Deciphering Ancient Papyrus Texts: A Machine Learning Approach with Varied Architectures and Encoders

1 author:

Pinxue Lin
Yeshiva University

Deciphering Ancient Papyrus Texts: A Machine Learning Approach with Varied Architectures and Encoders

Pinxue Lin
Yeshiva University
Katz School of Science and Health
plin3@mail.yu.edu

Abstract

This study explores the application of machine learning models for detecting ink on carbonized papyrus scrolls excavated from the Roman villa library in Herculaneum, providing invaluable insights into ancient texts preserved in their original form for almost two millennia. These fragile scrolls, buried and carbonized during the catastrophic eruption of Mount Vesuvius, have remained unreadable due to their delicate state. However, with the advancements in technology, particularly 3D X-ray scanning, a novel opportunity arises to unveil their contents without physical intervention, ensuring the preservation of these historically significant artifacts. The detection of ink on the scrolls plays a pivotal role in this research, encompassing a comprehensive three-step process. First, high-resolution 3D scans are generated using X-ray tomography, enabling the creation of detailed volumetric representations of the scrolls. Second, through sophisticated segmentation and flattening techniques, the layered structure of the rolled papyrus is meticulously identified and transformed into a flattened "surface volume" presenting a unique visual context for subsequent analysis. Finally, state-of-the-art machine learning models, including Unet, Deeplabv3, and Manet, are employed to identify and delineate the inked regions within the flattened surface volume, unraveling the hidden textual information concealed within these ancient artifacts. By experimenting with various encoder layers and optimizing model architectures, the research aims to achieve accurate and efficient ink detection. Preliminary results show promising progress, with the highest achieved dice score on the validation set reaching 48.44, This project demonstrates the ability to use machine learning to detect ink in 3D scans, and ultimately to decipher the contents of carbonized parchment.

1. Introduction

Understanding the past relies on our ability to access and understand historical artifacts. One valuable source of information is ancient manuscripts, which provide important knowledge about past cultures and civilizations. However, it is difficult to preserve and read these manuscripts, especially the carbonized papyrus scrolls from the Roman villa library in Herculaneum.

These scrolls, buried and carbonized during the catastrophic eruption of Mount Vesuvius nearly two millennia ago, have remained inaccessible due to their delicate state. Attempts to physically unroll these scrolls have often resulted in damage, leading to loss of potential knowledge. This predicament underscores the need for non-invasive methods that allow us to access the texts contained within these carbonized scrolls without compromising their physical integrity.

Advancements in technology have opened new avenues for examining these scrolls. One such development is 3D X-ray scanning, a technique that has the potential to unlock the contents of these scrolls without necessitating their physical unrolling. X-ray tomography can generate detailed volumetric representations of the scrolls, providing an unprecedented level of detail that paves the way for further analysis.

This research builds upon these technological advancements, focusing on the critical task of ink detection on the scrolls. This process forms a key component of a three-step methodology: 3D scanning, segmentation and flattening, and ink detection. Using sophisticated techniques, the layered structure of the rolled papyrus is identified and transformed into a flattened "surface volume" suitable for analysis. The last step, and the central focus of this study, involves the application of state-of-the-art machine learning models to detect and delineate the inked regions within the surface volume.

Machine learning has shown significant promise in a variety of complex tasks, including image recognition[12], segmentation[16], and classification[10]. In this project, we leverage this potential to optimize the process of ink detection on carbonized papyrus scrolls. We experiment with multiple machine learning models, including Unet[13], Deeplabv3[4], and Manet[8], integrating various encoder layers to enhance the performance and accuracy of our approach.

With an achieved dice score of 48.44 on the validation set, our initial results suggest promising progress towards our ultimate goal of reading these ancient texts. This research not only seeks to advance the field of digital paleography but also aims to uncover the hidden historical knowledge preserved within these precious artifacts, thereby contributing to our understanding of our shared past.

The EduceLab-Scrolls dataset[11] we used is a groundbreaking resource for scholars and researchers interested in the study of ancient texts. This comprehensive open dataset represents two decades of research effort on the problem of revealing the hidden texts of the Herculaneum papyri using X-ray CT images. The dataset contains a set of volumetric X-ray CT images of both small fragments and intact, rolled scrolls, along with aligned 2D labels. The EduceLab-Scrolls dataset is designed to be used in conjunction with this pipeline and represents a significant contribution to the field of digital humanities.

