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
{name.surname}@unimore.it

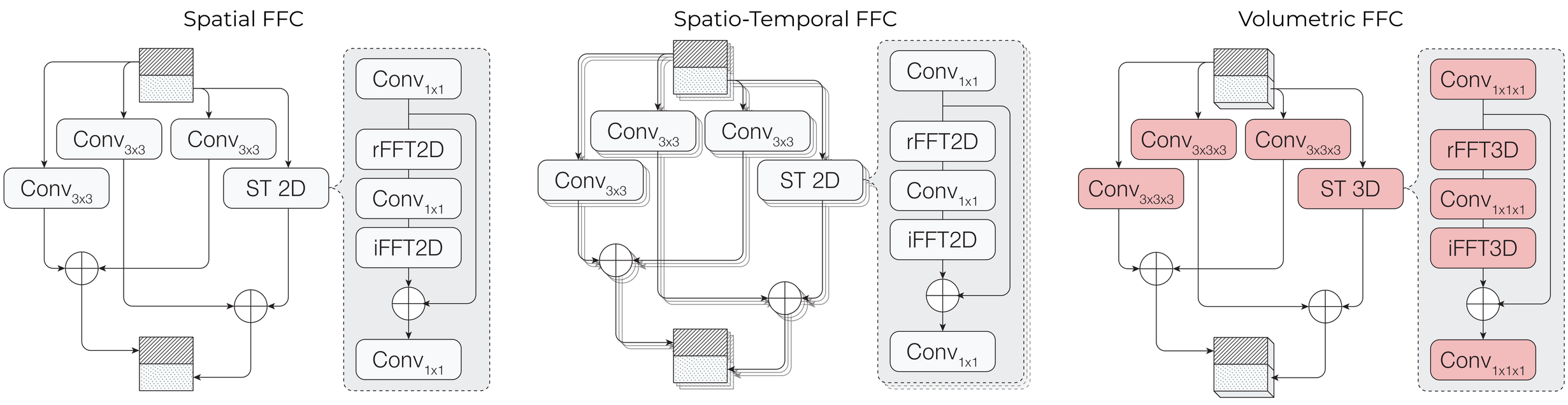
AImage Lab
UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
ICCV23 PARIS
4th Workshop on e-Heritage

Overview

Digital Document Restoration (DDR) aims to enable access to highly damaged documents. In case of physically unreachable documents (e.g., rolls or books too fragile to handle), Virtual Unwrapping is performed to reconstruct the 2D image from a 3D X-ray volumetric representation. However, traditional techniques are not effective for the challenging Herculaneum Papyri due to their carbon-based ink and its seemingly indistinguishable response from carbonized papyrus. In this work, we focus on fragments detached from the Herculaneum Papyri and propose a DDR strategy to detect ink on them. To this end, we devise a variant of the Fast Fourier Convolution (FFC) operator specifically designed to handle volumetric data.

Volumetric FFC

The vFFC splits the input tensor into two parts along the channel axis. A local branch encodes volumetric information with 3D convolutions, while a global branch operates on global patterns, mapping the input to the spectral domain with the FFT3D.

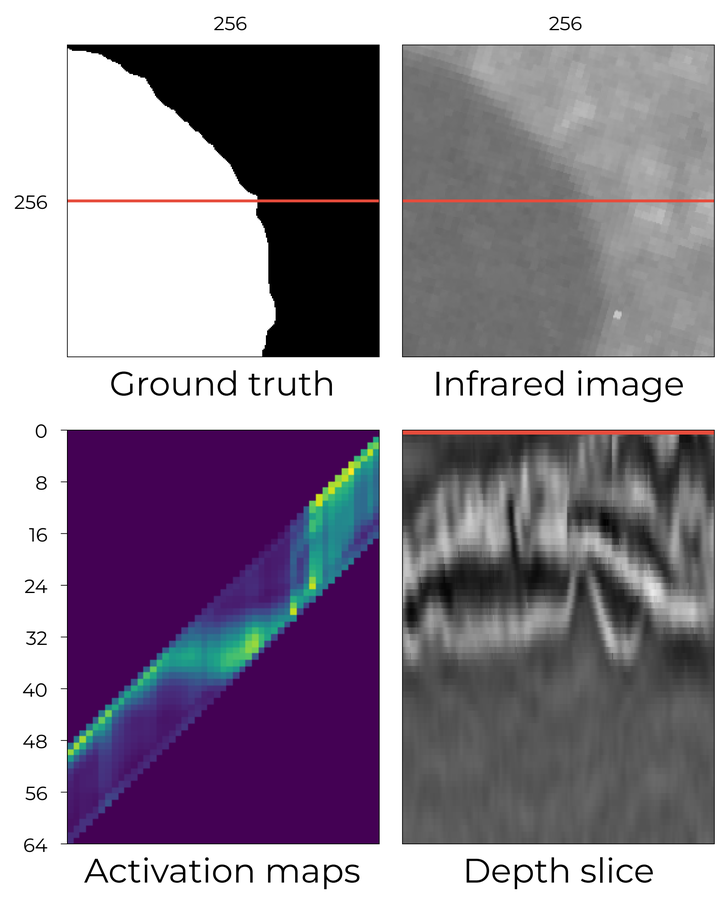


We use an encoder-bottleneck-decoder architecture composed of a 3D convolutional encoder (consisting of a 3D ResNet-34), a vFFC-based bottleneck (comprising 3 vFFC Residual Blocks, each combining two vFFC layers with a residual connection), and a 2D convolutional decoder (consisting of a 2D ResNet34). We employ skip connections between encoder and decoder.

Photo | 3D X-ray scan | Infrared Image | vFFC Bottleneck | Ground Truth

Depth Invariance

During training, we extract subvolumes of the original volumetric scans, sliding along the depth axis. This way, we enforce the network to manage patterns independently from specific depth locations so that it can account for segmentation misalignments in real-world scenarios. Indeed, via LayerCam, we show that the model is able to recognize informative slices independently from their relative position in the subvolume.



Results

We compare the proposed architecture with a number of baselines, all trained on different splits of the dataset, and show that the proposed vFFC in the bottleneck leads to the best performance. Qualitative results show that this is due to a more precise prediction of the background papyrus pixels containing pseudo-periodic patterns.

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**

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