Assignment 2: CNNs

Timo Brady
Trimester 3, 2023
Deep Learning Fundamentals

Abstract

This paper presents a methodology in analyzing papyrus scroll fragments from Herculaneum, accomplished through the use of three-dimensional convolutional neural networks (CNNs). Central to this methodology is a unique approach combining clustering-based ink detection methods with semi-supervised learning. Utilizing the 3D U-Net architecture for initial feature extraction, I have enhanced its capabilities through a semi-supervised clustering paradigm, effectively surpassing the constraints of traditional two-dimensional and basic three-dimensional techniques. The approach experiments with a more refined ink detection process, to hopefully advance from simple binary classification to a nuanced method of identifying ink traces in the future.

1. Introduction

Imagine having access to lost ancient texts that predate 79 AD, with content unknown to the modern world. The challenge? The texts are almost impossible to read. In deciphering ancient Herculaneum scrolls, not only do we contend with interference from noise, cracks, charred regions, and age-induced damages, but also the added complexity of the ink being radiolucent. Traditional methods are not equipped to handle these intricacies, especially when the ink itself is elusive to certain imaging techniques.

While ink detection serves as the starting point and the current state of affairs with attempting to read the scrolls, simply discerning between ink and non-ink is not sufficient. Papyrus is marred by noise that models can easily mistaken for ink, making the task of accurate reading a complex challenge. Moving past traditional techniques, I propose a preliminary clustering methodology to identify genuine ink traces amidst overwhelming noise, enabling a more refined and accurate interpretation in subsequent character recognition and script reading.

Traditional deciphering is not challenged by continuous text, but by discerning actual ink from noise. Clustering is pivotal in this endeavor, accurately segmenting genuine ink characters from the predominant noise. Clustering differentiates, helping identify regions that are disproportionately noisy, refining the ink detection process, and foregrounding genuine ink traces. Noise misinterpreted as ink leads machine learning models astray. By clustering and filtering out the noise, we're training the models on true ink traces in their natural context. This sharpens their capability to accurately interpret the characters in subsequent experiments.

The implementation of 3D imaging in the analysis and interpretation of ink on ancient papyrus fragments has proved invaluable. This methodology overcomes the high rate of false positives, a common limitation of binary ink classification in conventional approaches. The strategy of this paper redefines the ink detection challenge, integrating a specialized 3D U-Net architecture for feature extraction and enhancing it with a clustering algorithm. This dual approach, it is hoped, will, in the future, improve the detection of authentic ink traces, thus increasing the accuracy and precision in ink detection and facilitating the eventual reading of the scrolls *in toto*.

2. Related Work

The foundations of the method build upon the pioneering work of Parker *et al*. [2] and Chesler *et al*. [1] in applying machine learning to ink detection on papyrus. I have synthesized and improved upon these methodologies, addressing the specific limitations they presented. By incorporating refined feature extraction techniques and employing advanced clustering algorithms, my approach will eventually help to distinguish between ink and papyrus fibers more effectively and recognize complex patterns of noise and ink with greater accuracy. The experiments set out in the code base (see below section) are a step forward in ink detection with machine learning models.

3. Exploratory Data Analysis

The dataset consists of 4μm 3D computerized-tomography (CT) scans covering four papyrus fragments. With 65 2D slices per fragment, there are 260 slices. For the 2D image analysis, individual slices, associated labels,

