Deep Learning for More Expressive Virtual Unwrapping

Learning Transformations from Tomography to Other Modalities

Stephen PARSONS, University of Kentucky, United States
Kristina GESSEL, University of Kentucky, United States
Clifford PARKER, University of Kentucky, United States
William SEALES, University of Kentucky, United States

Keywords: *Machine Learning — Virtual Unwrapping — X-Ray Microtomography — Text Recovery*

CHNT Reference: Parsons, S., Gessel, K., Parker, C. S. and Seales, W. B. 2020. Deep Learning for More Expressive Virtual Unwrapping. Proceedings of the 25th International Conference on Cultural Heritage and New Technologies 2020, Vienna.

1 Introduction

This paper presents the general use of deep learning for texturing[1] within the virtual unwrapping model. Virtual unwrapping is a software pipeline for the noninvasive recovery of texts inside damaged manuscripts or scrolls via the analysis of tomography (Seales et al., 2016) and consists of three stages (Fig. 1). Segmentation isolates pages or layers as surface meshes, texturing paints these surfaces based on the local neighborhood in the tomography, and flattening produces legible images from the folded, rolled or warped meshes. The pipeline allows for the recovery and restoration of a variety of otherwise lost or hidden heritage objects.

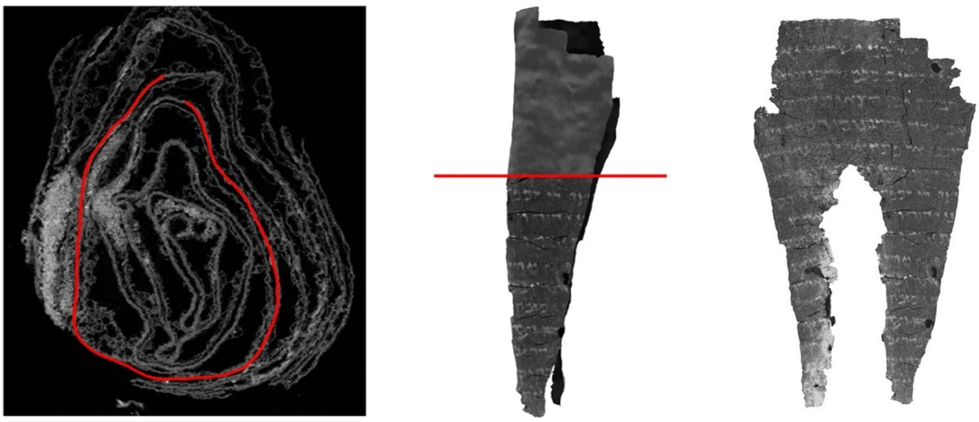

Fig. 1 Virtual unwrapping. (a) Segmentation isolates layers in the volume. (b) The result is a 3D mesh of the shape of a layer. This mesh is then "textured" based on some function of the tomography. (c) Flattening produces a 2D image.

The output of the virtual unwrapping pipeline is an image revealing the contents of a page or layer inside the scanned volume (Fig. 1c). Segmentation determines where each point on this surface originates in the volume, and flattening determines the layout of the final visualization. Texturing is the process responsible for rendering or painting the image itself onto the surface, guided by the original scan. It is critical that texturing reveal something of interest, as segmentation and flattening are of no use if the resulting image is featureless.

Traditionally, texturing methods have been based on simple functions designed to extract some information directly from the tomography. For example, Fig. 1c is obtained pixel by pixel by examining a neighborhood in the scan around a particular point of interest and selecting the maximum intensity observed inside that

[1] "Texturing" here means the application of a 2D image to the surface of a 3D mesh, as with "UV mapping" or "texture mapping," not "texture" as in the shape, consistency or feel of a surface.

neighborhood. This value is directly plotted on the output image. The image can be interpreted as a modified visualization of raw tomography data. What appears as a bright spot in the tomography likewise appears brightly in the image. Where this intensity difference corresponds with a desired signal, as above where the ink appears brighter in X-ray than the surrounding substrate, the image reveals the text.

It is not always the case, however, that the X-ray response readily provides useful texturing information. For example, carbon ink is not easily distinguished in tomography. Additionally, the scholar may not wish to see an image directly corresponding to the X-ray response for other reasons. It may be advantageous to look at a color image, as if the surface had been photographed. Parker et al. (2019) addressed these cases by creating neural networks to learn the carbon ink signal directly and to learn to recreate the page's color appearance.

This work further expands the generalization of the texturing stage. Neural networks can be trained to detect not only carbon inks, but any desired signal in tomography data. Additionally, they can output not only color images, but any modality or form desired by the scholar. The contributions of this paper are as follows:

- A conceptualization of texture mapping as a general function, perhaps learned, mapping tomography to any modality is established (Section 2).
- Several technical improvements to the previous neural network approach are discussed (Section 3).
- This framework is applied to new data, yielding state-of-the-art results (Section 4).

2 Expanding the Texturing Concept of Virtual Unwrapping

The heritage scientist and scholar presently have a wealth of tools at their disposal. For example, the pages of a manuscript could be photographed for the creation of a digital facsimile; imaged under multi-spectral lighting to reveal faded pigments; or undergo X-ray fluorescence (XRF) studies to analyze the ink composition.

