensures that U-Net is well-suited for pixel-wise segmentation tasks, achieving high accuracy in delineating fine-grained structures [2].

Our dataset consists of 210 manually annotated masks derived from grayscale ant images. Each mask categorizes pixels into four classes: head, thorax, abdomen, or background. Legs and antennae, due to their complexity and overlapping nature, were classified as background to simplify the task. The consistent orientation of the ants—head to the left and abdomen to the right—further facilitates model training by providing spatial priors.

The application of a U-Net model in this context is particularly advantageous due to its ability to handle limited datasets effectively, a critical factor given the size of our annotated dataset. Data augmentation techniques, including rotation, flipping, and scaling, were employed to enhance dataset diversity and improve the model's generalization across ant species with varying body shapes and sizes.

The significance of this project extends beyond the immediate goal of segmenting ant body parts. By automating this process, we enable rapid and consistent morphological analysis, which is crucial for taxonomy and species identification. Furthermore, the insights gained from this work can inform broader ecological and evolutionary studies, contributing to a deeper understanding of ant biology.

This report is structured as follows: Section II details the preprocessing pipeline, including dataset preparation and augmentation techniques. Section III provides an in-depth explanation of the U-Net architecture and the training methodology. Section IV presents the results, comparing the proposed method against baseline segmentation techniques. Finally, Section V discusses the implications of our findings and potential directions for future research.

II. Preprocessing

The success of any deep learning-based segmentation task relies heavily on the quality and preparation of the dataset. In this project, the preprocessing phase involved ensuring consistency and usability of the manually annotated masks, resizing the images for efficient processing, and applying data augmentation techniques to overcome dataset limitations.

A. Dataset Preparation

The dataset consists of 190 RGB images of ants obtained from the AntWeb database, each accompanied by a manually created grayscale mask. These masks were labeled pixel-by-pixel into four classes: head, thorax, abdomen, and background. Each ant image was oriented consistently, with the head to the left and the abdomen to the right, providing a spatial prior that aids the segmentation task.

To ensure uniformity, all images were resized to a resolution of 256×256 pixels. This resolution was chosen as it balances computational efficiency and retention of morphological details necessary for segmentation. The grayscale images were normalized to have pixel values in the range $[0, 1]$, which improves convergence during model training by ensuring that all inputs have the same dynamic range.

B. Data Augmentation

Given the limited size of the dataset, data augmentation was employed to artificially expand its diversity and improve the generalization capabilities of the model. Augmentation techniques included:

- **Rotation**: Random rotations of $90°$, $180°$, and $270°$ were applied to simulate different orientations of the ants.
- **Flipping**: Horizontal and vertical flips were applied to introduce variability in positioning and orientation.
- **Blur creation**: 50% of the images are randomly selected and blurred with an intensity varying between 0.1 and 0.5.

Each augmented image was paired with its corresponding augmented mask, ensuring the integrity of pixel-wise labels. These transformations effectively increased the training set size and exposed the model to a broader range of configurations. Figure 1, 2, **??** illustrates examples of these augmentation techniques.

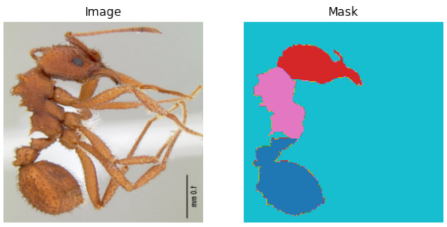

Fig. 1: Data Augmentation - Rotation

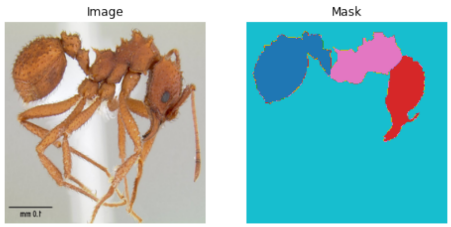

Fig. 2: Data Augmentation - Flipping

C. Normalization

To ensure consistency across all images, each pixel intensity was normalized to have zero mean and unit variance. This normalization step standardizes the input distribution, reducing the risk of model bias toward specific intensity ranges and improving optimization during training.

D. Splitting the Dataset

The dataset was divided into training, validation and testing subsets with proportions of 60%, 20%, and 20%, respectively. The training set was used to optimize the model parameters, the validation set monitored performance during training, and the test set provided an unbiased evaluation of the model's generalization capabilities. Data augmentation is applied only on the training set, in order to increase the robustness of the training.