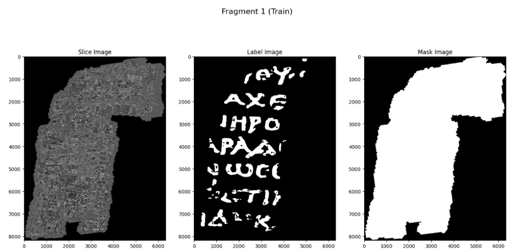

Figure 1. Sample of fragment slice, label, and mask for Fragment 1

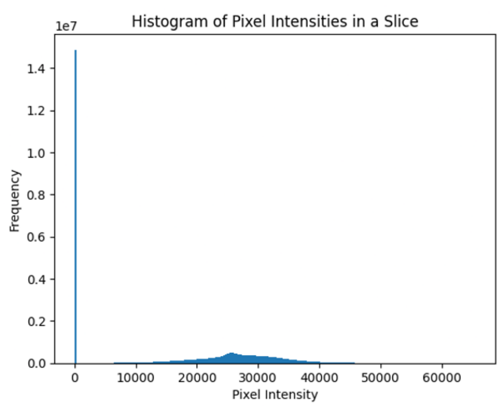

Figure 2. Histogram of Pixel Intensity in a Slice of Fragment 1

and masks from each fragment have been subjected to computational visualization (Figure 1). The initial histogram analysis revealed approximately 10 million zero-intensity pixels, indicating potential missing or erroneous data (Figure 2). A prototypical slice exhibited a mean intensity value of 16,941 with a standard deviation of 14,196.

These parameters underline the inherent data heterogeneity and indicate the need for an appropriately complex model. These zero-intensity pixels warrant further attention, especially in 3D contexts. Preliminary edge detection suggests that regions with these pixels often coincide with cracks or missing portions of papyrus, potentially adding contextual information for the model. Edge-detection is critical to isolating text from non-text elements. I have overlaid an ink label for context. When viewed in a 3D compilation, these pixels, the deep purple points, appear to provide additional textural complexity (Figures 3, 4, and 5).

A defining feature of ancient writing on papyrus, which, in our case is in ancient Greek, is that the characters by and large tend to be arranged in parallel with the papyrus fibers. I have overlaid the label for context and zoomed in and added a contrast filter to see a close up of how the strips of papyrus are laid out as well as an artifact in the upper half of the image that a model could mistake for ink (Figures 6 and 7).

Stacking multiple slices for 2D visualization reveals a decline in data interpretability, especially the intra-slice context, accentuating the limitations of 2D methodologies and the consequent necessity for higher-dimensional analyses. In 2D, when you stack multiple slices, any variations in texture or intensity are averaged out or overshadowed by the most dominant features, turning it into what appears to be a gray mass. In contrast, a 3D approach would consider the correlation in the Z-axis (that is, through the slices), which could help in identifying nuanced features that are lost in a 2D approach (Figure 8).