U-Net[13]: U-Net is a convolutional neural network architecture widely used in medical image segmentation tasks. It consists of an encoder-decoder structure that captures both local and global information. The U-shaped architecture allows for efficient feature extraction and precise segmentation. U-Net has been successfully applied in various medical imaging applications, such as tumor segmentation and cell detection, due to its ability to handle limited training data and produce accurate segmentation results.

DeepLabV3[4]: DeepLab is a deep learning-based approach for semantic image segmentation. It utilizes a modified convolutional neural network, such as ResNet, as the backbone network for feature extraction. DeepLab incorporates atrous convolution, which enables dense feature extraction at multiple scales without increasing the network's computational cost. By combining global context and fine-grained details, DeepLab achieves state-of-the-art performance in tasks like object segmentation, scene parsing, and human pose estimation.

MANet[8]: MANet is a proposed deep convolutional neural network architecture for semantic segmentation of remote sensing images. It utilizes a dot-product attention mechanism to refine multi-scale features and explore long-range dependencies of feature maps, resulting in improved feature representations associated with each semantic class. Numerical experiments on three large-scale fine resolution remote sensing images captured by different satellite sensors demonstrate the superior performance of MANet compared to other benchmark approaches such as DeepLab V3+, PSPNet, FastFCN, DANet, and OCRNet.

About the encoder we use, ResNet50[5] is a deep residual network with 50 layers, known for its skip connections that address vanishing gradients. It excels in image classification and object detection tasks. MobileNet[6], designed for resource-constrained environments, employs depthwise separable convolutions to reduce computational cost while maintaining good accuracy. It is ideal for real-time applications on devices with limited power. EfficientNet[17] is a scalable and efficient architecture achieved through compound scaling, balancing model size and accuracy. It outperforms other architectures in image classification. These networks - ResNet50, MobileNet, and EfficientNet - are widely used in computer vision, each offering unique benefits for various applications.

2. Related Work

Virtual unwrapping[15, 14], an innovative approach merging heritage science and computational methods, employs non-invasive volumetric imaging and a multi-step computational pipeline to delineate 3D surfaces, reveal hidden text, and translate the results to 2D images. This technique has proven adaptable to various challenges and materials, as demonstrated by its successful application to lab-made manuscripts. These test runs have showcased the method's versatility across diverse imaging methods, substrates, inks, and manuscript forms such as scrolls and folded sheets. From initial trials on proxy materials with exaggerated features, the field has evolved to handle intricate materials like bamboo scrolls, metallic inks on parchment, and rolled and folded papyri. These advancements have paved the way for successful text recovery from authentic heritage manuscripts, including those made from metallic inks on parchment, paper, papyri, etched metal scrolls, and lead amulets.

Noninvasive study of Herculaneum papyri[2] has primarily used spectral imaging in infrared bands[3], enabling the visual contrast of exposed writings, but with limited penetration into the hidden layers of a rolled scroll. The research has been centered on understanding the chemical composition of the ink used in these papyri, to inform future imaging methods that could capture clear ink contrast from an intact scroll. Findings indicate varying levels of elements such as calcium, lead, and strontium in the ink, but not in the papyrus. Despite the presence of lead in the ink being not fully understood, X-ray fluorescence has yielded clear imaging contrast. The varying composition of inks can be attributed to the fact that the Herculaneum papyri were authored by different scribes, using homemade inks, over three centuries. X-ray CT imaging has revealed the internal structure of the scrolls but not the ink contrast. Machine learning-based methods have been used for ink detection.

Volumetric datasets[9, 7], while common in medical imaging, often suffer from lower resolution due to constraints such as radiation dosage and patient motion. However, methods developed for these medical images, particularly surface segmentation, can be applied to the virtual unwrapping of Herculaneum papyri.

The study by Parsons et al. [11] presents a significant contribution to the field of digital humanities by providing a complete software pipeline for revealing the hidden texts of the Herculaneum papyri using X-ray CT images. The authors combine machine learning with a novel geometric framework linking 3D and 2D images to detect carbon ink inside Herculaneum scrolls. They also present EduceLab-Scrolls, a comprehensive open dataset representing two decades of research effort on this problem, which contains volumetric X-ray CT images of both small fragments and intact, rolled scrolls, along with aligned 2D labels. The authors validate their method using multiple metrics, including an assessment by Herculaneum scholars which confirms the accuracy of their generated images. This work builds on previous efforts in this field, including the development of virtual unwrapping algorithms for reading unopened objects and the use of X-ray CT to reveal text in other ancient scrolls. Overall, Parsons et al.'s study represents an important step towards reading the intact Herculaneum scrolls using current imaging technology and computational methods.

