Volumetric Fast Fourier Convolution for Detecting Ink on the Carbonized Herculaneum Papyri

Fabio Quattrini, Vittorio Pippi, Silvia Cascianelli, Rita Cucchiara
University of Modena and Reggio Emilia
Via Pietro Vivarelli, 10, Modena (Italy)
{name.surname}@unimore.it

Abstract

Recent advancements in Digital Document Restoration (DDR) have led to significant breakthroughs in analyzing highly damaged written artifacts. Among those, there has been an increasing interest in applying Artificial Intelligence techniques for virtually unwrapping and automatically detecting ink on the Herculaneum papyri collection. This collection consists of carbonized scrolls and fragments of documents, which have been digitized via X-ray tomography to allow the development of ad-hoc deep learning-based DDR solutions. In this work, we propose a modification of the Fast Fourier Convolution operator for volumetric data and apply it in a segmentation architecture for ink detection on the challenging Herculaneum papyri, demonstrating its suitability via deep experimental analysis. To encourage the research on this task and the application of the proposed operator to other tasks involving volumetric data, we will release our implementation (https://github.com/aimagelab/vffc).

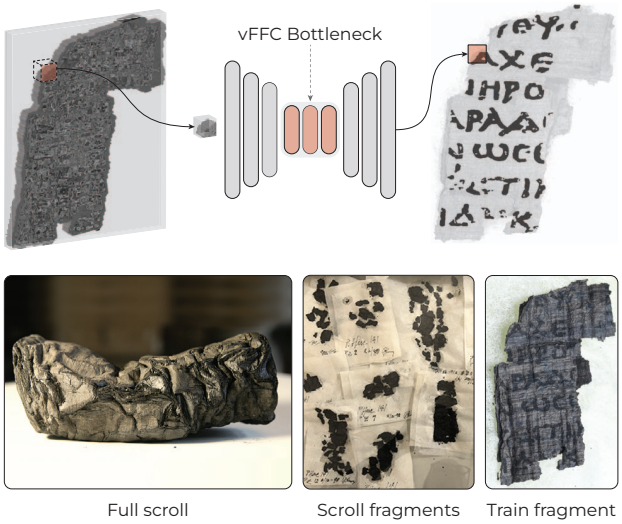


Figure 1. We propose an architecture featuring our devised volumetric Fast Fourier Convolution operator for the ink detection task on the fragments of carbonized Herculaneum papyri (parts of this image are from the official project website https://scrollprize.org/).

1. Introduction

Some of the most valuable sources of information we have about ancient cultures and populations are the manuscripts and, in general, the artifacts with writings and pictures that survived history [37, 8, 26, 35]. For this reason, even the smallest of such objects is precious for its potentially unique and impactful content. Due to the fragility of the medium and their long history, most of the ancient manuscripts found by archaeologists can be extremely degraded and irreversibly damaged, and thus, challenging to handle and analyze. These challenges are even more critical for those ancient documents that were written on scrolls, which also necessitate being unrolled for reading.

Digital Document Restoration (DDR) aims to provide access to these kinds of documents and has therefore gained great interest from researchers and practitioners [12, 6, 34, 7, 55, 13]. Among the DDR techniques, Virtual Unwrapping can be applied when the textual content is physically unreachable, as in the case of fragile scrolls. Starting from a digital 3D volumetric representation obtained with X-ray micro-computed tomography [21], this method entails reconstructing the 2D text image, allowing researchers, scholars, and the general public to visually inspect and study these historical artifacts. Early works on the application of Virtual Unwrapping were limited to specific use cases and entailed semi-manual pipelines [43, 23]. Later, fully automated digital unwrapping was applied to undamaged or partially damaged scrolls made of parchments [40, 50, 49, 39], bamboo [48, 47], papyri [3, 2], silver [16, 4], and the famous En-Gedi scroll [44], charred and charcoaled by a fire in early middle ages.

When discussing ancient irreparably damaged scrolls, few examples carry as much significance and present as

many challenges as the Herculaneum papyri. This collection of more than 1800 manuscripts, recovered from the Herculaneum Villa of the Papyri, presents a particularly important case due to its unique preservation state and the special interest of scholars [19]. Carbonized and buried in ashes during the Vesuvius eruption of 79 C.E., they represent the only intact surviving library from antiquity discovered in its original location [19]. Among the posed challenges, the ink is carbon-based, different from the metal-based ink in the En-Gedi scroll. Thus, the response of the ink to X-rays is not seemingly distinguishable from that of the carbonized papyrus support, making traditional Virtual Unwrapping ineffective.

In the sight of this, in 2023, Parsons *et al.* [33] proposed to push forward DDR on this challenging collection by resorting to modern Computer Vision, Document Analysis, and Artificial intelligence techniques. To this end, they developed and released EduceLab-Scrolls, an open dataset containing volumetric scans of rolled scrolls and detached fragments from the Herculaneum papyri, which is the object of a dedicated ongoing competition[1]. The goal is to develop advanced Virtual Unwrapping strategies to unroll these challenging scrolls and then perform ink detection on the unrolled sheets. Meanwhile, the fragments can be used to develop ink detection algorithms. In fact, for the fragments, it has been possible to obtain infrared images and, thus, ink maps that can be used for training supervised deep learning models. Thanks to their effort, this novel task on such challenging data is receiving increasing interest from the Artificial Intelligence community.