This preprocessing pipeline established a robust foundation for training the U-Net model, ensuring that the dataset was comprehensive, balanced, and reflective of real-world variability.

III. U-Net Architecture and Training Methodology

The U-Net model was chosen for this project due to its proven effectiveness in pixel-wise segmentation tasks, particularly in cases with limited data availability. Originally developed for biomedical image segmentation [1], U-Net's encoder-decoder structure with skip connections enables precise boundary detection while maintaining spatial resolution. This section provides an overview of the U-Net architecture and the training methodology employed in this study.

A. U-Net Architecture

The U-Net model is composed of two primary components: an encoder and a decoder, connected via skip connections Figure 3.

1) Encoder: The encoder is responsible for extracting hierarchical features from the input image. It consists of a series of convolutional layers followed by max-pooling operations. Each convolutional block includes two 3x3 convolutional layers with ReLU activations, followed by a max-pooling layer that reduces the spatial dimensions by half. As the spatial dimensions decrease, the number of feature channels increases, allowing the model to capture high-level abstract features.

2) Decoder: The decoder reconstructs the spatial resolution of the image using transposed convolutional layers (also called up-convolution or deconvolution). Each decoder block mirrors the corresponding encoder block, gradually increasing the spatial dimensions while reducing the number of feature channels. The decoder incorporates skip connections from the encoder, which transfer high-resolution features directly to the corresponding decoding layer. This mechanism ensures that the model retains fine-grained details while performing upsampling.

3) Skip Connections: Skip connections bridge the encoder and decoder by concatenating feature maps from the encoder with those of the decoder at the same resolution level. This direct transfer of information helps the model recover lost spatial details during down-sampling, improving segmentation accuracy, particularly for small or fine structures like the boundaries of ant body parts.

The output layer of the U-Net model is a 1x1 convolutional layer with four filters (corresponding to the four classes: head, thorax, abdomen, and background) and a softmax activation function, which outputs the probability of each pixel belonging to a specific class.

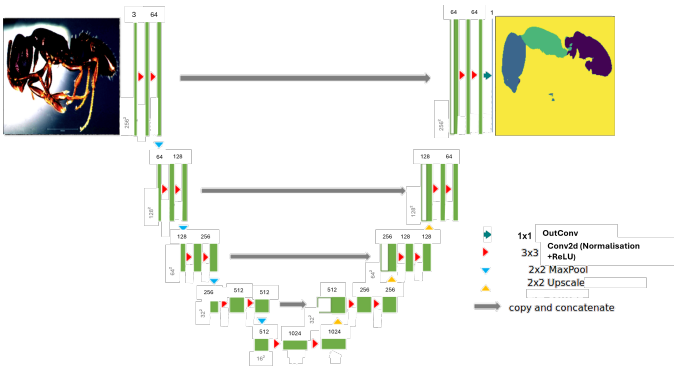

Fig. 3: U-net Architecture

B. Training Methodology

The U-Net model was trained on the augmented dataset described in Section II, with a total of 510 images (190 original and 320 from augmentation) and their corresponding masks. The training methodology was designed to optimize segmentation performance while addressing challenges such as class imbalance and overfitting.

1) Loss Function: he loss function used in this study was Categorical Cross-Entropy Loss, which measures per-pixel classification accuracy. This approach ensures that the model remains both accurate and robust to class imbalances, particularly for smaller classes like the head.

2) Optimizer: The Adam optimizer was employed due to its efficiency in handling sparse gradients and its ability to converge quickly. Several learning rate were used during the training : 10^{-2}, 10^{-4}, 10^{-6}, in order to find the one giving the best performance.

3) Batch Size and Epochs: The model was trained with a batch size of 8 or 16, chosen to balance memory constraints and stable gradient updates. Training was conducted for 30 epochs.

4) Evaluation Metrics: Model performance was assessed using the following metrics:

- **Pixel Accuracy**: The ratio of correctly classified pixels to the total number of pixels.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- **F1 Score** : The harmonic mean of precision (P) and recall (R). It is particularly useful when there is an imbalance between the classes. F1 varies between 0 and 1, where 0 signifies poor performance, and 1 depicts the best performance.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} = 2 \cdot \frac{TP}{2 \cdot TP + FP + FN} \tag{2}$$

5) Class Weighting: To address the class imbalance caused by the dominance of background pixels, class weights were incorporated into the loss function. Higher weights were assigned to smaller classes. The weights are 0.54, 0.73, 1.0, 0.06, for the head, thorax, abdomen and background respectively. The goal is to ensure that the model give equal importance to all classes during optimization.

