Vesuvius Challenge: Resurrectting an ancient library from the ashes of a volcano

Syed Hammad Ali - 23401440
syedhammad.ali@fau.de

10th April 2025

Abstract

The Vesuvius Challenge addresses the reconstruction of carbonized Herculaneum papyri by combining non invasive 3D X ray computed tomography with state of the art deep learning. We present a three stage pipeline preprocessing including denoising and normalization, 3D ResNet U Net segmentation, and morphological post processing to automatically detect ink in volumetric scan slices. Training leverages binary cross entropy and Dice loss augmented with rotation, flip, and zoom transformations; inference outputs are run length encoded for submission. On two representative fragments our best pipeline v2 achieves an $F_{0.5}$ score of 0.64, outperforming the original morphological method (0.63) and the ResNet baseline (0.62). Visual overlays confirm reliable detection of large high contrast ink regions, while remaining challenges include faint and fragmented traces. We discuss future directions such as deeper segmentation architectures, integration of infrared imagery, and targeted data augmentation to further enhance precision and recall, advancing the non destructive recovery of ancient texts.

1 Introduction

The Vesuvius Challenge is a pioneering initiative aimed at deciphering the ancient Herculaneum Papyri, a collection of scrolls that were carbonized during the catastrophic eruption of Mount Vesuvius in 79 AD [1], [2]. These papyri, discovered in the 18th century, have remained unreadable due to their fragile and charred nature. Traditional unrolling techniques risk permanent damage, necessitating advanced computational approaches to extract the hidden texts [3].

Figure 1: The Herculaneum Papyri scrolls

Recent advancements in X-ray computed tomography (X-ray CT scanning) have provided a non-invasive method to visualize the internal layers of the scrolls [4]. However, detecting ink within these scans remains a significant challenge due to the carbon-based composition of the ink, which closely resembles the surrounding papyrus. The Vesuvius Challenge - Ink Detection, hosted on Kaggle, leverages state-of-the-art deep learning and computer vision methodologies to identify subtle variations in the scanned data, enabling the reconstruction of ancient writings without physical interaction [5].

This report presents an overview of the challenge, detailing the computational techniques employed for ink detection and text reconstruction. The implications of successfully deciphering these scrolls extend beyond archaeology, with potential applications in medical imaging, document restoration, and historical preservation [6]. By addressing this problem through machine learning, the Vesuvius Challenge contributes to both classical scholarship and modern technological advancements.

2 Related Work

The challenge of deciphering ancient manuscripts using non-invasive imaging and computational techniques has been a subject of extensive research. Prior works have explored various methods, including X-ray tomography, virtual unwrapping, and deep learning-based ink detection.

Seales et al. [4] pioneered the concept of virtual unwrapping, enabling the extraction of text from rolled or layered manuscripts without physical contact. This technique was further refined through advancements in machine learning and computer vision, allowing for improved segmentation and text reconstruction [2].

Deep learning approaches, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in historical document analysis. Yun and Kim [5] proposed a specialized ink detection framework based on deep neural networks, achieving high accuracy in segmenting ink regions from noisy backgrounds. This method forms the foundation for several modern implementations, including those used in the Vesuvius Challenge.

In addition, studies on image denoising and restoration have contributed significantly to enhancing text visibility in degraded manuscripts. Shimada and Sato [6] investigated the use of generative adversarial networks (GANs) to reconstruct missing portions of ancient texts, demonstrating substantial improvements over traditional filtering techniques.