3. Methods

3.1. Data set

Our comprehensive training data set is comprised of three distinct components, each playing an important role for the data set.

The first component is the surface volumes. These volumes are generated through the innovative technique of 3D X-ray scanning of papyrus. Each surface volume is represented by a collection of 65 grayscale images, meticulously captured and stored as uint16 data type. These images provide an invaluable insight into the intricate details and structural composition of the papyrus, enabling us to unravel its hidden secrets and unravel its historical significance.

The second component of our data set is the ink labels. Recognizing the importance of precise and reliable annotation, it has been manually labeled and marked the exact locations of ink . These ink labels serve as a key reference point for our data set, facilitating the identification and analysis of various writing patterns, textual content embedded within the papyrus.

Lastly, we have the mark images. As the background of our images is non-transparent and predominantly black, the mark images act as a visual guide, highlighting and demarcating the specific regions that have been subjected to the scanning process.

Figure 1. train set 1's surface volumes

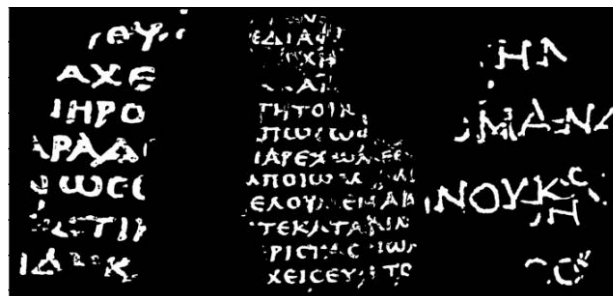

Figure 2. Spliced ink labels for three datasets

Figure 3. Spliced masks for three datasets

Together, these three components synergistically form our training data set, providing a comprehensive and multifaceted resource for advancing the field of papyrus analysis.

3.2. EDA

Because AJLAND draws charts about each data set's mean, median, standard deviation and median absolute deviation information[1], use these charts can help us to analyze our data set.

The figure shows two primary indicators: 'ink label' representing the data within the ink labels mask area, and 'masked region' representing data within the mask area.