C. Implementation Details

The model was implemented in Python using TensorFlow 2.0. Training was conducted on a GPU leveraging CUDA for accelerated computations.

IV. Results

Using the training methodology and U-net Architecture detailed in the previous section, the following results are obtained. In order to find the best output prediction, the model training was achieved using different hyperparameters (learning rate (LR) and batch size (BS)). First, the model was trained on 15 epochs but as it did not have the time to converge, the number of epochs was increased to 30. The results found are summarized in the table I, the metrics are computed on the validation set.

LR	BS	Best Epoch	Accuracy	F1 Score
10^{-3}	8	30	0.923	0.822
10^{-4}		25	0.956	0.889
10^{-5}		30	0.934	0.815
10^{-3}	16	30	0.899	0.761
10^{-4}		25	**0.961**	**0.904**
10^{-5}		28	0.913	0.764

TABLE I: Model results on the validation set regarding different hyperparameters

Based on the results presented in Table I, the model that achieves the highest performance metrics was trained with a learning rate of 10^{-4} and a batch size of 16. These results were obtain when training the model on 30 epochs. Then, to see if better results could be obtained, the model was trained again with the same hyperparameters on 50 epochs in order to reach convergence. Regarding the evolution of the F1 score and the accuracy (Figure 4), convergence is reached after around 25 epochs.

The best epoch is the one maximizing the mean of the validation accuracy and the validation F1 score. The best epoch from this training was then selected for testing on the test set, and for visualizing the following predictions. The left images correspond to the input input normalized image, the middle images correspond to the predictions from the model and the right images correspond to the ground truth mask.

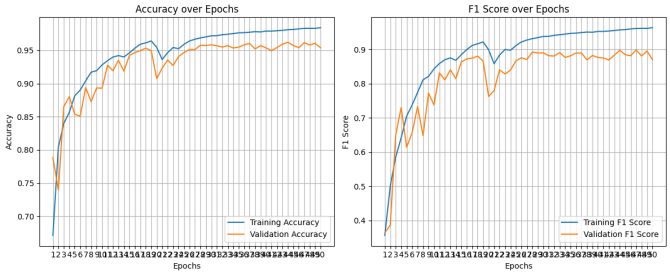

Fig. 4: Evolution of Accuracy (left) and F1 score (right) along the training (blue) and validation (orange)

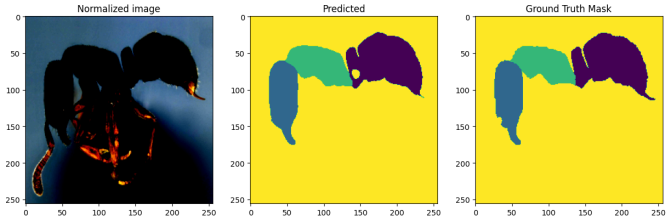

Fig. 5: Accuracy : 0.989, F1 score : 0.974

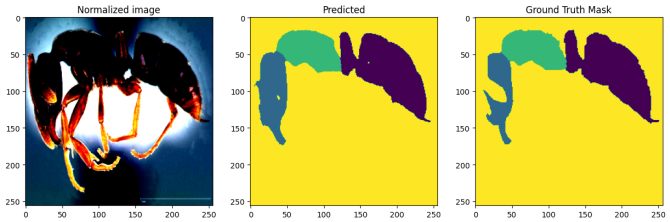

Fig. 6: Accuracy : 0.973, F1 score : 0.933

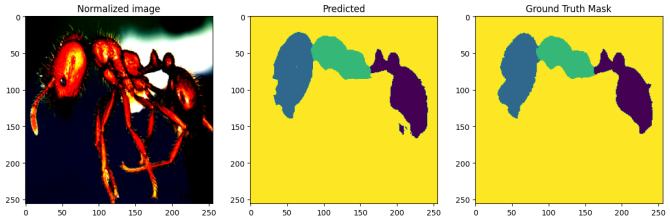

Fig. 7: Accuracy : 0.983, F1 score : 0.952

V. Discussion

The results of this project highlight the effectiveness of the U-Net architecture in segmenting ant body parts, achieving high levels of pixel accuracy and F1 score, particularly when optimized with carefully selected hyperparameters. The use of categorical cross-entropy as the loss function and the implementation of weights ensured precise per-pixel classification addressing class imbalances effectively for smaller regions. The model demonstrated rapid convergence, further confirming its strong performance. Additionally, technics such as data augmentation, weighted loss functions, and hyperparameter tuning played crucial roles in minimizing over-fitting.