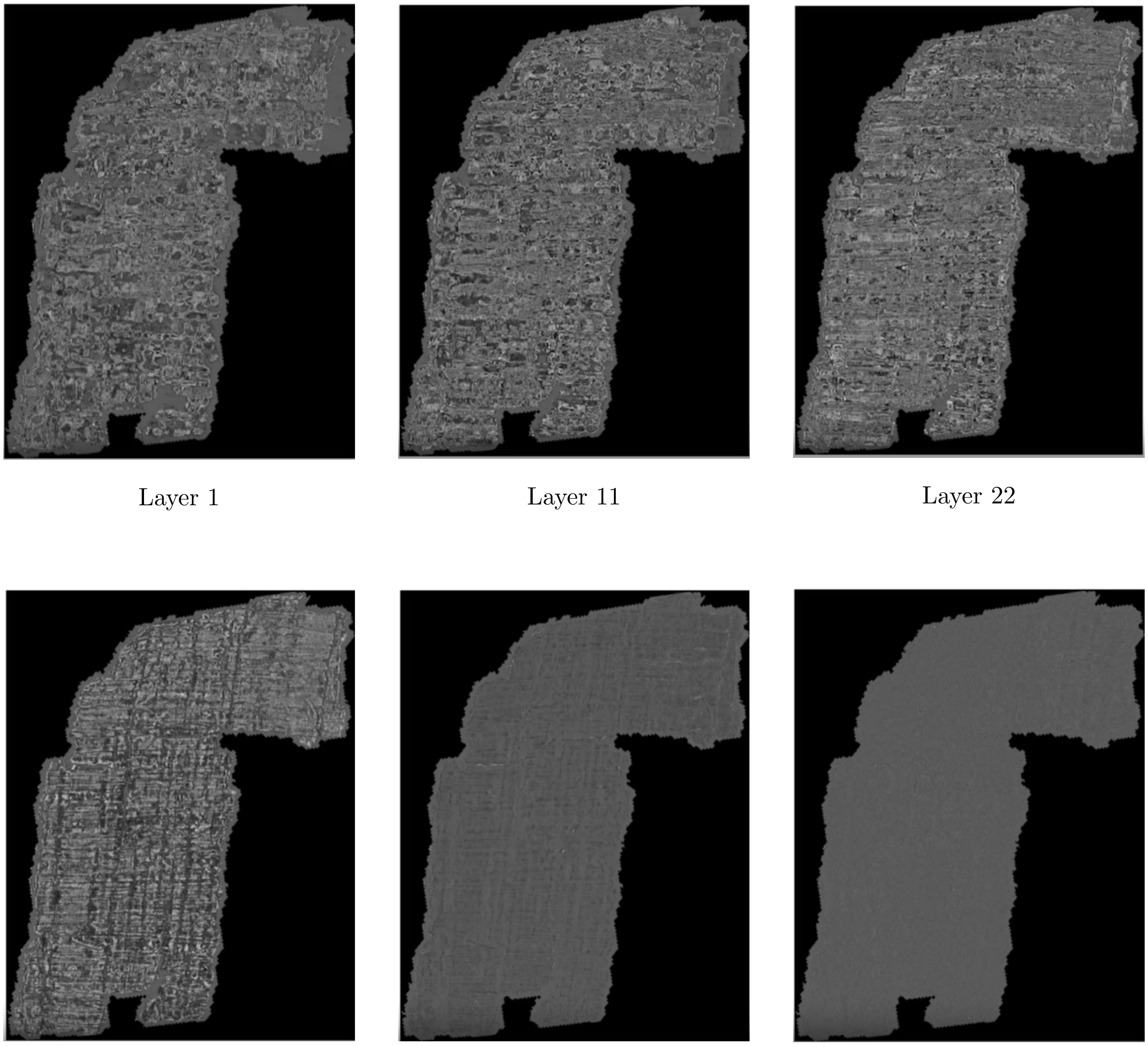
The application of computational imaging in archaeology extends beyond ink detection. Brent [3] provided a comprehensive review of methods used for analyzing the Herculaneum Papyri, highlighting the challenges associated with differentiating ink from carbonized papyrus. Parker [1] further emphasized the importance of integrating historical knowledge with computational tools to ensure the accurate interpretation of recovered texts.