In this respect, we propose to tackle the task via an efficient, fully-Convolutional model featuring blocks inspired by Fast Fourier Convolutions (FFCs) [9], which we modify to handle volumetric data (Figure 1). Note that the FFC operator exploits the spectral information of the input to expand its receptive field in the frequency domain and to handle pseudo-periodic patterns. Spatial FFC has been employed for tasks on 2D images but, to the best of our knowledge, it has never been applied to volumetric data. For this reason, we propose a modification to the original FFC operation that makes it able to handle volumetric data and thus be applied to the Herculaneum papyrus fragments scans for ink detection. Through a deep evaluation analysis, we provide useful insights on the challenging task of ink detection on Herculaneum papyri and on the proposed approach, which we demonstrate to be suitable for the task.

2. Related Work

Digital Document Restoration. DDR encompasses Digital Imaging, Image Processing, and Computer Vision techniques for non-invasive content recovery from severely damaged ancient documents [12, 6, 34, 7, 55, 13]. One of these techniques is Virtual unwrapping [42], which is mostly applied to documents that are too fragile to handle and analyze, as in the case of scrolls. First, the entire scroll is scanned, typically with X-ray micro-computed tomography (micro-CT) [21]. Then, each scroll sheet in the volumetric representation is segmented and projected into a 2D image. Depending on the specific use case, an additional texturing step (such as ink detection) can be performed. Early works operated on small contrived samples with semi-manual pipelines [43, 23]. The first fully-automated solution was proposed by Samko *et al.* [40], who developed a novel graph cut method for parchment scrolls sheets segmentation. In [16, 4], the authors read the contents of a damaged silver scroll by exploiting the specific characteristics of the material, such as the engraved and ruled text. Finally, the carbonized En-Gedi scroll was read in 2015 by Seales *et al.* [44].

Herculaneum papyri. This work focuses on DDR of the Herculaneum papyri, a document collection that poses unprecedented challenges [41] to the application of a classical virtual unwrapping pipeline. Firstly, the papyrus layers have been compressed, crumpled, and deformed by the carbonization process, making sheet segmentation not straightforward. Secondly, the text is mostly written using carbon-based ink, which is almost invisible on the carbonized papyrus support in the X-ray scans. Thus, traditional virtual unwrapping, which heavily relies upon the visibility of ink or simple layers structure, is not applicable. In a recent work, Parker *et al.* [32] proved the detectability of carbon-based ink in micro-CT scans of the Herculaneum papyri using a 3D Convolutional Neural Network (Conv) [17] trained on subvolumes to detect ink in the central voxel. A further step towards DDR of Herculaneum papyri is due to Parsons *et al.* [33], who proposed an open dataset containing volumetric scans of such collection. In this work, we tackle the problem of ink detection on the surface volumes of Herculaneum papyri fragments.

Spectral Analysis. Previous DDR approaches, both classical [29] and learning-based [1, 24], have successfully leveraged the spectral information of document images to enhance their visual quality. In recent works, the Discrete Wavelet Transform [27] has been used for Document Binarization to represent the document image in different frequency sub-bands, either to perform Segmentation [1] or to obtain a ground truth ink map to train a Generative Adversarial Network [24]. Our method exploits the Fourier Transform, localized only in frequency, to operate on the periodic structures common in the writing substrate. Unlike previous works, we do not use spectral information to represent the images but as an operator inside our proposed end-to-end Convolutional architecture. In particular, we extend the FFC operator proposed by Chi *et al.* [9] to make it process

[1] https://scrollprize.org/

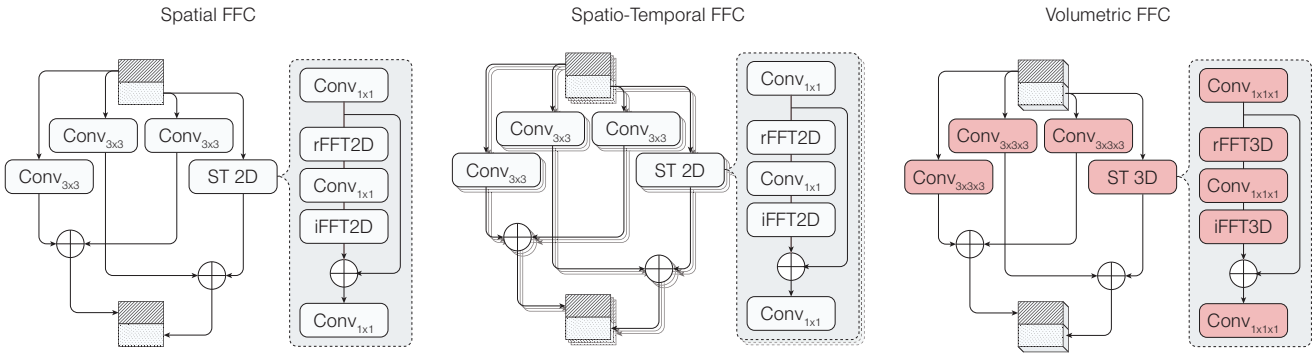


Figure 2. The standard spatial FFC (left) can be adapted to work on spatio-temporal data by replicating it along the channel dimension (center). Our proposed volumetric FFC (right) is designed to handle 3D volumetric data directly by combining 3D convolutions and 3D FFTs. For simplicity, we omit the batch normalization and ReLU operations in the schemes.

