EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2023)
A. Bucciero, B. Fanini, H. Graf, S. Pescarin, and S. Rizvic (Editors)

High Resolution 2D-/3D-Scanning and Deep Learning Segmentation for Digitization of Fragmented Wall Paintings

O. Kroeger[1], O. Krumpek[1], P. Koch[1], M. Pape[1], J. Schneider[3] and Prof. Dr.-Ing. J. Krueger[1,2]

[1]Fraunhofer Institute for Production Systems and Design Technology IPK, Germany
[2]Technische Universität Berlin, Germany
[3]MFB MusterFabrik Berlin GmbH, Germany

Abstract

The preservation and study of mural wall paintings often involve the collection of numerous fragments with unknown context. In this paper the authors present a case study involving a Roman wall painting discovered in 2013 at the European Cultural Park Bliesbruck-Reinheim. The objective of this work was to develop a semi-automated assistance system for the digitization, visualization, and digital repositioning of the Roman wall painting fragments. Therefore an easy-to-use scanner system was developed, that captures high-resolution 2D images of the front and back surfaces of the fragments, along with a height map of the backside. The contributing partners also developed a control and operating software for the scanner, as well as an automated software platform for visualization and repositioning of the digital fragments. The contributions of this paper include the introduction of a ML-based algorithm for background subtraction and segmentation of the front surface of the fragments. The technical realisation for fast and accurate image acquisition of the fragments, including sensor registration and high-resolution capture, has been worked out. The system calibration process, hardware setup and data correction techniques are described in detail. Additionally, the challenges of pixel-wise image segmentation for distinguishing between background, inner contour (wall painting), and outer contour (fragment surface without painting) are discussed. Our proposed approach overcomes the limitations of training ML algorithms on high-resolution images by employing patch-wise training and leveraging small features instead of large-scale features. The digitization and segmentation process demonstrated promising results in preserving and reconstructing the roman wall painting fragments. The findings of this study contribute to the field of cultural heritage preservation and provide valuable insights as the developed equipment and methods are highly transferable to future digitization projects.

CCS Concepts

• ***Applied computing*** → *Fine arts;* • ***Computing methodologies*** → *Image segmentation; 3D imaging; Camera calibration;*

1. Introduction

The study and preservation of scientific and cultural heritage usually begins with the collection of countless fragments of unknown context. As more and more pieces are uncovered, research often becomes a giant jigsaw puzzle whose unknown solution promises interesting insights into a bygone era. Since fragments of lost art objects can have the most diverse manifestations, the contributing partners and authors of this paper, dealt with a puzzle of about 12.000 sandstone pieces (see figure 1). Each piece being part of a Roman wall painting that was discovered during excavations at the European Cultural Park Bliesbruck-Reinheim in the year 2013.

The cultural and archaeological park, located at the border between Germany and France, with its Roman villa of Reinheim and the Roman small town (vicus) of Bliesbruck represents not only regional history but also a part of Europe's history. The fragments, discovered in the area of the Roman villa are very well preserved. However, reliable archaeological analyses, such as determining the dating and origin of the fragments as well as understanding the social-historical background, can only be conducted after the wall paintings have been reconstructed. Archaeologists and restorers face the challenge that any manual handling of the fragile fragments, which consist of lime mortar, leads to additional damage to the fragments themselves [Saa].

In 2019 a research community was formed to tackle the challenges of digital archiving. The research partnership consisted of the Landesdenkmalamt Saarland, the MFB (MusterFabrik Berlin GmbH), the Fraunhofer IPK, as well as the European Cultural Park Bliesbruck-Reinheim. The team, leveraging the many years of experience of the IPK in the industrial application of optical inspection and the MusterFabrik Berlin's expertise in algorithmic processing of fragmented glass mosaics, aimed to create a digital twin of the puzzle for further algorithmic and human-guided processing. The joint project work, carried out under the name "DigiGlue",

DOI: 10.2312/gch.20231152

comprised the development of an automated IT assistance system for the digitization, visualization and digital repositioning of the Roman wall painting fragments. The final system can be seen in figure 2 at its destination, near the excavation site.