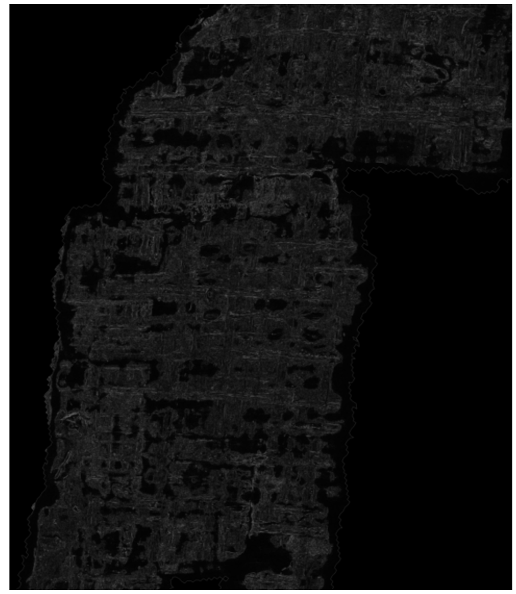

Figure 3. Edge Detection with High Contrast Filter

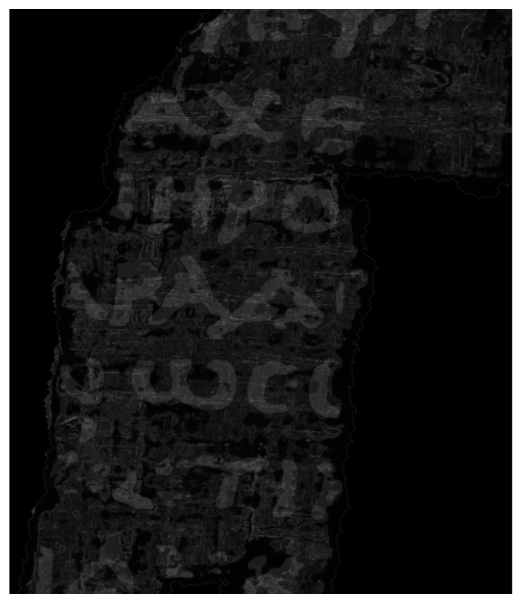

Figure 4. Edge Detection with Ink Label

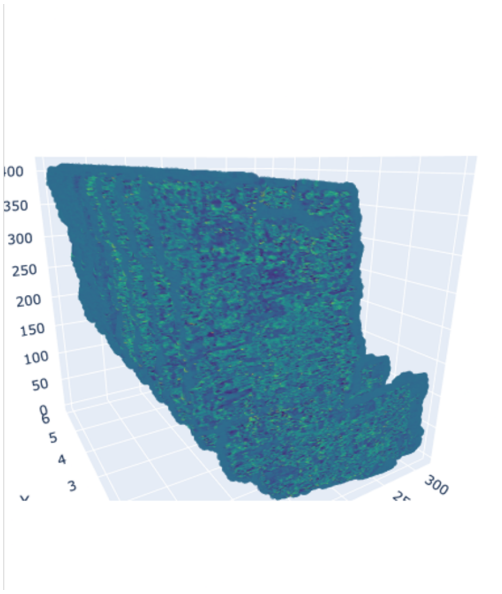

Figure 5. Rendering of a 3D Volume of the 2D Slices

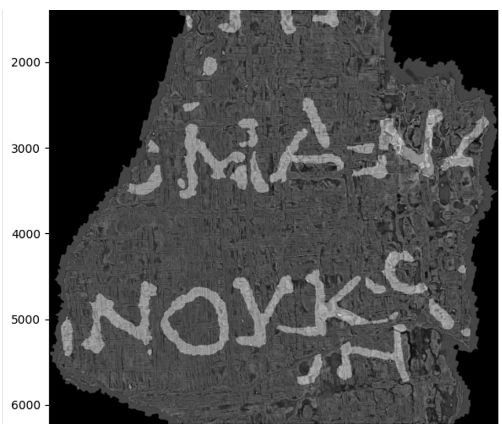

Figure 6. Zoomed in with Ink Label Overlay

Distinct peaks in intensity variation across depth have been identified between slices 25-30 in the fragments. This could indicate that the most relevant information lies in those slices. The outer layers of the scrolls were obviously more exposed to environmental factors, causing them to fade or degrade more than the central layers. This is a conjecture (Figures 9, 10, and 11).

A peak in contrast followed by a drop indicates that the texture within these fragments changes significantly at smaller angular distances. The high contrast, especially for fragment 3, could suggest that there are pronounced edges, corners, or other features that result in this contrast peak and quite possibly the presence of ink when compared to blank papyrus as in the image. High contrast regions warrant closer inspection (Figures 12, 13, 14, and 15).