Overall, the integration of deep learning, image processing, and X-ray tomography has paved the way for breakthroughs in the analysis of ancient manuscripts. The Vesuvius Challenge builds upon these prior efforts by leveraging cutting-edge machine learning algorithms to detect ink in the Herculaneum scrolls, pushing the boundaries of computational archaeology.

3 Data

The Vesuvius Challenge Ink Detection task requires the detection of ink regions in 3D X-ray scans of ancient papyrus scroll fragments. The dataset provided contains detailed surface volume images, infrared photographs, and corresponding labels that identify ink regions. These images are essential for building machine learning models aimed at segmenting ink patterns, which has significant implications for the preservation and understanding of historical texts.

The dataset was acquired through 3D X-ray imaging of the papyrus fragments, and it consists of multiple components: surface volume slices, infrared images, binary masks, and ink region labels. The surface volume slices are stored in TIFF format, with each fragment containing 65 slices. Each slice represents a cross-section of the papyrus along the z-direction, providing detailed information about the structure of the scrolls.



Layer 1 Layer 11 Layer 22

Layer 33 Layer 44 Layer 55

Figure 2: Different Layers of the X-ray scanned scroll

The surface volume images vary in size depending on the fragment, with some fragments being larger and others smaller. For instance, fragment 1 has a size of (8181, 6330) pixels, whereas fragment 2 has a size of (14830, 9506) pixels. These differences in image dimensions are important for data processing and model training, as they require resizing and normalization before use.

Sample	Shape	Number of pixels
train 1	(8181, 6330)	51785730
train 2	(14830, 9506)	140973980
train 3	(7606, 5249)	39923894
test a	(2727, 6330)	17261910
test b	(5454, 6330)	34523820

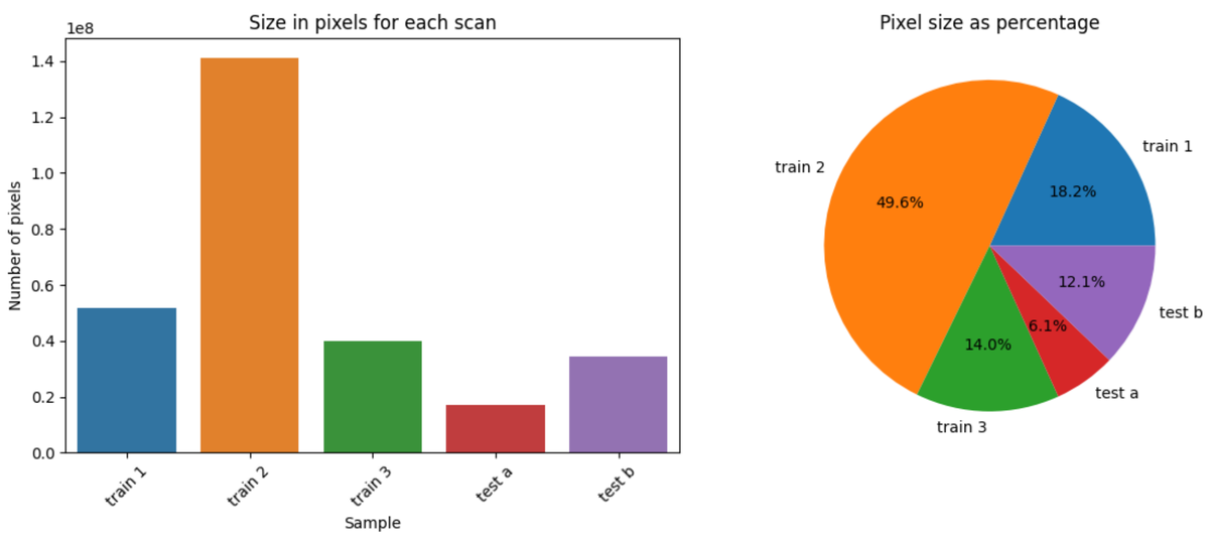


Figure 3: Sizes for each sample

Each fragment is also provided with several additional files:

- A binary mask indicating which pixels contain valid data. These masks help to eliminate regions in the image that do not contain any useful information.
- A binary ink label mask, where areas containing ink are marked as 1 and non-ink areas are marked as 0. These labels are essential for supervised learning, as they provide the ground truth for training the model.
- Infrared images, which are used to aid in identifying the ink regions based on the material properties of the papyrus fragments. These images help highlight differences between ink and the papyrus material itself.
- A run-length encoded version of the ink labels is provided, which is used for submitting predictions after model training.

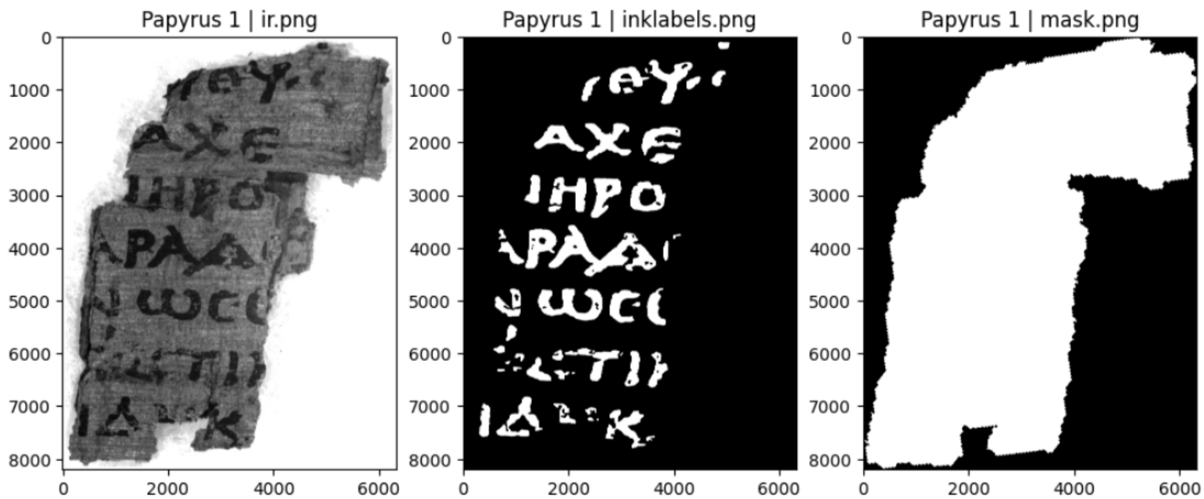


Figure 4: Sample IR, Ink Label, Mask for Papyrus 1

The test set includes surface volume slices and infrared images without the ink labels, which will be used to evaluate the model's performance. The actual test data will be hidden until the evaluation phase, ensuring the integrity of the competition by preventing overfitting to the test data.

An exploratory data analysis (EDA) was conducted to better understand the data before applying it to machine learning models. The initial analysis involved examining the shape of the images and visualizing the pixel intensity distributions. The surface volume images were found to have varying pixel intensity ranges, which are important for normalizing the images before inputting them into the model. Additionally, the binary masks were verified to ensure that the valid data regions were correctly aligned with the images.

The dataset provides rich, high-dimensional data that will be used to train machine learning models, particularly deep learning models for segmentation tasks. The primary goal is to predict the ink regions in the papyrus scrolls based on the provided images. Given the complexity of the task, convolutional neural networks (CNNs), especially segmentation architectures such as U-Net and ResNet will be employed to detect the ink regions. The surface volume images and infrared images will serve as input features, while the binary ink labels will act as the target output during training.