Figure 1: *Original fragments of the Roman wall paintings discovered in the European Cultural Park Bliesbruck-Reinheim © Landesdenkmalamt Saarland. The image clearly shows that the size of the fragments varies greatly. The sandy and friable structure of the material is also visible.*

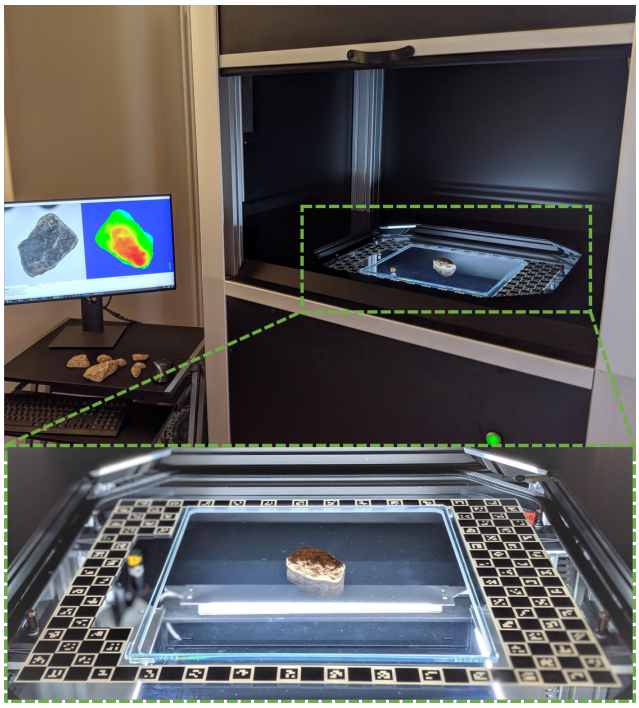

Figure 2: *Photograph of the final digitization equipment (Scanner), set up in the rooms of the Bliesbruck-Reinheim site. The screen on the left shows the result of a single fragment acquisition. The fragment is placed on the glass plate of the scanning area with the flat side down.*

Within the project, the Fraunhofer IPK was responsible for the development of the core technologies for an easy-to-use scanner system that simultaneously captures the front and back sides of the mural fragments, segment the outer- and inner- contour on the front surface (see figure 3) as well as a height map of the backside and the implementation of a pixel-wise data fusion of the 2D front surface to the height map of the back surface. The objective of Muster-Fabrik Berlin (MFB) was the development of a software platform including UI and several tools for persistent data storage (including meta data) as well as handling and positioning of the digitised fragments based on a semi-automatic 2D reconstruction approach. The MFB was also responsible for parts of the technical specifications and the assembly of electrical and mechanical components in the later deployed system.

The algorithm used for the automatic reconstruction approach was mainly based on the extraction of 2D features within the flat area on the front side where the image is visible. However, the 3D topography of the backside was captured to validate the positioning results. As validation parameters, we assumed continuity in the prints of significant architectural features, traces of straw, wooden beams and other structural support elements. It should be noted that the positioning algorithm is not part of the methods and developments presented here. It should also be noted that the selection of the camera systems used was based on specific requirements for image quality and resolution, as set out, for example, in the German Research Foundation's guidelines for digitisation. [ABB*23] In particular, high quality lenses with low distortion and sensors that allow a resolution of at least 400 ppi (pixels per inch) were chosen. At the same time, the project partners expected the size of the recording area dimensions to be at least 30×30 cm in width and height to allow the simultaneous recording of multiple fragments or single, large sized fragments.

In the project context we described above, we made the following contributions to the digitization of cultural heritage:

- Introducing new technologies for easy and fast 2D image acquisition of the front and 3D acquisition of the back of mural fragments in a fast one shot process. The 2D images of the two sides are colour calibrated and have very high-resolution. The captured data is also stored in a memory-efficient way, as 3D data points are only captured from the relevant perspective. The fragments only need to be placed in the scanner once. Depending on the size and texture of the fragments, the entire image acquisition takes approximately 60 seconds for each fragment.
- The entire scanner system is mobile and can be set up near an excavation site and operated there by non-experts. The hardware is relatively inexpensive, even at the very high resolution.
- An ML-based algorithm for background subtraction and segmentation of the inner and outer contour of the front surface of the fragments capable of efficiently handling the very high image resolutions.
- Our unique scanning process and ML-based segmentation allows multiple objects to be placed and digitised simultaneously within the large scanning area of our scan.
- A special laser line helps the user to label the motif side of the fragment, contributing to fast learning of the segmentation.