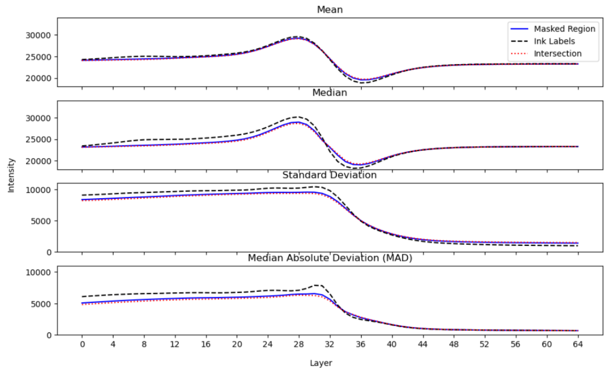

Figure 4. train set 1's surface volumes

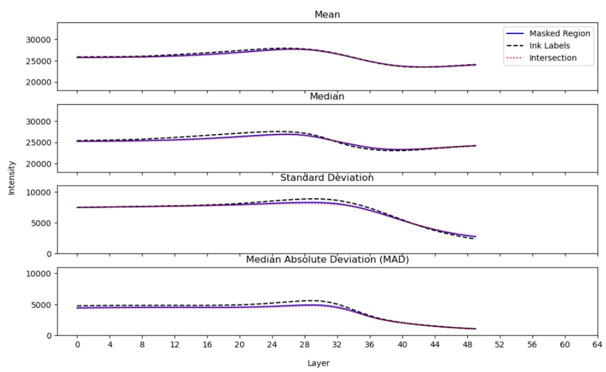

Figure 5. train set 2's surface volumes

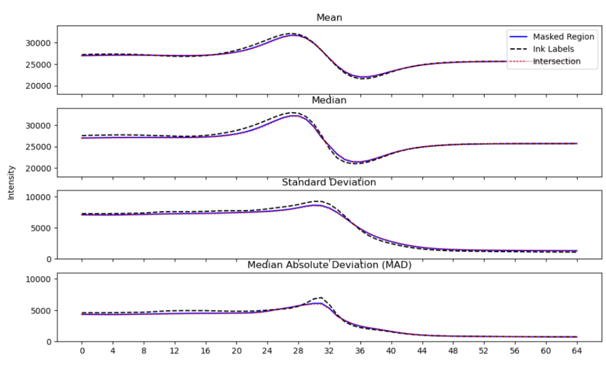

Figure 6. train set 3's surface volumes

The horizontal axis below represents the surface volume index. Observing these images, it's noticeable that in some volumes, the ink labels and masked region do not overlap, indicating a discrepancy. This suggests that there might be additional information within the ink labels areas of these volumes, which could aid in ink detection.

Observing the graph of 'train set 1', the standard deviation initially shows a consistent gap, but a significant change occurs at the 32nd layer, indicating a substantial shift in information in the middle of the data. This change is precisely what the model needs as it can help the model better predict the ink region.

The changes in 'volumes 2' are smoother, showing a very different data trend compared to 'volumes 1 and 3'. This suggests that the information characteristics of 'volumes 2' are distinct from 'volumes 1 and 3', and using it as a training set can allow the model to learn more ink characteristics. Similarly, the difference between the ink labels and masked region shows a substantial change at the middle layer.

'Volumes 3' and 'volumes 1' are very similar, possibly originating from the same material, with the significant changes in information occurring in the middle layer.

Based on this analysis, when training the model, selecting layers close to the middle of the 65 layers can enhance the model's performance. For this study, I chose the middle 12 layers, layers 26 to 37, for training.

3.3. Data Pre-processing

Due to the high resolution of our images, such as volume 1 with dimensions of 6130x8181, it is essential to manage the scale of data for this project. We divided these large images into smaller segments with a resolution of 224x224, each serving as individual samples in our dataset. This approach ensures that the model can process the dataset efficiently, while still retaining critical details for ink detection on the papyrus scrolls.

Figure 7. Image after splitting, in order to get more patches, the stride are set to 112

To augment our data set and improve the model's robustness and generalization ability, we utilized the Albumentations package, a powerful library for image augmentation. The chosen augmentation strategies include random flipping and cropping, brightness variation, Grid Distortion, random blurring, and Shift-Scale-Rotate transformations.

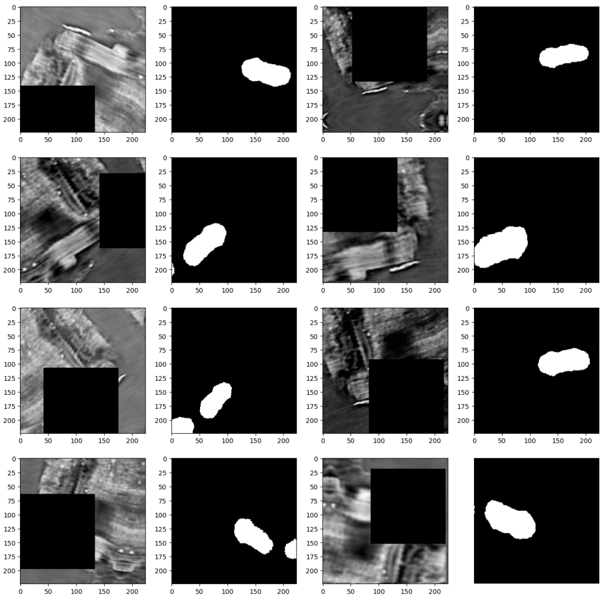

Figure 8. sample image augmentation for same image

Random flipping and cropping are fundamental techniques that help increase the diversity of the data, providing the model with different perspectives and scales of the images. Brightness variation emulates different lighting conditions, making the model more resistant to variations in image illumination. Grid Distortion can simulate non-linear deformations of the image, preparing the model for possible irregularities in the scrolls. Random blurring imitates out-of-focus situations, whereas Shift-Scale-Rotate transformations provide the model with more varied positional data.

By employing these augmentations, the model can learn more generalized features from the images, thereby enhancing its performance and robustness, leading to more accurate and reliable ink detection on the carbonized papyrus scrolls.