I prepared separate 3D volumes of each fragment from

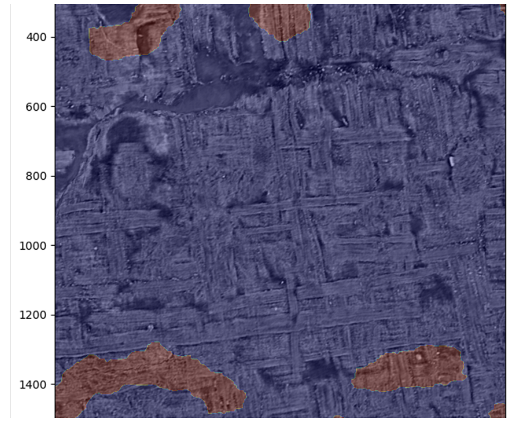

Figure 7. Zoomed in on Lower Right Quadrant

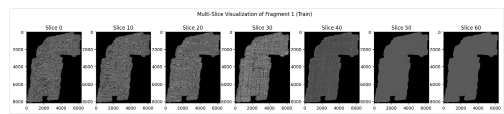

Figure 8. Stacking of Multiple Slices

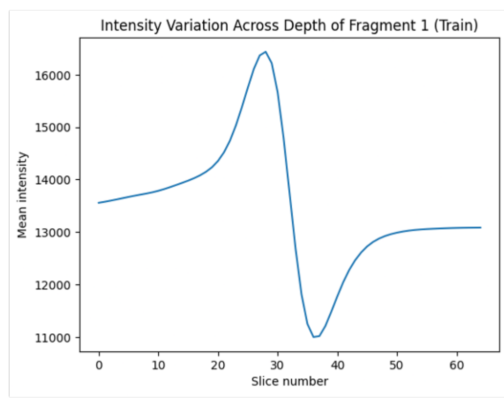

Figure 9. Intensity Variation Across Depth of Fragment 1

the 2D slices. Because the data type of the 2D slices is a big-endian unsigned 16-bit integer, it was necessary to convert them into float32; each slice was downsampled by a factor of 2 in order to effectively manage computational resources. Padding was added to fit a target shape of so that fragments

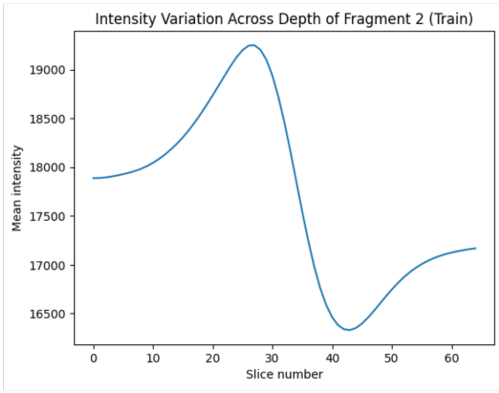

Figure 10. Intensity Variation Across Depth of Fragment 2

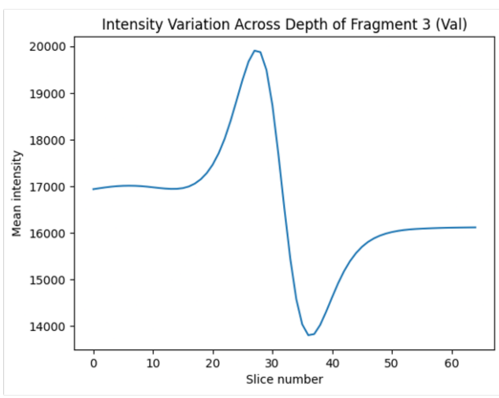

Figure 11. Intensity Variation Across Depth of Fragment 3

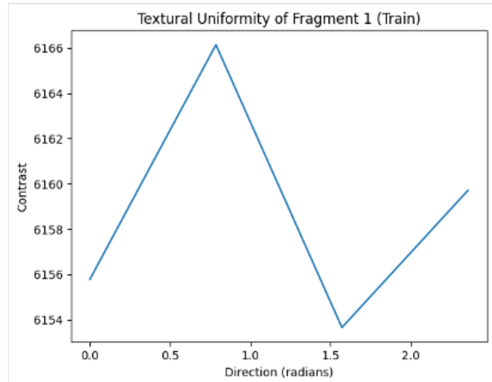

Figure 12. Textural Uniformity of Fragment 1

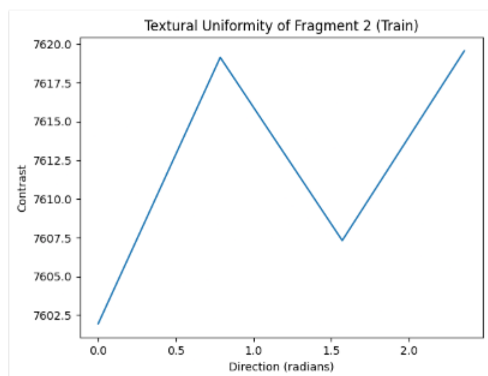

Figure 13. Textural Uniformity of Fragment 2

of varying dimensions could be combined in the future for model training.

4. Methodology: Feature Extraction

4.1. Data Pre-Processing for 3D U-Net

The dataset, derived from specialized 3D CT scans of the Herculaneum scrolls, was meticulously processed. By converting 2D slices into 3D volumes, I have preserved the intricate intra-slice context and spatial hierarchies, crucial for detailed feature extraction. The conversion of the data to float32 and strategic downsampling enabled efficient computational performance without compromising on the quality of feature extraction.