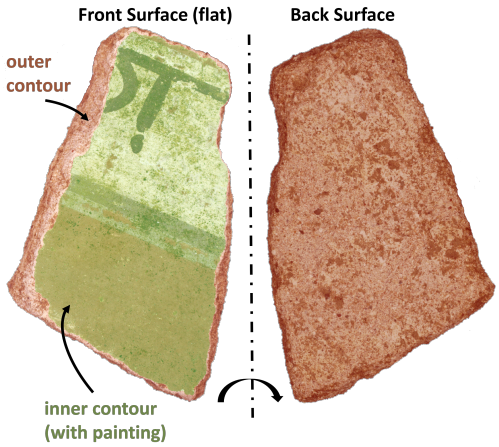

Figure 3: *Example of a digitized fragment. The flat/planar front surface of the wall painting (left) has an inner contour with the actual painting (here coloured in transparent green) and an outer contour (here coloured in transparent red). The back surface (right) has only an outer contour (also coloured in transparent red).*

2. Related Work

Brown B. J. et al. [BTFN*08] present an acquisition system for the digitization of excavated fragments at Akrotiri, Santorini (Thera). The fragments pass sequentially, but manually, through an acquisition workflow using flatbed scanners and two 3D scanners to capture the front and back surfaces. The software and a processing pipeline then ensure that the fragments are automatically merged from the single images. This results in a complete 3D model of the fragments. The entire system can easily be operated by non-experts, which was the goal of the project. The throughput is about 10 fragments per hour or 6 min per fragment.

In the field of cultural preservation, both classical computer vision approaches and machine learning-based approaches are successfully employed. Projects like the Fraunhofer CultArm3D use robot based digitization with high-resolution cameras like the PhaseOne up to 100 MP [HvWLW20]. The system was already applied for the 3D mass digitization in the year 2017 [SRFF17]. Additionally, the combination of robotics and artificial intelligence has shown promising results for the reconstruction of cultural heritage, as demonstrated by the EU-funded project RePAIR. This project focuses on the reconstruction of ancient artworks in the destroyed city of Pompeii, using shape, 3D information and decoration to find relationships between individual pieces, and later employing soft robotics for actual manipulation [Rsi21]. In contrast to our proposed work, the RePAIR project also focuses on the heavily laborious task of physical reconstruction. Furthermore, especially for full 3D reconstruction of archaeological finds or museum archives, imaging systems based on changing light conditions, such as the multi-light reflectance approach, are used for digitisation. Often based on photometric stereo or reflectance transformation imaging, systems including physical dome light setups [HH23] or portable, robot-guided systems such as the LightBot [LCC*22] can be used directly at the excavation site.

In contrast to the aforementioned systems and many of the corresponding use cases, the concept of the presented system does not include a complete 3D reconstruction, since the goal of the image reconstruction can essentially be described by the interpretation and processing of the 2D information of the flat object surface.

Data driven ML heuristics have dominated SOTA segmentation-benchmarks [GLU12,ZZP*17,LMB*14,COR*16,SRT*16] within pixel-wise image segmentation in recent years. However, training neural networks on high-resolution images remains challenging due to the substantial hardware resource requirements (GPU-Memory). This is particularly evident for powerful transformer based NN architectures, or big CNNs. Lin T. et al. [LHL*21] have addressed this issue by scaling Transformer-based image processing towards high-resolution images of 1536^2 pixel. However, the image resolution used in this project, 9568×6376 pixels, still exceeds that (3.87%) by a significant margin. With respect to convolutions models, method such as deformable convolution [DQX*17], stride, dialation can be used to increase the convolutions receptive field (kernel size), without increasing the number of model parameters. This allows for a reduction in the input space while extracting long-range features. Likewise, pooling layers are often used to decrease the latent space drastically. However, this contradicts the reason for high-resolution images. Early work on ML-based pixel-wise segmentation of medical images [RFB15] uses a patch wise training, where small and trackable image patches are sampled from the high-resolution input space. During inference, the trained model could be executed on the original image size on a single GPU without running out of memory, thanks to the missing gradient tracking for backpropagation and the scale invariant feature extraction of CNNs. Recent work has uses a two-stage approach where a first stage selects (attends to) high-resolution patches based on a low-resolution input [KF19]. Thus, only a fraction of the high-resolution data is processed. While this approach can crop out background information, it comes at the cost of an increased processing time due to sequential computation.