3.4. model architecture

In this project, I opted to experiment with five distinct models to evaluate their performance in the task of ink detection on the carbonized papyrus scrolls. These models can be classified into three groups based on their architecture and encoder structure:

Unet Models with ResNet50, EfficientNet-b0, and MobileNet: Unet is a powerful architecture that excels in image segmentation tasks. I utilized three different encoders for the Unet model to compare their performance and efficiency. These encoders include ResNet50, EfficientNet-b0, and MobileNet. Each of these encoders brings unique strengths to the model. ResNet50, renowned for its deep residual networks, is great at learning complex patterns. EfficientNet-b0, designed with a balance between depth, width, and resolution, ensures efficiency. MobileNet, known for its lightweight structure, provides a balance between computational load and model performance.

DeeplabV3 with EfficientNet-b0: DeeplabV3 is a state-of-the-art model for semantic image segmentation, which utilizes an EfficientNet-b0 encoder. With its advanced atrous convolution and spatial pyramid pooling modules, DeeplabV3 is capable of capturing multi-scale context and accurately segmenting objects at various scales.

MANet with EfficientNet-b0: MANet stands for Multi-scale Attention Net, which employs an EfficientNet-b0 encoder as well. It is equipped with a multi-scale attention mechanism to capture features at different scales and locations. It emphasizes the region of interest and suppresses the irrelevant areas, thereby improving the model's focus on the inked regions.

By comparing and analyzing the performance of these models, we aim to identify the most effective architecture for the unique task of detecting ancient ink on carbonized papyrus scrolls.

3.5. loss function

The Binary Cross Entropy (BCE) loss function was used as the main metric in the project. BCE is commonly employed in binary classification tasks, measuring the difference between actual labels and model predictions. It quantifies dissimilarity using this formula.

$$L = -[y \cdot \log(p) + (1 - y) \cdot \log(1 - p)] \tag{1}$$

Where y represents the true label and p is the predicted probability. The loss adjusts based on the true label: when y is 1, the loss decreases as the prediction approaches 1, indicating accuracy, and when y is 0, the loss penalizes deviations from 0. BCE aids in training model optimization by penalizing incorrect predictions.

3.6. evaluation method

By mentioned by TMYOK[18], this project used a modified dice coefficient to evaluate the model's prediction.

The Dice coefficient, also known as the Dice similarity coefficient (DSC), is a statistical measure of the similarity between two sets. In the context of image segmentation, it is used to quantify the overlap between the predicted and actual binary images.

In this custom function, a beta parameter is introduced to balance the precision and recall. This is similar to the F-beta score in traditional classification problems, where beta > 1 favors recall and beta < 1 favors precision.

The function first applies a sigmoid activation to the predictions to map them to a range between 0 and 1. After flattening the prediction and target tensors, it calculates the true positive (ctp), false positive (cfp), and true count values.

Finally, it calculates the custom Dice coefficient using the following formula:

$$Dice = \frac{(1+\beta^2) \times (c_{precision} \times c_{recall})}{\beta^2 \times c_{precision} + c_{recall} + smooth}$$

where

$$c_{precision} = \frac{ctp}{ctp + cfp + smooth}$$

$$c_{recall} = \frac{ctp}{y_{true\_count} + smooth}$$

Here, 'smooth' is a small constant added to the denominators to prevent division by zero. The resulting Dice coefficient ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap.

4. Results

Table 1. result with same encoder efficientnet-b0

model name	val score
Unet	0.4844
MANet	0.4582
DeepLabV3	0.4567

Table 2. result with same architecture Unet

encoder	val score
efficientnet-b0	0.4844
resnet50	0.4637
mobilenetV3	0.4306

The first table provides a comparison of different models, all using the efficientnet-b0 as the encoder. The Unet model achieved the highest validation score of 0.4844, followed by the MANet and DeepLabV3 models, which scored 0.4582 and 0.4567, respectively. This indicates that when efficientnet-b0 is used as the encoder, the Unet model outperforms the other models.

The second table compares the performance of the Unet model with different encoders. Here, the efficientnet-b0 encoder again leads with a validation score of 0.4844, followed by resnet50 and mobilenetV3, which achieved validation scores of 0.4637 and 0.4306, respectively. This suggests that the Unet model with efficientnet-b0 as the encoder provides the most accurate results among the tested configurations.

5. Discussion

In this project, we examined the impact of different models and encoders on performance, with the ultimate goal of interpreting the text on papyrus. However, the current performance of the model is still far from being able to fully decipher the text. It can decipher some parts of the text, but there is a significant gap to completely understanding the text.