Unfortunately, in the case of damaged or fragile materials, the use of these imaging tools is precluded for conservation reasons, and the heritage scientist is instead restricted to non-invasive tomographic imaging methods. This limitation is a significant constraint, primarily for the changes in dimensionality, visualization, and resolution. Manuscript pages look quite different when folded or warped and imaged in three dimensions than they do when photographed. Virtual unwrapping overcomes these limitations by extracting the layers from the volume and presenting them in a more familiar arrangement. But a second limitation must also be overcome, which involves the modality of the resulting images and the information they convey. Note that Fig. 1c clearly reveals the text but does not resemble any modality familiar to the manuscript scholar. Rather than a photograph or other image, the scholar is viewing essentially a projection of intensity values from X-ray.

One can imagine this is not the scholar's ideal visualization of the pages. The key principle of this work is that texturing in virtual unwrapping can be generalized to simulate any modality. Though the scholar is constrained to acquire tomographic images, in this case X-ray, the visualization of the resulting data can take on any form.

Machine learning provides the ability to learn virtually any transformation for which training data can be acquired. The training data is acquired using reference materials which have exposed surfaces. To continue the example above, consider a manuscript which cannot be opened but has exposed text on the top surface. This can be imaged using any traditional tools, and the entire manuscript can be scanned with tomography. A mapping can thus be learned between tomography and these modalities. The real benefit is when this mapping is then used to visualize the hidden pages as if they had been scanned using spectral photography, XRF, etc.

3 Technical Improvements to the Neural Network

The initial neural network was trained to output a value representing a single pixel of the output image (Parker et al., 2019). To create an output image, a subvolume corresponding to each individual pixel location was fetched and then fed to the trained network. The resulting output was stored one pixel at a time in the output image. This approach was a breakthrough for being the first to at all reveal carbon ink in tomography, but left room for additional improvement. First, the method is slow and scales poorly to the high-resolution images that are often desirable. Second, independently processing adjacent pixels ignores their spatial correlation. As a result, output images can appear noisy where they should be smooth. In this work the use of 2D neighborhoods as labels and outputs is explored. The neural network is trained to take a subvolume as input and output a label corresponding to the region of the output image spanned by that subvolume. The output images can be generated much more efficiently this way and are additionally smoother. This neural network can be considered a hybrid architecture similar to a 3D U-Net contracting path (Çiçek et al., 2016) with a 2D U-Net expanding path (Ronneberger, 2015). Finally, additional refinements to the performance at inference time have been developed, such as inference augmentation with prediction averaging and output region overlap with blending. Finally, alternative loss functions and their respective performance improvements have been explored.

4 Results on the Morgan M.910 Manuscript

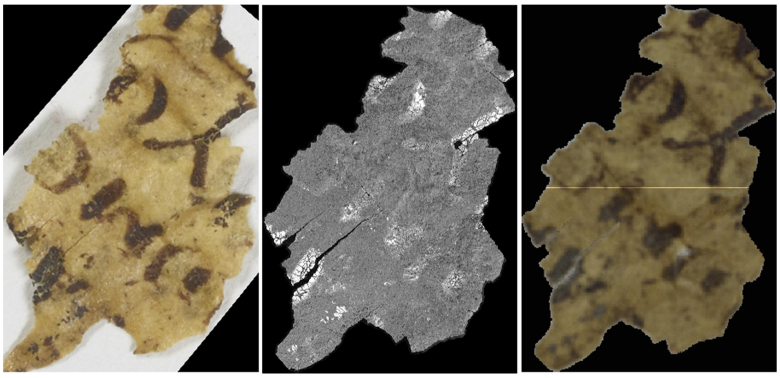

Fig. 2. Morgan M.910 Manuscript fragment. (a) Photograph. (b) Method from Seales et al., 2016. (c) Proposed method.

Fig. 2 shows the results of a neural network trained to reproduce an RGB photograph from tomography data. The image on the right, generated purely from X-ray tomography, highly resembles the real photograph (left). This is a much more natural way of visualizing the tomography compared with the previous method (center).

5 Conclusion

This work has presented a general approach to texturing within virtual unwrapping that allows the scholar to generate virtually any visualization for which there is training data. This framework allows virtual unwrapping to become a powerful lens for looking inside tomographic scans, magnifying or revealing previously buried signals. Along with technical refinements, these concepts are enabling state-of-the-art results on real data.

Ongoing and future work. There are ongoing efforts to improve the ability of the trained networks to generalize across various scans, so they can be trained on one set of data and then used to reveal the contents of entire scans for which the network has never seen labeled data. One initial approach to be tried is the use of autoencoder regularization for semi-supervised learning. These methods are also contingent on accurate segmentation. Other work focuses on leveraging deep learning to improve the segmentation performance.

Acknowledgements. The authors would like to thank Dr. Paul Dilley from the University of Iowa for the study of the revealed texts; Maria Fredericks, the Sherman Fairchild Head of the Thaw Conservation Center at the Morgan Library & Museum who made the scans possible and worked with us to ensure the manuscript's safety; and MicroPhotonics for the use of the micro-CT equipment to scan the manuscript.

References

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016, October. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.

Parker, C.S., Parsons, S., Bandy, J., Chapman, C., Coppens, F. and Seales, W.B., 2019. From invisibility to readability: Recovering the ink of Herculaneum. PloS one, 14(5), p.e0215775.

Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Seales, W.B., Parker, C.S., Segal, M., Tov, E., Shor, P. and Porath, Y., 2016. From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science advances, 2(9), p.e1601247.