3. Methods

In the following we present our methods. We introduce the methods used for calibration and go into detail about the hardware of our system. Afterwards we present our methods for data post-processing, and pixel wise image segmentation. Figure 4 gives a general overview of the process flow with our scanner.

3.1. System Calibration and Sensor Registration

The complete system used for digitizing the stone fragments consisted of a heterogeneous multi-camera setup as shown in figure 6. The setup contained two 61 MP RGB cameras (Sony α7R IV 35 mm full-frame camera (ILCE-7RM4)) and an active stereo system. The RGB cameras were placed at the top and bottom of the system, aiming for centred views with collinear optical axes, orthogonal to a centered glass plate and the flat surface of the fragment, respectively. The stereo system was positioned close to the top RGB camera. To achieve the most useful visualization of the recorded data (equally sized and symmetrically positioned representations see figure 3 and the left side of figure 5), a pixel-wise registration

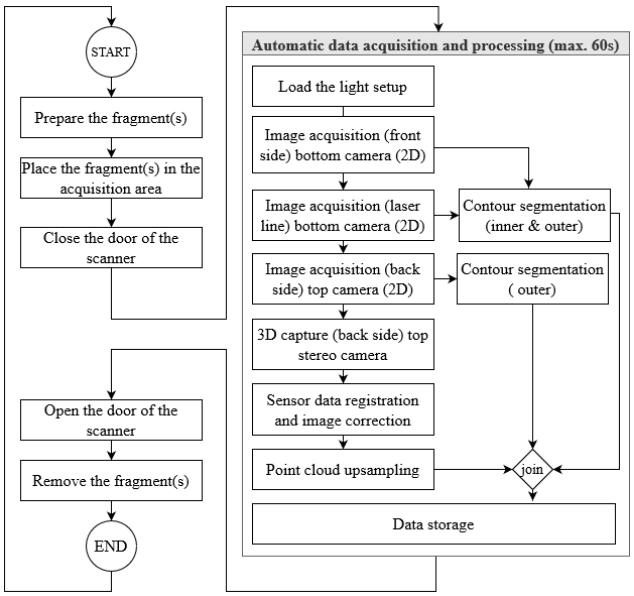

Figure 4: *Workflow of a single acquisition process. The left side of the diagram describes the user interaction, while the right side of the diagram describes the automated process steps. The complete process for a single capture is completed in less than 60 s.*

for all three sensors was performed by accounting the differences in perspective, resolution and scale. The full procedure can be described as two main parts. Firstly, the image data of the top RGB camera and the 3D data from the stereo system needed to be registered, resulting in a pixel-wise overlay of 3D information on the image plane of the RGB camera. Secondly, the RGB image of the camera in the bottom had to be transformed to fit the target and image position (see figure 5). For both RGB cameras, we also used a calibration target to calibrate the white balance of all processed images. Another, two sided color calibration target was persistently placed in the visible area of both cameras to allow possible post-processing correction steps.

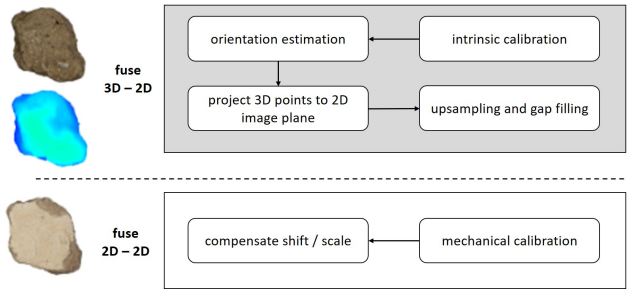

Figure 5: *Illustration of the calibration pipeline used for image and sensor registration. The left side of the figure shows the pixel-wise registered results of a single fragment, as used in the program's user interface. The user is able to switch between the plots without changing the size or outline of the visualisation.*