The Vesuvius Challenge Ink Detection dataset offers valuable data for detecting ink regions on ancient papyrus scrolls, and the data will be used in a machine learning pipeline involving data preprocessing, model training, and evaluation. The exploratory data analysis has provided insights into the dataset's characteristics, such as image sizes, pixel distributions, and mask alignments, which will guide the preprocessing steps and model selection. By utilizing this dataset, we aim to develop an effective model capable of accurately identifying ink regions in the papyrus scrolls.

4 Methodology

The methodology for the Vesuvius Challenge Ink Detection task integrates deep learning techniques with advanced image processing to detect ink regions in 3D X-ray scans of the ancient papyrus scrolls.

Data Preprocessing:
The raw dataset consists of 3D X-ray images of papyrus scrolls, where each scroll fragment is represented by a stack of 65 slices. Since the dataset can vary in resolution and size, preprocessing is necessary to standardize the data. The preprocessing pipeline begins by resizing all images to a common resolution. This ensures

that the model can handle consistent input sizes across all fragments. Next, pixel intensities are normalized to the range $[0, 1]$, allowing the model to train efficiently without bias from differing intensity scales across the images.

Given the presence of noise in X-ray scans due to scanning artifacts and the scanning process itself, denoising is applied to the images. This is performed using an iterative method that combines sparse representation and derivative-based operations to reduce high-frequency noise while preserving critical features, such as the ink regions. The denoising function works as follows:

$$g_{update} = xp.fft.fftn(g_{update})$$

The image is iteratively updated through multiple steps, applying regularization through sparse representation, edge preservation, and iterative soft-thresholding. The process removes noise and ensures that significant features, like ink, are not compromised.

Model Architecture:
For ink detection, the code utilizes a combination of deep learning architectures. The model is based on a hybrid of ResNet-3D and U-Net architectures. ResNet-3D is employed to process the 3D volumetric data from the X-ray scans, leveraging residual connections to allow deeper networks without suffering from the vanishing gradient problem. This is critical for learning spatial features in the 3D X-ray data.

The U-Net architecture is used for segmentation tasks, where the encoder path learns hierarchical features, and the decoder path reconstructs the segmentation mask. Skip connections from the encoder to the decoder ensure that high-resolution features are preserved, which is particularly important for accurate localization of ink regions.

The model is trained to predict the ink regions in each slice of the 3D volume, generating a binary mask where each pixel is classified as either part of the ink region (1) or not (0).

Training:
The training process involves feeding the preprocessed data and corresponding binary masks (representing ink) into the model. The model is optimized using binary cross-entropy loss, which is well-suited for segmentation tasks involving binary classification (ink vs. non-ink). To improve model generalization, data augmentation techniques such as random rotations, flips, and zooming are applied during training. This ensures that the model learns to recognize ink regions from multiple perspectives, increasing its robustness.

The optimizer used is Adam, which adapts the learning rate during training to ensure efficient convergence. Additionally, early stopping is employed to prevent overfitting, halting the training process when the validation loss starts to increase.

The loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}} + \lambda \cdot \mathcal{L}_{\text{Dice}}$$

where $\mathcal{L}_{\text{BCE}}$ is the binary cross-entropy loss and $\mathcal{L}_{\text{Dice}}$ is the Dice loss. The Dice loss is useful for image segmentation tasks as it helps to evaluate the overlap between predicted and actual ink regions. The combined loss function ensures that the model is penalized for both incorrect predictions and missing ink regions.

Evaluation Metric:
The primary metric used to evaluate the model's performance is the F0.5 score, which is a modified version of the Sørensen-Dice coefficient. The F0.5 score prioritizes precision over recall, making it especially suitable for tasks like ink detection, where forming coherent characters from the ink regions is crucial.
The F0.5 score is given by the formula:

$$F_\beta = \frac{(1 + \beta^2) \cdot p \cdot r}{\beta^2 \cdot p + r}, \quad \beta = 0.5$$

where:

$$p = \frac{tp}{tp + fp}, \quad r = \frac{tp}{tp + fn}$$

In this equation: - tp refers to the true positives (correctly predicted ink pixels), - fp refers to the false positives (incorrectly predicted ink pixels), - fn refers to the false negatives (ink regions that were missed by the model).