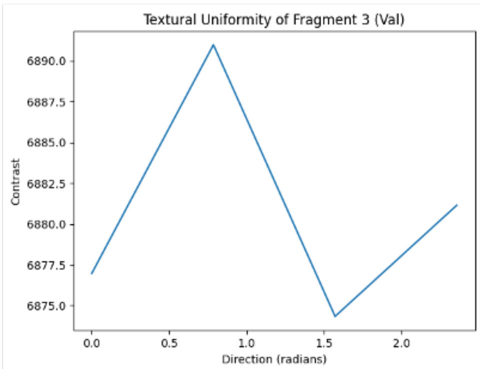

Figure 14. Textural Uniformity of Fragment 3

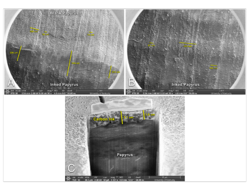

Figure 15. Detail Showing Blank vs Inked Papyrus (Parker, *et al.* [2]

Feature Shape	(30, 1, 16, 2, 926, 594)
Mean	0.259
Standard Deviation	0.490
Silhouette Score	0.813

Table 1. Preliminary Feature Extraction Results

4.2. Preliminary Feature Extraction

The U-Net architecture, inspired by Chesler *et al.* [1], was tailored for optimal feature extraction. The model, balancing computational efficiency and the capacity to detect complex features, has shown promising results. The preliminary extraction outcomes, indicated by the feature shape, mean, standard deviation, and silhouette score, confirm the effectiveness of the approach in setting the stage for more advanced ink detection in the future.

4.3. Modified 3D CNN Architecture

The implementation of the U-Net model introduces key modifications that have significantly enhanced its performance. The architecture, augmented with a series of convolutional blocks and pooling layers, has been tailored to more accurately capture and differentiate between ink and noise.

1. First Convolutional Block: Featuring 8 filters, I observed a marked improvement in handling the vanishing gradient problem with the ReLU activation function. Experimentation with leaky ReLU also yielded promising results, offering better performance in certain scenarios.

2. Second Convolutional Block: Comprising 16 filters, these were fine-tuned using advanced techniques such as Bayesian Optimization, resulting in a more nuanced understanding and detection of ink features.

3. Max Pooling Layers: The choice to retain max pooling proved effective in reducing computational complexity while capturing essential features. Exploratory trials with alternative approaches like average pooling also provided insights into their potential in retaining more diverse information types.

4. Feature Transformation Block: The implementation of a feature transformation block effectively captured at least 32 essential features, leading to a more detailed representation of noise and ink, within the computational constraints.

5. Output Layer: With dimensions set to (batchsize, 32, 1024, 1024), the output layer's granularity greatly enhanced the model's capability in clustering noise and ink features, thus fine-tuning the semi-supervised model for better performance.

By meticulously selecting and fine-tuning the filters in the convolutional layers, I have significantly improved the model's sensitivity to ink and noise-specific features. The additional feature transformation block has further elevated this capability, making it an indispensable component of the model.

4.4. 3D U-Net Evaluation Metrics

The performance of the 3D models for feature extraction was comprehensively evaluated using the silhouette score. The silhouette score (Equation 1) particularly provided valuable insights into the effectiveness of the clustering approach.

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \tag{1}$$

The model exhibited a consistently high silhouette score, affirming the robustness of the feature extraction process and the effectiveness of the fine-tuning strategies.

Augmentation Technique	Amount Applied
horizontal and vertical flips	50%
90-degree rotations	75%
brightness contrast	50%
1-2 channel (depth) dropout	25%
shift scale rotate	10%
noise and blur	10%
coarse dropout	10%
grid distortion	10%

Table 2. Successful Data Augmentation Techniques from Chesler *et al*. [1]

5. Methodology: Current Considerations

5.1. Clustering

The feature vectors extracted from the 3D U-Net served as a precise input for the clustering phase. The direct pipeline from the feature transformation block to the clustering algorithm ensured seamless integration and maximized the effectiveness of the high-dimensional feature vectors.