volumetric data. Indeed, the FFC operator combines 2D convolutions and 2D discrete Fourier Transform (DFT) [10] and has been applied to spatial data for computer vision tasks such as inpainting [15], super-resolution [45, 54], and semantic segmentation [5]. Some attempts have been made to apply the FFC to spatio-temporal data [9]. Nonetheless, volumetric data are different from spatio-temporal ones, and thus, require ad hoc solutions [53, 36, 14, 31]. Therefore, we argue that a solution based on spatio-temporal FFCs is not optimal for modeling the correlations between dimensions in volumetric data. In sight of this, we propose the volumetric FFC operator, which features 3D convolutions and 3D discrete Fourier Transform and better handles this kind of data.

3. Proposed Approach

Our goal is to detect the presence of ink in the volumetric representation of a carbonized papyrus sheet. Our model takes as input such volumes and is expected to output an ink map with the same width and height of the volume, whose elements contain the probability of ink being in the corresponding papyrus surface. To tackle this task, we devise a U-net-like architecture, whose details are given in Section 3.2, featuring a variant of the FFC operator that we devise to handle volumetric data as described below.

3.1. Volumetric Fast Fourier Convolution

The FFC proposed by Chi *et al.* [9], which here we refer to as spatial FFC, is a neural operator that combines convolution and Fourier Transform to perform local reasoning in the space domain and non-local reasoning in the frequency domain. This information processing expands its receptive field and makes it suitable for handling pseudo-periodic patterns in the data. The operator has also been applied to spatio-temporal data [9]. Here, we refer to this variant as Spatio-Temporal FFC (stFFC). The Volumetric FFC variant devised in this work presents the same properties of the spatial FFC operator but can directly handle volumetric data. A visual comparison between these mentioned operators is reported in Figure 2.

The vFFC operator takes as input a tensor $\mathbf{X} \in \mathbb{R}^{D\times H\times W\times 2C}$, where $2C$ is the number of channels, D is the depth dimension, and H and W are the spatial dimensions. This tensor is split into two parts along the channel dimension, which are fed to two interconnected branches: a local branch and a global branch. Splitting the input tensor into two chunks allows the encoding of different information in separate regions of the tensor. In addition, this approach permits the global and local branches to specialize in different aspects because they do not share the same inputs. The *local branch* contains two 3D convolutional layers with kernel size 3 and is in charge of modeling local volumetric information. In the *global branch*, the input is mapped into the spectral domain to model global information. This branch consists of a 3D convolution with kernel size 3 and a 3D Spectral Transform (ST 3D) block.

The ST 3D block exploits a three-dimensional real Fast Fourier Transform (FFT3D) [10]. Specifically, the FFT3D is performed across the depth, height, and width dimensions of the input tensor:

$$\text{FFT3D}(\mathbf{X}) = \mathbf{Z} = \mathbf{R} + i\mathbf{I},$$

where $\mathbf{Z}$ is a complex tensor, with real and imaginary parts $\mathbf{R}, \mathbf{I} \in \mathbb{R}^{D\times H\times \frac{W}{2}\times C}$. Then, $\mathbf{R}$ and $\mathbf{I}$ are stacked together along the channel axis, thus obtaining a tensor

$$\mathbf{R}||\mathbf{I} \in \mathbb{R}^{D\times H\times \frac{W}{2}\times 2C}$$

that is then fed to a $1\times1\times1$ convolutional layer. Batch normalization and ReLU activation functions are applied to the output of the convolution. The resulting tensor is reshaped into a complex-valued tensor $\mathbf{Z}' \in \mathbb{C}^{D\times H\times \frac{W}{2}\times C}$, and the inverse real Fast Fourier Transform (iFFT3D) is computed to obtain the global branch output:

$$\text{iFFT3D}(\mathbf{Z}') = \mathbf{X}' \in \mathbb{R}^{D\times H\times W\times C}.$$

The output of each branch is summed to the output of the other, and the resulting tensors are fed to separate batch normalization and ReLU operation before being stacked together to obtain the final vFFC output tensor.

3.2. Ink Detection Architecture

Rather than having a different response to X-rays than the papyrus support, the carbon-based ink in the Herculaneum scrolls ink modifies the substrate structural patterns in subtle ways, as shown in [32]. In particular, parts containing ink have different types of cracks and densities and are thicker and smoother than empty ones. In this challenging case, it is important to consider both local and global contexts. The subtle patterns are, in fact, very localized, but with a global context, the model is able to consider holistically information from the whole subvolume. Moreover, we argue that the vFFC operator, which is able to handle pseudo-periodic patterns, is suitable for modeling this kind of data. We treat ink detection as a pixel-wise classification task, and we use an encoder-bottleneck-decoder architecture composed of a 3D convolutional encoder, a vFFC-based bottleneck, and a 2D convolutional decoder. To accurately localize the signal and fuse high-level and low-level features, we employ skip connections between the encoder and the decoder, similar to U-Net [38], and vFFC layers in the latent space of the network.