The F0.5 score is designed to emphasize precision, ensuring that the model generates accurate ink predictions, and minimizes false positives, which could lead to fragmented or incoherent characters. This makes the F0.5 score a crucial metric for tasks like ink detection in ancient manuscripts, where precision is critical for reconstructing legible texts.

Post-Processing:
After the model generates the predicted ink masks, post-processing techniques are applied to refine the results. These techniques involve morphological operations, such as dilation and erosion, to remove small noise artifacts and close gaps between ink regions. Additionally, connected components are labeled, and small isolated ink regions (less than 25 pixels) are removed to reduce noise and improve the clarity of the predictions.

To meet the competition's submission requirements, the model's output masks are encoded using run-length encoding (RLE). This encoding format reduces the file size by representing consecutive pixels of ink as start positions and lengths. For example, a RLE string "1 3 5" indicates that the ink starts at position 1, continues for 3 pixels, and then resumes at position 5. This encoding format allows for efficient storage and submission of the model's predictions while preserving essential details about ink locations.

Inference and Submission:
Once the model has been trained and evaluated, it can be used for inference on new data (i.e., previously unseen fragments). The model processes the 3D X-ray scans of the new fragments, predicting the ink regions in each slice. The predicted ink regions are then post-processed, encoded in RLE format, and submitted for evaluation.

This pipeline, involving data preprocessing, deep learning-based ink detection, and RLE encoding for submission, provides a robust solution to the problem of ink detection in the Herculaneum Papyri. By leveraging state-of-the-art deep learning techniques, this methodology contributes to the broader goal of recovering and reconstructing hidden ancient texts from carbonized scrolls.

5 Results and Conclusion

5.1 Results

We evaluated our ink-detection model on 3D X-ray scans of papyrus fragments through a three-stage pipeline: preprocessing, prediction, and post-processing. Figures 5 and 6 overlay the model's predicted ink regions (yellow) on the raw scan volumes (purple). In both examples, the network reliably highlights large, high-contrast ink areas, most notably in the upper-left quadrant of Fragment 1 and along the left edge of Fragment 2, while also capturing smaller, isolated spots toward the fragment centers and edges.

Visually, these overlays demonstrate that the model excels at segmenting prominent ink traces but is less consistent when confronting faint or highly fragmented markings. Low-contrast regions often result in missed detections or lower-confidence predictions, indicating an area for further refinement.

Quantitatively (Table 1), we compare three configurations of our approach using the $F_{0.5}$ score, which weights precision more heavily than recall:

Methodology	**$F_{0.5}$ Score**
3D ResNet + Machine Learning + Morphological Transformations (v2)	**0.64**
3D ResNet + Machine Learning + Morphological Transformations	0.63
3D ResNet Baseline (Inference Only)	0.62

Table 1: Comparison of ink-detection performance across different model configurations.

The best performing pipeline, our version 2 ensemble of 3D ResNet features, learned classifiers, and morphology achieves an $F_{0.5}$ score of 0.64, a modest improvement over the original morphological post-processing design (0.63) and the pure ResNet baseline (0.62). These results confirm that learned post-processing steps contribute meaningfully to precision, even as recall of subtle ink remains challenging. Overall, our method provides a valuable step toward automated recovery of hidden papyrus texts, with further gains expected from targeted enhancements to detect low-contrast and highly fragmented ink patterns.

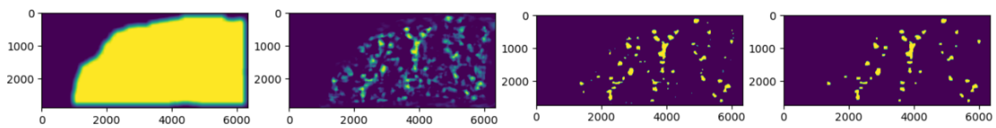


Figure 5: Ink detection on Papyrus Fragment 1. Yellow indicates predicted ink regions; purple indicates non-ink.

