

Deep Learning for Ink Detection in Ancient Herculaneum Scrolls

André Thomas Gil Cifuentes
Jhonatan David Ardila Gomez

Deep Learning
Universidad de Antioquia, Colombia

October 4, 2025

1 Introduction

The *Vesuvius Challenge – Ink Detection* is a competition organized by *Kaggle* in collaboration with the international *Scroll Prize* project. Its purpose is to apply advanced deep learning techniques to reveal hidden texts in carbonized papyrus scrolls from *Herculaneum*, a Roman city destroyed by the eruption of *Mount Vesuvius* in 79 AD.

The papyrus fragments have been scanned using **X-ray computed tomography (CT)**, producing high-resolution three-dimensional volumes of their internal structure. However, both the papyrus substrate and the ink are primarily composed of carbon, which makes them visually indistinguishable in standard X-ray imagery. As a result, detecting the subtle variations in density that correspond to ink requires sophisticated image processing and machine learning models capable of learning these fine-grained differences.

2 Objectives

The main goal of this project is to apply deep learning techniques to solve the ink detection segmentation task. Specifically, we aim to develop a model capable of accurately identifying the presence of ink within the three-dimensional X-ray scans of papyrus fragments. By effectively distinguishing inked from non-inked regions, the model contributes to the broader effort of virtually reconstructing and reading the ancient scrolls.

3 Dataset Description

As describe in [1], the dataset for this project consists of **3D X-ray scans of detached fragments of ancient papyrus scrolls**.



Figure 1: Example of a 3D X-ray fragment used in the Vesuvius Challenge dataset.

Each papyrus fragment is represented as a **stack of 65 grayscale images** (.tif format), corresponding to slices along the z -axis. Together, these slices form a 3D volume of voxel data with dimensions `width × height × 65`.

The dataset is divided into two main parts:

- **Training set:** 3 fragments, each containing 65 image slices, along with associated ground truth labels.
- **Test set:** 2 fragments, each also containing 65 image slices, provided without ink labels for model evaluation.

4 Evaluation Metrics

To evaluate the performance of the model, we use a modified version of the Sørensen–Dice coefficient, specifically the **$F_{0.5}$ score**. This metric places a higher emphasis on *precision* than on *recall*, which encourages the model to form coherent ink patterns rather than scattered false positives.

The F_β score is defined as:

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

where

$$p = \frac{tp}{tp + fp}, \quad r = \frac{tp}{tp + fn}, \quad \beta = 0.5$$

Here, tp , fp , and fn represent the number of true positives, false positives, and false negatives, respectively. Using $\beta = 0.5$ increases the weight of precision relative to recall, leading to more reliable identification of actual ink areas.

At the time of evaluation, the model output is expected to be a **binary mask**, where 0 indicates “no ink” and 1 indicates “ink”.

References

- [1] Vesuvius Challenge. The vesuvius challenge – ink detection.
<https://scrollprize.org/>, 2023.