Specifically, we employ a 3D ResNet-34 [15] to encode the subvolumes. The output from the last block is fed to 3 vFFC Residual Blocks, each combining two vFFC layers with a residual connection. The $\mathbb{R}^{D\times H\times W\times C}$ feature maps from the bottleneck are collapsed to $\mathbb{R}^{H\times W\times C}$ by averaging the depth dimension. The same applies to the encoder activations in the skip connections. Then, a 2D decoder reconstructs the ink map. This decoder is obtained by stacking four blocks made of 2-fold bilinear interpolations and convolutions, operating on the concatenation of the output of the previous layer and the features from the corresponding encoder layer. Since the fragments have very high resolution but represent small objects, the decoder does not reconstruct the full-scale ink map but rather a 4-fold downscaled one. This approach also limits the computational weight of the model.

3.3. Training and Inference

We train our model on subvolumes (3D patches) of the fragments scans both to limit the computational complexity of our model and to obtain more samples from the same fragment. Note that this strategy is customary for other data-constrained tasks such as document binarization [51] or medical images segmentation [20, 11]. Moreover, we apply a number of data augmentation operations (described in Section 4) in order to reduce the risk of overfitting and help the model generalize to unseen data. Our adopted loss function is a combination of the Dice loss and weighted binary cross entropy (WBCE). The Dice score coefficient has been proposed for unbalanced segmentation [28], while the weighted binary cross-entropy is customary for classification. Let N be the number of voxels, $p_i \in P$ the predicted binary segmentation map, and $g_i \in G$ the ground-truth ink image. The loss can be written as:

$$\mathcal{L} = \text{Dice} + \text{WBCE},$$

with

$$\text{Dice} = \frac{2\sum_n^N p_n g_n + \epsilon}{\sum_n^N p_n^2 + \sum_n^N g_n^2 + \epsilon},$$

$$\text{WBCE} = -\frac{1}{N}\sum_{n=1}^{N} w[g_n \log p_n + (1-g_n)\log(1-p_n)]$$

where w is the weight attributed to the ink class. The predicted ink map contains values ranging from 0 to 1, representing the ink presence probability. From this, we obtained a binarized image by applying a threshold of 0.5.

4. Experimental Analysis

In this section, we present the experiment setup concerning the ink detection task on the fragments of the recently proposed EduceLab-Scrolls dataset and detail our training strategy. Moreover, we present an extensive analysis of the proposed components, both to explore their contribution and give some intuitions on what is more suitable for this task.

4.1. Experiment Setup

Dataset. As mentioned above, in this work, we focus on the surface volumes of the fragments in the EduceLab-Scrolls [33] (see Figure 3). These scans have been performed with X-ray micro-CT and thus are very detailed. With a $3.2\mu m$ voxel size, they have resolutions in the order of thousands of pixels and weigh several Gygabites. In particular, we use the three publicly available fragments released for the Vesuvius Challenge on Ink Detection related to the EduceLab-Scrolls dataset project. These documents, broken and detached from the scrolls during destructive physical unrolling tentatives, have visible ink and, thus, it has been possible to obtain a ground truth ink map, also thanks to infrared analysis. The scans are all composed of 65 slices but have different spatial sizes: 6330×8181 for fragment 1, 9506×14830 for fragment 2, and 5249×7606 pixels for fragment 3. In our experiments, we use fragment 1 for test and the other two for training.

Evaluation Metrics. Considering the novelty of the task and its similarities to other tasks, we use a combination of metrics to evaluate our models. First of all, we adopt the

Figure 3. Fragments contained in the considered dataset, in different modalities.

score proposed in [33] to evaluate the ink detection performance, *i.e.*, the F_β score, which weighs more the precision than the recall and is defined as follows:

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

where p and r are precision and recall, respectively, and the parameter β=0.5. Moreover, we employ some scores commonly used in Document Binarization, namely the pseudo-FMeasure (pFM) [30] and the Peak Signal-to-Noise Ratio (PSNR), which measures the similarity of two images.

Implementation Details. Considering that most of the information is in the central slices and to reduce computational impact, we use 16 slices for training. Note that using more slices would not straightforwardly improve performance: in fact, fragments have hidden text, in papyrus layers fused in the back, that could be present in the surface slice if we go too deep. Taking also into account that the ink is a localized signal, we train on 16×256×256 subvolumes.