3.2. Hardware Setup

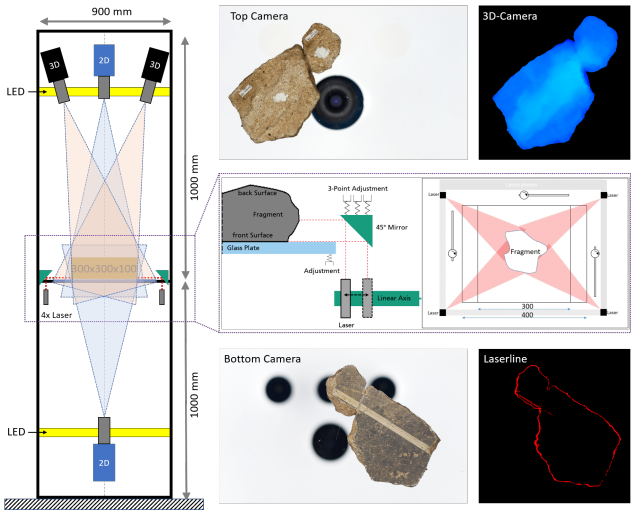

Figure 6: *Illustration of the scanner setup with cameras, illumination, laser and glass plate with object volume in the centre. The illustration also includes visual information about the installed laser system, used to highlight the relevant area of the scene, including deflection mirrors at the four corners of the glass plate.*

Initially, the calibration of the top RGB camera was performed using a standard calibration routine to estimate the intrinsic parameters [Zha00]. The relative orientation of the stereo system and the calibrated camera was determined by exploiting the fact that pixel-wise correspondences between the rectified image of the left stereo camera and the projected 3D points were available due to the sensor principle. ChAruco markers were detected to achieve a set of 2D correspondences C that could be found in the set of detected markers A and B of both cameras.

For each given corresponding image point pair in C, the 2D-3D correspondence of the stereo system could be used to form a Perspective-n-Point (PnP) problem, resulting in the simple relation $sp_{rgb} = [R|T]Kp_{st}$ between the 3D points of the stereo system, the corresponding RGB image points, the estimated intrinsic parameters and the desired orientation. With the data from the upper sensors aligned, the registration of the merged data and the second RGB camera was easily achieved by compensating for the differences in scale, rotation and shift caused by the geometry of the setup and the inaccuracy of the mechanical calibration step (see figure 6). As a first step a Hough circle detector was used to estimate the scale factor between the projected level of the glass plate in both images. The scale factor was given by the ratio of the average diameter of detected circles on a calibration pattern, placed on the glass surface. Shift and rotation were also estimated using another calibration pattern, permanently glued to the glass plate and visible to both RGB cameras (see figure 6).

3.3. Data Correction and Post-Processing

After transforming the point cloud into the coordinate system of the RGB camera, the 3D points were projected onto the image

plane, with the pixel value representing the z-component of the projected point. Due to the difference resolution between the sensors (9568×6376 to 1280×1024) and the absence of 3D information, the coverage of the resulting depth map was less than 2% of the available pixels. This resulted in an uneven distribution of the projected points in the image. An example is shown in figure 7, where the area around the available depth information has been extended for better visibility.

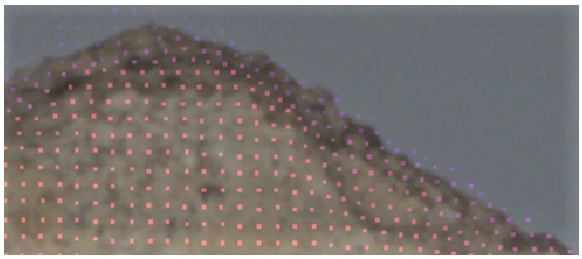

Figure 7: *Illustration of the unevenly distributed sparse depth information after projection to the image frame of the top RGB camera. The illustration describes the reason for the upsampling of the depth information.*

To upsample the depth information, a 3D reconstruction algorithm was used to create a surface model of the sensor's raw point map, again using the given neighbourhood information for the 3D data [HB13]. Additionally a random point sampling was performed to generate the missing depth information [Vit84].

3.4. Pixel Wise Image Segmentation

A training data set plays a crucial role in the development and training of any heuristic ML-based image segmentation algorithm. Here related work commonly drives on large scale annotated image data sets ([COR*16] 5k high quality pixel-wise annotated frames). However, due to the nature of our segmentation problem, we do not require large scale feature extraction (cars, trees) but rather small features (edges and texture), which helps to distinguish between A) Background, B) inner contour (wall painting), and C) outer contour (fragment surface without wall painting) (see figure 3). Here, large scale features (e.g. a given wall painting) enables the heurisitic algorithm to overfit on its training data rather then learning to generalize the problem. Likewise to cancer cell segmentation, our segmentation problem is based on high-resolution images with relatively small features. Therefore patch wise training is conducted in related work on medical imaging [RFB15], which helps address the hardware requirements of high-resolution images and prevents overfitting caused by excessive extraction of large-scale features.