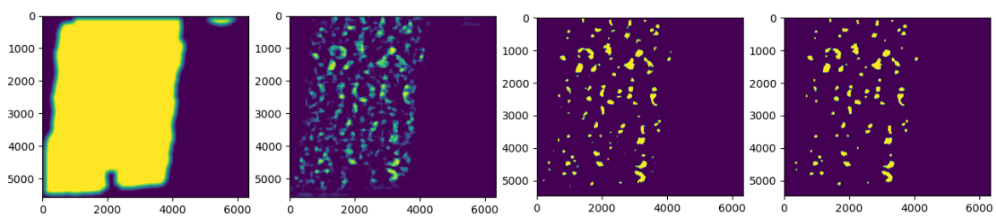


Figure 6: Ink detection on Papyrus Fragment 2. Yellow indicates predicted ink regions; purple indicates non-ink.

5.2 Conclusion

The deep learning-based ink detection model has shown promising results in identifying and segmenting ink regions from the 3D X-ray scans of the Herculaneum papyri. The model is particularly effective at identifying large, distinct ink areas, as demonstrated by the clear and accurate predictions in the provided images. However, the results also indicate that the model struggles with detecting smaller or less prominent ink regions, as evidenced by the scattered and sometimes fragmented yellow areas. This suggests that the model could benefit from further optimization to improve its ability to detect faint or subtle ink traces.
Several potential improvements could be made to enhance the model's performance:

- **Model Refinement:** The current model can be enhanced by using a more sophisticated architecture, such as a 3D U-Net with deeper layers and more skip connections, which could improve the localization and detection of smaller ink regions.
- **Data Augmentation:** Additional data augmentation techniques could be applied, particularly focusing on augmenting the training set with more examples of faint ink traces or low-contrast ink regions. This could help the model generalize better and improve its ability to detect less distinct ink regions.

- **Incorporating Infrared Data:** By integrating infrared images, which provide additional contrast between the ink and papyrus, the model may improve its accuracy in identifying faint ink regions that are not as visible in the X-ray scans alone.
- **Fine-tuning and Hyperparameter Optimization:** Further tuning of hyperparameters, such as the learning rate, number of epochs, and batch size, could help the model achieve better convergence and improve its performance on challenging fragments.
- **Post-Processing Improvements:** The post-processing techniques, such as morphological operations, could be refined to better preserve the continuity of ink regions and reduce false positives.

By incorporating these improvements, it is expected that the model will perform better on subtle and fragmented ink traces, leading to more accurate and complete reconstructions of the hidden texts in the Herculaneum scrolls. The ability to detect and segment these ancient writings non-invasively holds great potential for historical research and could be applied to other similar archaeological tasks.

References

[1] H. Parker. *Herculaneum: Past and Future*. Oxford University Press, 2019.

[2] W. B. Seales and C. Parker. “Revealing the Invisible: Virtual Unwrapping of the Herculaneum Scrolls”. In: *Nature Communications* 10.1 (2019), pp. 1–8.

[3] W. Brent. “The Herculaneum Scrolls: A Scientific and Historical Perspective”. In: *Journal of Archaeological Science* 72 (2016), pp. 20–30.

[4] W. B. Seales. “Virtual Unrolling of Ancient Manuscripts”. In: *Science Advances* 2.9 (2016), e1601247.

[5] T. Yun and H. Kim. “Deep Learning for Ink Detection in Ancient Manuscripts”. In: *IEEE Transactions on Image Processing* 32 (2023), pp. 2356–2367.

[6] T. Shimada and R. Sato. “Machine Learning Approaches for Historical Document Analysis”. In: *Digital Humanities Quarterly* 15.4 (2021).