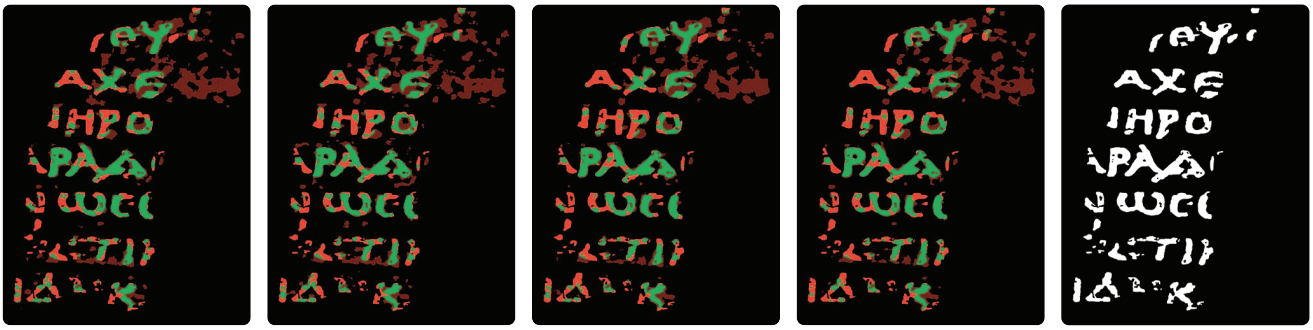
For optimization, we use the AdamW [25] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.95$, and the OneCycle scheduler [46] with learning rate 0.003. We set the batch size to 4 and apply floating point 16 mixed precision to speed up training. For regularization, we use Drop Path [22] with rate 0.1, Channel Dropout [52], with rate and maximum dropped channels both set to 0.5, and additional data augmentation strategies as described in Section 4.1. We set the ink weight in the WBCE to 1 and train the models for 20 epochs, which takes 24 hours with an NVIDIA RTX A5000. At test time, for the entire test fragment, we perform the prediction on 16×256×256 subvolumes and keep only the prediction on the inner 16×128×128 part to reduce border artifacts.

Data Augmentation. We perform data augmentation in order to enhance model generalization and robustness. In particular, we obtain the training subvolumes starting from a 3D lattice of the whole fragment scans, with cells of size 32×512×512 and stride 64. During each training iteration, we extract subvolumes of size 16×256×256 from a random 3D position of the lattice cells. The rationale behind this strategy is twofold. In the spatial dimension, the random crop increases the variability in the training data, preventing the network from being exposed to the same patches repetitively. In the depth dimension, exposing the network to different slices ensures that the model learns to discern patterns across all depth coordinates. Indeed in a real-world application, the surface volumes would be extracted from segmented sheets in a scroll. This process is prone to errors and misalignment. Thus, there is rarely correspondence between depth coordinates in different samples. By enforcing depth invariance, the network is able to model patterns independently of specific depth locations. Moreover, we randomly apply horizontal and vertical flip, 90° random rotation, and transposition to obtain the corresponding Dihedral group D_4 transformations, further increasing diversity in spatial patterns.

4.2. Results

In this section, we report an analysis aimed at identifying the most relevant elements of the training procedure and assessing the effectiveness of our proposed vFFC-based model for the ink detection task.

Training Startegy. First, we analyze the contribution of the applied data augmentation strategies. The results of this analysis are reported in Table 1. It can be observed that the Dihedral transformations bring the most benefits to the training by inducing the larger variability. From these ex-



no bottleneck | stFFC bottleneck | 3D-Conv bottleneck | vFFC bottleneck | Ground truth

Figure 4. Qualitative results of the considered variants on the test fragment. Red indicates the missed ink predictions, green indicates the pixels correctly identified as ink, and maroon indicates the background pixels incorrectly classified as ink.

Dihedral Transform	Random Crop	Channel Dropout	F_β	pFM	PSNR
-	-	-	0.37	0.53	9.37
✓	-	-	0.46	0.57	**10.38**
✓	✓	-	0.46	0.57	9.62
✓	✓	✓	**0.47**	**0.58**	10.09

Table 1. Ablation analysis on the augmentation strategies adopted.

Loss	w	F_β	pFM	PSNR
Dice	-	0.40	0.56	9.47
WBCE	1	0.44	0.56	**10.34**
WBCE + Dice	5	0.41	0.53	8.20
WBCE + Dice	2	0.45	0.56	9.86
WBCE + Dice	1	**0.47**	**0.58**	10.09

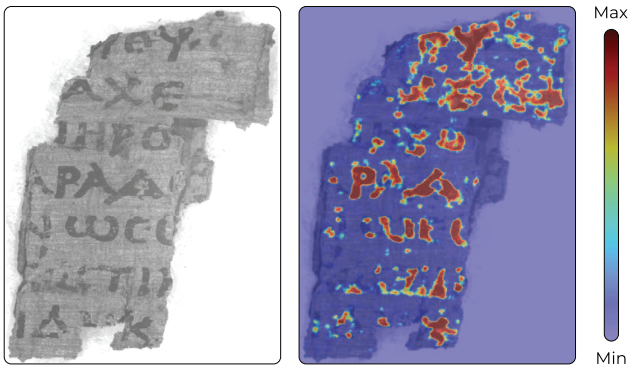
Table 2. Ablation analysis on the training loss function.

periments on fragment 1, the benefit of the random crop is not evident. Nonetheless, when testing on the public and the private test set of the Kaggle Ink Detection Challenge associated with the EduceLab-Scrolls dataset, this regularization strategy has been proven beneficial for increasing the generalization capability of our model in handling fragments in which the relevant information is localized in different slices. The effect of the random crop, especially on the depth axis, is visible from the activation maps discussed in Section 4.2.1. Then, we perform an ablation on the training loss terms, whose results are reported in Table 2. It emerges that the WBCE leads to good performance, and the combination with the Dice helps refine the results. Moreover, giving more weight to the ink class leads to worse performance, despite the ink class being much less represented than the background.