Taking this into consideration, the fragments are digitized (see figure 6) following the outlined procedure in section 3.1. Hereby we gain three in pose varying sets of high-resolution images from 79 fragments front- and back surface (front surface has wall painting, see figure 3). With a complementary set without any fragment present in the images, we can use background subtraction in order to find the outer contour of each fragment (in figure 8 we outline the background subtraction algorithm). Thereby, we generate

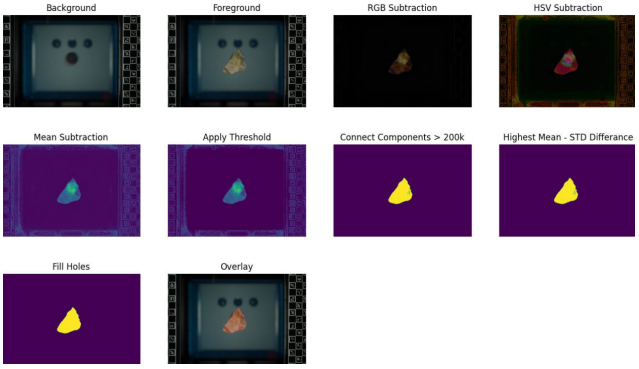

Figure 8: *Outline of the background subtraction algorithm. We use connected component analysis to find objects $>$ 200k pixels and select the component with the largest mean absolute difference.*

a segmentation mask for each fragment. However, this segmentation mask alone can not distinguish between a wall painting and a pure fragment texture that does not contain any information about the wall painting. We use an open source annotation tool to mask the wall paintings in each front surface fragment image by hand. In order to cope with the large images, we crop the images via the mask generated by our background subtraction. This cropping enables us to isolate the fragment surface outer contour (without wall painting, see figure 3) by subtracting the fragment mask generated by the background subtraction. A major annotation problem arises when the boundary between the mural and the fragment is blurred, making it difficult to draw a clear annotation. Likewise the wall painting masking accuracy is crucial for the training as high recall is important. In doubt fragments are rather annotated as wall painting to ensure a high recall over precision (see figure 9).

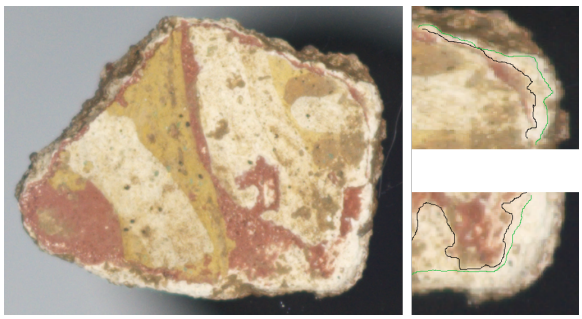

Figure 9: *Visualisation of the inner contour (wall painting) annotation. Black line is a precise annotation for high precision training. The green line is more conservative annotation which yields a high recall but comes with a cost of lower precision. High precision annotation is often not possible due to gradual shifts between the painting and fragment or leaps between two patches of wall painting where any annotation becomes ambiguous.*

With the annotated high-resolution training data, we can now conduct patch-wise training. State-of-the-art methods for pixel-wise segmentation models employ transformer architec-

tures [XWY*21] or deformable CNNs [WDC*23] for complex feature extraction and segmentation. However, our problem at hand neither requires big nor complex features. In addition, a fast runtime of a few seconds is required for smooth operation. Therefore, we decide to use a simple UNet [RFB15] with a ResNet 32 [HZRS15] backbone pretrained on ImageNet [DDS*09] for pixel-wise image segmentation (see figure 10).