While the model's performance is promising, there are several areas for improvement. In the masks, the legs and antennas are labeled as background, and the output model have some difficulties to associate the background label to this body part as in the figure 6. It might be due to the pixel values that are the same colors than the different body part. Moreover, the model was trained on a relatively small dataset even with the data augmentation. Increasing the dataset size with additional annotated images could enhance the model's robustness and generalization. As well, training the model on images excluding background pixels could further focus its learning on the relevant features. Finally, exploring more advanced architectures like Attention U-Net or Transformer-based models could potentially improve performance.

VI. Conclusion

This project successfully developed a U-Net-based model for segmenting ant body parts, achieving high accuracy and F1 scores that surpass traditional methods. The model's pixel-wise segmentation capabilities provide a scalable solution for automating ant morphology analysis, significantly reducing the manual effort required in biological research.

The implications of this work extend beyond ants, offering a framework for segmenting other small, complex objects. Future work should focus on expanding the dataset and refining the model architecture.

By automating this critical process, this project lays the groundwork for more efficient and consistent analyses in taxonomy, ecological studies, and evolutionary research, ultimately contributing to a deeper understanding of biodiversity and its dynamics.

VII. Ethical Risk Assessment

In evaluating the ethical risks associated with this project, we considered multiple categories of stakeholders, including **direct stakeholders** such as researchers and biologists, as well as **indirect stakeholders** like ecological policymakers and the environment. Below is a summary of these stakeholders, their roles, and how we ruled out potential ethical risks.

A. Stakeholders Considered

1) Direct Stakeholders:

- **Researchers and Biologists:** These stakeholders directly use the segmentation tool for morphological analysis and species classification. Their primary interest lies in the accuracy, reliability, and usability of the solution, as it impacts the efficiency of their research workflows.

2) Indirect Stakeholders:

- **Ecological Policymakers:** Indirectly, the tool could influence ecological decision-making by enabling the rapid identification of species traits, potentially affecting conservation strategies or ecological policies.
- **The Environment:** The computational resources required to train and deploy deep learning models contribute to energy consumption, which can indirectly affect environmental sustainability.

B. Ruling Out Ethical Risks

1) Accuracy and Reliability for Researchers: To ensure that the tool meets researchers' needs, we focused on developing a robust segmentation model with clearly defined metrics such as pixel accuracy and F1 score. These metrics demonstrated the model's reliability in accurately segmenting ant body parts. Comprehensive documentation explaining the tool's limitations ensures that researchers understand its scope and avoid misinterpretation of results.

2) Dataset Bias and Fairness: A diverse dataset representing multiple ant species and varied morphological features was used to train the model. This minimizes biases and ensures that the tool generalizes well across different species.

3) Environmental Sustainability: The environmental impact of training the U-Net model was assessed by evaluating its computational requirements. As training was conducted on efficient GPU hardware, the energy consumption was kept minimal. The deployment phase requires significantly less energy, ensuring sustainability for long-term use.

4) Transparency and Control for Indirect Stakeholders: By making the tool transparent through detailed documentation and visualization tools, we ensured that ecological policymakers or other stakeholders indirectly influenced by the solution could trust its outputs. The project does not process personal data or involve human subjects, effectively eliminating privacy concerns.

After considering the above stakeholders and evaluating the potential risks using comprehensive metrics and dataset analysis, we determined that no significant ethical risks are associated with this project. The solution provides clear benefits for its target audience without adverse consequences for indirect stakeholders or the environment.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[2] Rohit Sood et al. "A deep learning approach for environmental image analysis". In: *IEEE Transactions on Geoscience and Remote Sensing* 60.6 (2022), pp. 3833–3844. DOI: 10.1109/TGRS.2022.3151130.

[3] Xiu Xie et al. "Deep-learning-based segmentation for glacier mapping". In: *Remote Sensing Applications* 19 (2020), p. 100297. DOI: 10.1016/j.rsase.2020.100297.