Architecture. We compare the proposed architecture with a number of baselines in Table 3. To assess the effectiveness

Bottleneck	F_β	pFM	PSNR
-	0.46	**0.58**	9.87
stFFC	0.45	**0.58**	9.76
3D-Conv	0.46	**0.58**	10.04
vFFC	**0.47**	**0.58**	**10.09**

Table 3. Quantitative comparison between our model featuring vFFCs in the bottleneck and baselines with different bottlenecks.



Infrared image | Probability map

Figure 5. Prediction probability map of our proposed model on the test fragment.

of vFFCs, we compare our proposed model against variants featuring different kinds of bottlenecks. In particular, we consider a variant without bottleneck layers, one with stFFCs, and one with 3D-Convs. Overall, using the proposed vFFC in the bottleneck leads to the best performance. Arguably, this is due to the more precise prediction of the background pixels, as can also be observed from the qualitative comparison in Figure 4 (see, *e.g.*, the top-right part of the prediction map).

Kaggle Ink Detection Challenge Results. We participate in the Kaggle Vesuvius Challenge - Ink Detection[2] with a

[2] www.kaggle.com/competitions/vesuvius-challenge-ink-detection

boosted version of our model. In particular, we maintain the architecture but employ a different training and inference strategy to fully exploit the training fragments available. In particular, we split each available fragment into two parts in order to obtain a total of six subsets. Then, we train six versions of our model in a k-fold strategy and set the prediction threshold to 0.8. We then submit an ensemble of these models, combined via majority voting on each pixel with four votes out of six. The submission scored silver medal (*i.e.*, within the top-5% submissions) in the competition, scoring around F_β=0.70 on the public test set and F_β=0.60 on the private one.

4.2.1 Visualizations

For reference, in Figure 5, we report the ink prediction probability of our model. Although we set the prediction threshold to 0.5, we argue that such a non-thresholded visualization can be helpful when visually inspecting the fragments.

Finally, we study the model class activation map by using LayerCam [18]. In particular, we consider the whole depth size for a given 2D patch on the papyrus surface, obtaining a volume $\mathbf{V}$ with shape $65 \times 256 \times 256$. Then, we extract subvolumes of size $\mathbf{v} \subset \mathbf{V}$ with shape $16 \times 256 \times 256$ and stride set to 1, covering the whole Z axis, and feed them to the network. We compute the mean activation value on the spatial dimensions for each depth slice and show the result in Figure 6. On the x-axis of the activation maps plot, we report the value of the starting $z \in Z$ coordinate of the subvolume. On the y-axis, we plot the absolute $z \in Z$ depth coordinate. As we can see, the plot is divided into three regions, corresponding to the upper ($0 \leq z < 24$), middle ($24 \leq z < 40$), and lower ($40 \leq z \leq 65$) parts of the papyrus volume $\mathbf{V}$. When we feed to the model slices corresponding to small values of z, they are all informative; thus, the network does not look specifically at any of them but rather combines the patterns. In the center-most region of the plot, the network consistently focuses on the slice with $z \approx 30$ for several slidings of the subvolume. We argue that this is the main manifestation of the depth invariance that we instill with the subvolume random crop. This way, the network can account for segmentation misalignment in real-world scenarios, learning to recognize important slices independently from their relative position in the subvolume. Then, with growing values of z, the depth slices contain less and less information, so the network focuses on the slices with relatively lower values of z. We provide visualizations of the volume $\mathbf{V}$ in Figure 6: from the depth slice, we can see that the information is for the most part in the low z coordinates, while the infrared image and the ground truth show where ink is present in the surface and changes the inner texture of the papyrus support.

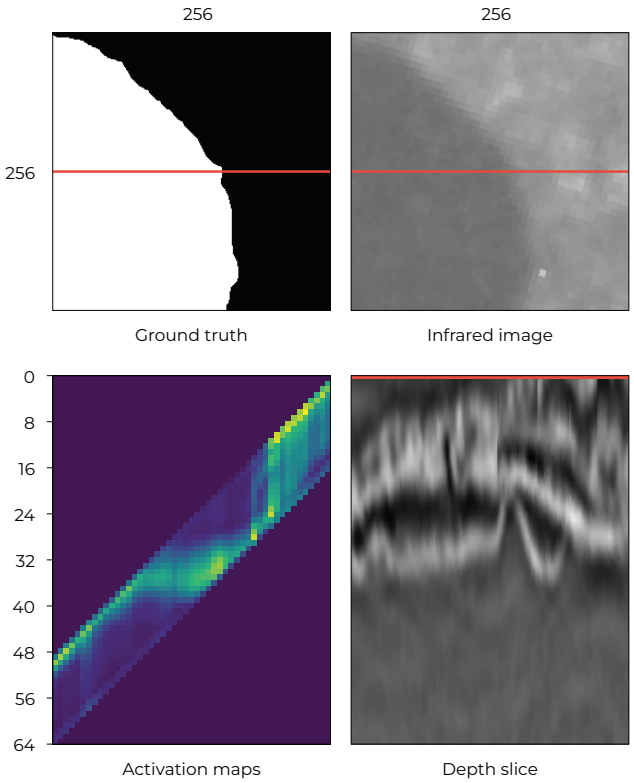


Figure 6. Activation maps by varying the depth and keeping the spatial coordinates of the subvolume fed to our model (blue represents the minimum value, yellow the maximum). For reference, we report the corresponding ground-truth ink map, the infrared image of the surface, and the coronal slice of the subvolume.