5.2. Augmentation Techniques for Clustering

Recognizing the limited nature of the dataset (after all, there are only 4 fragments), I implemented data augmentation strategies. Modifications to Chesler *et al*.'s initial techniques were pivotal in achieving enhanced discrimination between ink and other elements such as noise. Explorations with rotations, flips, and a combination of other augmentation techniques significantly bolstered the model's resilience and accuracy.

5.3. Clustering Algorithm Selection and Evaluation

The investigation into various clustering algorithms, including K-means, DBSCAN, and Hierarchical Clustering, was guided by considerations of intra-cluster and inter-cluster distances along with computational efficiency. The silhouette score and explained variance (Figure 16) emerged as key metrics in determining the most effective clustering model. The final selection significantly outperformed others in the validation dataset.

5.4. Semi-Supervised Learning and Evaluation

In the final phase of the methodology, the clusters formed were projected back onto a 2D plane. This approach allowed for leveraging t-distributed Stochastic Neighbor Embedding (t-SNE) for generating pseudo-labels, enhancing the accuracy of the binary ink detection. The concentration of clusters towards the origin suggests that most of the features reside within a small numerical range (Figure

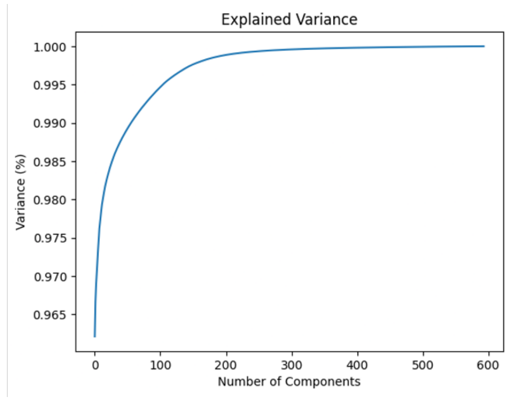

Figure 16. Explained Variance Plot

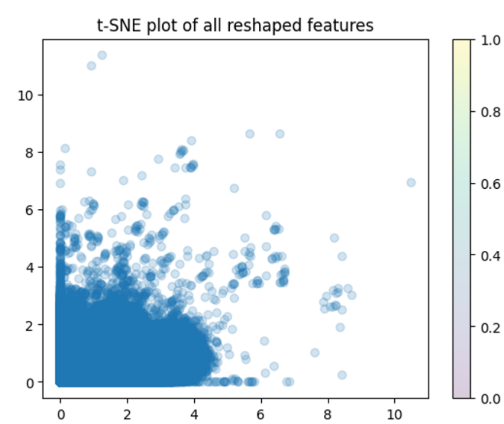

Figure 17. t-SNE Plot of All Reshaped Features

17). I recommend experimenting more with feature extraction, including with fragment 2, and potentially reverse engineering the model to glean the patterns and features that are detected. However, it is indeed clear that textural variation, pixel intensity, and region of high contrast are critical features.

6. Future Considerations

In order to move the methodology forward in the future, I recommend that the 3D U-Net architecture be reconfigured for backpropagation, featuring a secondary output layer for binary ink detection. This adaptation, coupled with the utilization of both original ink labels and pseudo-labels, may lead to substantial improvements in precision, recall, and F1-scores. A tailored False Positive Rate (FPR) measure, relying on the probability of cluster membership as a form of binary thresholding, may offer a more nuanced and robust assessment of model performance.

We often stand on the peripheries of history, sensing its vastness but rarely its depth. The Herculaneum scrolls are a testament to our past, relics awaiting revelation.

7. Code

The supporting code for this paper can be accessed at the following private respository: https://github.com/bluebindu/DeepLearningAssignments/blob/main/DeepLearningFundamentalsAssignment2.ipynb. Access has been granted to the following individuals: jinan.zou@adelaide.edu.au, haiyao.cao@adelaide.edu.au, and jiaxin.wang01@student.adelaide.edu.au.

References

[1] Chesler et al. https://www.kaggle.com/competitions/vesuvius-challenge-ink-detection/discussion/417496. *Vesuvius Challenge - Ink Detection: 1st place solution*, 2023. 1, 5, 6

[2] Parker et al. From invisibility to readability: Recovering the ink of herculaneum. *PLoS One*, 14(5), 2019. 1, 5