For future work, our focus should be on improving the performance of the model. Here are a few potential directions: 1. Use model fusion techniques: Since we have three training sets, we can train two at a time to form three different models. When making predictions, we can use an ensemble vote model to make a decision. 2. Try more model combinations: We can experiment with more encoders and model architectures. 3. Hyper-parameter tuning: Adjusting the learning rate, batch size, and other hyper-parameters could lead to better model performance. 4. Custom loss functions: We can experiment with other loss functions tailored for this specific task. 5. Incorporating domain knowledge: If certain patterns or features are known to be important for ink detection, these could be incorporated into the model design or training process.

6. Conclusion

In this study, we explored the impact of different machine learning models and encoders on the task of deciphering text on ancient papyrus. We found that while we can successfully identify some sections of the text, fully decoding it remains a challenging task. This suggests that there is substantial room for improvement and offers several promising avenues for future research.

We learned that the performance of the models significantly depends on the choice of architecture and the encoder used. Among the tested setups, the U-Net model with EfficientNet-b0 as the encoder demonstrated the best results, which highlights the potential of this combination for future explorations.

However, we must acknowledge that we are still far from fully decoding the text on the papyrus. This serves as a reminder of the complexity of the task at hand and the challenges associated with the interpretation of ancient texts. But with the rapid advancement in machine learning techniques, we are optimistic about the potential for progress in this field.

In future work, we will investigate techniques such as model fusion, hyper-parameter tuning, advanced data augmentation, and the use of larger, pre-trained models. We believe that these strategies, among others, will help us enhance the performance of our models, bringing us closer to our goal of fully decoding the ancient papyrus text.

References

[1] Ajland. [eda] a slice-by-slice analysis. Kaggle, 2023. Retrieved May 12, 2023. 3

[2] M.L. Amadori, S. Barcelli, G. Poldi, F. Ferrucci, A. Andreotti, P. Baraldi, and M.P. Colombini. Invasive and non-invasive analyses for knowledge and conservation of roman wall paintings of the villa of the papyri in herculaneum. *Microchemical Journal*, 118:183–192, 2015. 2

[3] Steven W. Booras and Douglas M. Chabries. The herculaneum scrolls. In *Image Processing, Image Quality, Image Capture Systems Conference*, 2001. 2

[4] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017. 2

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2

[6] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017. 2

[7] K. Li, X. Wu, D.Z. Chen, and M. Sonka. Optimal surface segmentation in volumetric images - a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134, 2006. 3

[8] Rui Li, Shunyi Zheng, Ce Zhang, Chenxi Duan, Jianlin Su, Libo Wang, and Peter M. Atkinson. Multiattention network for semantic segmentation of fine-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. 2

[9] Khushboo Mehra, Hassan Soliman, and Soumya Ranjan Sahoo. Data augmentation using feature generation for volumetric medical images, 2022. 3

[10] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey, 2020. 2

[11] Stephen Parsons, C. Seth Parker, Christy Chapman, Mami Hayashida, and W. Brent Seales. Educelab-scrolls: Verifiable recovery of text from herculaneum papyri using x-ray ct, 2023. 2, 3

[12] Priya Rani, Shallu Kotwal, Jatinder Manhas, Vinod Sharma, and Sparsh Sharma. Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments. *Archives of Computational Methods in Engineering*, 29(3):1801–1837, May 2022. 1

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 2

[14] W. Brent Seales and Y. Lin. Digital restoration: Principles and approaches. In *Proceedings of DELOS/NSF Workshop on Multimedia in Digital Libraries*, 2003. 2

[15] W.B. Seales. Reading the invisible library: A retrospective. In Carl Brune and Caroline Foutch, editors, *Modern Alchemy: New Technology for Museum Collections*. Gilcrease Museum, 2017. 2

[16] Pratap Chandra Sen, Mahimarnab Hajra, and Mitadru Ghosh. Supervised classification algorithms in machine learning: A survey and review. In Jyotsna Kumar Mandal and Debika Bhattacharya, editors, *Emerging Technology in Modelling and Graphics*, pages 99–111, Singapore, 2020. Springer Singapore. 2

[17] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. 2

[18] TMYOK. Modified dice coefficient implementation. Kaggle, 2023. Retrieved May 12, 2023. 5