5. Conclusion and Future Work

In this work, we have devised the vFFC operator, a modification of the standard FFC designed to handle volumetric data directly. Moreover, we have proposed to incorporate the vFFC in an architecture for the DDR sub-task of ink detection on volumetric scans of carbonized papyri fragments from the EduceLab-Scrolls dataset. Through experimental analysis, we have assessed the effectiveness of our approach and hopefully contribute to the emerging research interest in DDR on the challenging Herculaneum papyri data. Finally, we argue that the vFFC operator could also be applied to other tasks and scenarios involving volumetric data (*e.g.*, medical imaging), which we leave for future work.

Acknowledgement

This work was supported by the "AI for Digital Humanities" project (Pratica Sime n.2018.0390), funded by "Fondazione di Modena" and the PNRR project Italian Strengthening of Esfri RI Resilience (ITSERR) funded by the European Union – NextGenerationEU (CUP: B53C22001770006).

References

[1] Younes Akbari, Somaya Al-Maadeed, and Kalthoum Adam. Binarization of degraded document images using convolutional neural networks and wavelet-based multichannel images. *IEEE Access*, 2020. 2

[2] Dario Allegra, Enrico Ciliberto, Paolo Ciliberto, Filippo Luigi Maria Milotta, Giuseppe Petrillo, Filippo Stanco, and C Trombato. Virtual unrolling using x-ray computed tomography. In *EUSIPCO*, 2015. 1

[3] Dario Allegra, Enrico Ciliberto, Paolo Ciliberto, Giuseppe Petrillo, Filippo Stanco, and Claudia Trombatore. X-ray computed tomography for virtually unrolling damaged papyri. *Applied Physics A*, 122:1–7, 2016. 1

[4] Daniel Baum, Felix Herter, John Møller Larsen, Achim Lichtenberger, and Rubina Raja. Revisiting the jerash silver scroll: A new visual data analysis approach. *Digital applications in archaeology and cultural heritage*, 2021. 1, 2

[5] Bruno Berenguel-Baeta, Jesus Bermudez-Cameo, and Jose J Guerrero. FreDSNet: Joint Monocular Depth and Semantic Segmentation with Fast Fourier Convolutions. *arXiv preprint arXiv:2210.01595*, 2022. 3

[6] Michael S Brown and W Brent Seales. Document restoration using 3d shape: a general deskewing algorithm for arbitrarily warped documents. In *ICCV*, 2001. 1, 2

[7] Huaigu Cao, Xiaoqing Ding, and Changsong Liu. A cylindrical surface model to rectify the bound document image. In *ICCV*, 2003. 1, 2

[8] Silvia Cascianelli, Vittorio Pippi, Maarand Martin, Marcella Cornia, Lorenzo Baraldi, Kermorvant Christopher, and Rita Cucchiara. The LAM Dataset: A Novel Benchmark for Line-Level Handwritten Text Recognition. In *ICPR*, 2022. 1

[9] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *NeurIPS*, 2020. 2, 3

[10] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Math. Comput.*, 1965. 3

[11] Jeffrey De Fauw, Joseph R Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O'Donoghue, Daniel Visentin, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.*, 2018. 4

[12] Andrei Doncescu, Alain Bouju, and Veronique Quillet. Former books digital processing: image warping. In *DIA*, 1997. 1, 2

[13] Basilios Gatos, Ioannis Pratikakis, and Konstantinos Ntirogiannis. Segmentation based recovery of arbitrarily warped document images. In *ICDAR*, 2007. 1, 2

[14] Felix Gonda, Donglai Wei, Toufiq Parag, and Hanspeter Pfister. Parallel separable 3D convolution for video and volumetric data understanding. *arXiv preprint arXiv:1809.04096*, 2018. 3

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3, 4

[16] Gry Hoffmann Barfod, John Møller Larsen, Achim Lichtenberger, and Rubina Raja. Revealing text in a complexly rolled silver scroll from jerash with computed tomography and advanced imaging software. *Sci. Rep.*, 2015. 1, 2

[17] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE Trans. PAMI*, 2012. 2

[18] Peng-Tao Jiang, Chang-Bin Zhang, Qibin Hou, Ming-Ming Cheng, and Yunchao Wei. Layercam: Exploring hierarchical class activation maps for localization. *IEEE Trans. Image Process.*, 30:5875–5888, 2021. 7

[19] Seabrook John. The invisible library, 2015. 2

[20] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.*, 2017. 4

[21] Craig J Kennedy and Tim J Wess. The use of x-ray scattering to analyse parchment structure and degradation. In *Physical techniques in the study of art, archaeology and cultural heritage*. Elsevier, 2006. 1, 2

[22] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *arXiv preprint arXiv:1605.07648*, 2016. 5

[23] Yun Lin and W Brent Seales. Opaque document imaging: Building images of inaccessible texts. In *ICCV*, 2005. 1, 2

[24] Yu-Shian Lin, Rui-Yang Ju, Chih-Chia Chen, Ting-Yu Lin, and Jen-Shiun Chiang. Three-stage binarization of color document images based on discrete wavelet transform and generative adversarial networks. *arXiv preprint arXiv:2211.16098*, 2022. 2

[25] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[26] Martin Maarand, Yngvil Beyer, Andre Kåsen, Knut T Fosseide, and Christopher Kermorvant. A Comprehensive Comparison of Open-Source Libraries for Handwritten Text Recognition in Norwegian. In *DAS*, 2022. 1

[27] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI*, 1989. 2

[28] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV*, 2016. 4

[29] Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, and Mohamed Cheriet. Phase-based binarization of ancient document images: Model and applications. *IEEE Trans. Image Process.*, 2014. 2

[30] Konstantinos Ntirogiannis, Basilis Gatos, and Ioannis Pratikakis. Performance evaluation methodology for historical document image binarization. *IEEE Trans. Image Process.*, 22(2):595–609, 2012. 5

[31] Aniello Panariello, Angelo Porrello, Simone Calderara, and Rita Cucchiara. Consistency-based Self-supervised Learning for Temporal Anomaly Localization. In *ECCVW*, 2022. 3

[32] Clifford Seth Parker, Stephen Parsons, Jack Bandy, Christy Chapman, Frederik Coppens, and William Brent Seales. From invisibility to readability: recovering the ink of herculaneum. *PloS one*, 2019. 2, 4

[33] 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. *arXiv preprint arXiv:2304.02084*, 2023. 2, 4, 5

[34] Maurizio Pilu. Undoing paper curl distortion using applicable surfaces. In *CVPR*, 2001. 1, 2

[35] Vittorio Pippi, Silvia Cascianelli, Christopher Kermorvant, and Rita Cucchiara. How to Choose Pretrained Handwriting Recognition Models for Single Writer Fine-Tuning. 2023. 1

[36] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and Multi-view CNNs for Object Classification on 3D Data. In *CVPR*, 2016. 3

[37] Rizwan Qureshi, Muhammad Uzair, Khurram Khurshid, and Hong Yan. Hyperspectral document image processing: Applications, challenges and future prospects. *Pattern Recognit.*, 90:12–22, 2019. 1

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015. 4

[39] Paul L Rosin, Yu-Kun Lai, Chang Liu, Graham R Davis, David Mills, Gary Tuson, and Yuki Russell. Virtual recovery of content from x-ray micro-tomography scans of damaged historic scrolls. *Sci. Rep.*, 8(1):11901, 2018. 1

[40] Oksana Samko, Yu-Kun Lai, David Marshall, and Paul L Rosin. Virtual unrolling and information recovery from scanned scrolled historical documents. *Pattern Recognit.*, 2014. 1, 2

[41] Brent Seales and Daniel Delattre. *Virtual unrolling of carbonized Herculaneum scrolls: Research Status (2007-2012)*. Macchiaroli editore, 2013. 2

[42] WB Seales. Reading the invisible library: A retrospective. *Modern Alchemy: New Technology for Museum Collections. Gilcrease Museum*, 2017. 2

[43] W Brent Seales and Yun Lin. Digital restoration using volumetric scanning. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, 2004. 1, 2

[44] William Brent Seales, Clifford Seth Parker, Michael Segal, Emanuel Tov, Pnina Shor, and Yosef Porath. From damage to discovery via virtual unwrapping: Reading the scroll from en-gedi. *Science Advances*, 2016. 1, 2

[45] Abhishek Kumar Sinha, S Manthira Moorthi, and Debajyoti Dhar. NL-FFC: Non-Local Fast Fourier Convolution for Image Super Resolution. In *CVPR*, 2022. 3

[46] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pages 369–386. SPIE, 2019. 5

[47] Daniel Stromer, Vincent Christlein, Xiaolin Huang, Patrick Zippert, Tino Hausotte, and Andreas Maier. Virtual cleaning and unwrapping of non-invasively digitized soiled bamboo scrolls. *Sci. Rep.*, 9(1):2311, 2019. 1

[48] Daniel Stromer, Vincent Christlein, Andreas Maier, Patrick Zippert, Eric Helmecke, Tino Hausotte, and Xiaolin Huang. Non-destructive digitization of soiled historical chinese bamboo scrolls. In *DAS*, 2018. 1

[49] Daniel Stromer, Vincent Christlein, Christine Martindale, Patrick Zippert, Eric Haltenberger, Tino Hausotte, and Andreas Maier. Browsing through sealed historical manuscripts by using 3-d computed tomography with low-brilliance x-ray sources. *Sci. Rep.*, 8:15335, 2018. 1

[50] Daniel Stromer, Vincent Christlein, Tobias Schön, Wolfgang Holub, and Andreas Maier. Browsing through closed books: Evaluation of preprocessing methods for page extraction of a 3-d ct book volume. In *IOP Conference Series: Materials Science and Engineering*, 2017. 1

[51] Chris Tensmeyer and Tony Martinez. Document image binarization with fully convolutional neural networks. In *ICDAR*, 2017. 4

[52] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *CVPR*, 2015. 5

[53] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *ICCV*, 2015. 3

[54] Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, and Zhezhu Jin. SwinFIR: Revisiting the SWINIR with fast Fourier convolution and improved training for image super-resolution. *arXiv preprint arXiv:2208.11247*, 2022. 3

[55] Zheng Zhang, Chew Lim Tan, and Liying Fan. Restoration of curved document images through 3d shape modeling. In *CVPR*, 2004. 1, 2