Due to the scale-invariant feature extraction withing the ResNet CNN architecture, patch-wise training can be performed while processing the complete high-resolution image in a single step during inference. This allows fast computation within the given 6 GB GPU RAM hardware limitation of the target system. At training time we draw 512×512 pixel patches from the high-resolution images (see figure 12). Here we focus to evenly draw samples centered on pixels belonging to label B and C (wall painting and fragment). 10% of the samples are selected at random, including background only samples. This ensures that the model can handle the complete input, including background, during inference. Due to this ability we can disable e.g. the background subtraction pre-processing step for RoI (Region of Interest) crops, which during longtime deployment is unstable. Also a single forward pass is relatively fast within the GPU. We use a subset of 80% of our data for training, while the remaining 20% is used to validate the model performance after training (the validation data consists equally of unseen images of known fragments, and unseen images of unknown fragments). The model is trained using the Jaccard loss and Adam optimization 200 epochs with a learning rate of $10^{-}3$ and batch size 64. During each epoch we draw 20 samples from each fragment included in the training data (2k samples per fragment in total).

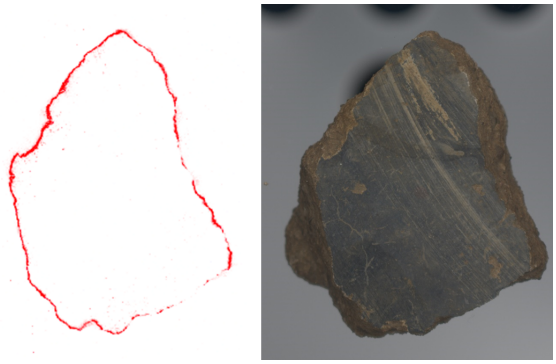

Figure 11: *Visualisation of the laser line on top of the RGB camera. The right image shows the raw RGB data, the left image shows the result of the laser line detection on the edge of the inner contour. The data can be used as additional guidance for the annotation step or/and for the segmentation algorithm.*

We report the mean intersection over union (mIoU) and training results in the following section.

The laser line image included in our digitisation process for the front side of the fragment holds additional information about the fragment's inner rim. Within the red channel of this RGB image, a laser light is reflected where the fragment is elevated above the glass plate. These elevated areas, highlighted in the red color channel, do not belong to the wall painting. We experiment to concatenate this red channel as a 4^{th} input channel to the usual front side RGB image where the complete wall painting information is present. This additional information is intended to further guide the segmentation process. However, reflecting nature is imprecise and unstable and cannot be used alone. A visualisation of the laser line as green overlay on top of the RGB image can be seen in figure 11. We experimented with the use of the laser line for annotation and/or for the segmentation process.

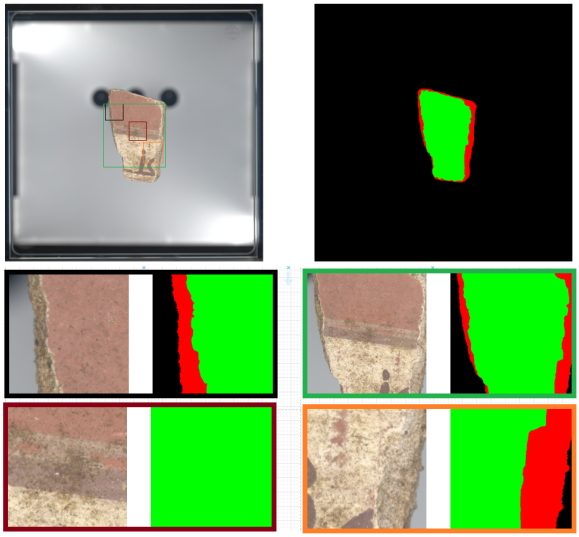

Figure 12: *Draw train patches the original high-resolution image (top left) and the pixel wise annotation (top right) and resize to* 512×512 *pixel.*

4. Results

One of the results of the "DigiGlue" project is an automated scanner capable of simultaneously capturing the front and back of wall fragments while creating an elevation map. This scanner is part of a novel system designed for the digital recovery of approx. 12.000 Roman wall painting fragments found in the European Cultural Park Bliesbruck-Reinheim. Another part of the system consists of an automated repositioning software platform which is based on technologies developed by MusterFabrik Berlin for material-preserving, non-contact scientific and restoration processing of highly sensitive cultural assets. This platform includes several tools for the visualization and repositioning of the fragment images captured by the scanner. For this purpose, the scanned fragments and other meta-information are stored in a database. The digital fragments can then be processed on digital workstations as part of the software platform according to various motifs and contour characteristics. These functions include grouping, aligning, measuring, labelling, sorting and - if the pieces fit - digitally gluing the fragments together.

The novel "DigiGlue" system was set up by the project partners in the fall of 2021 in an exhibition building, a replica of a Roman

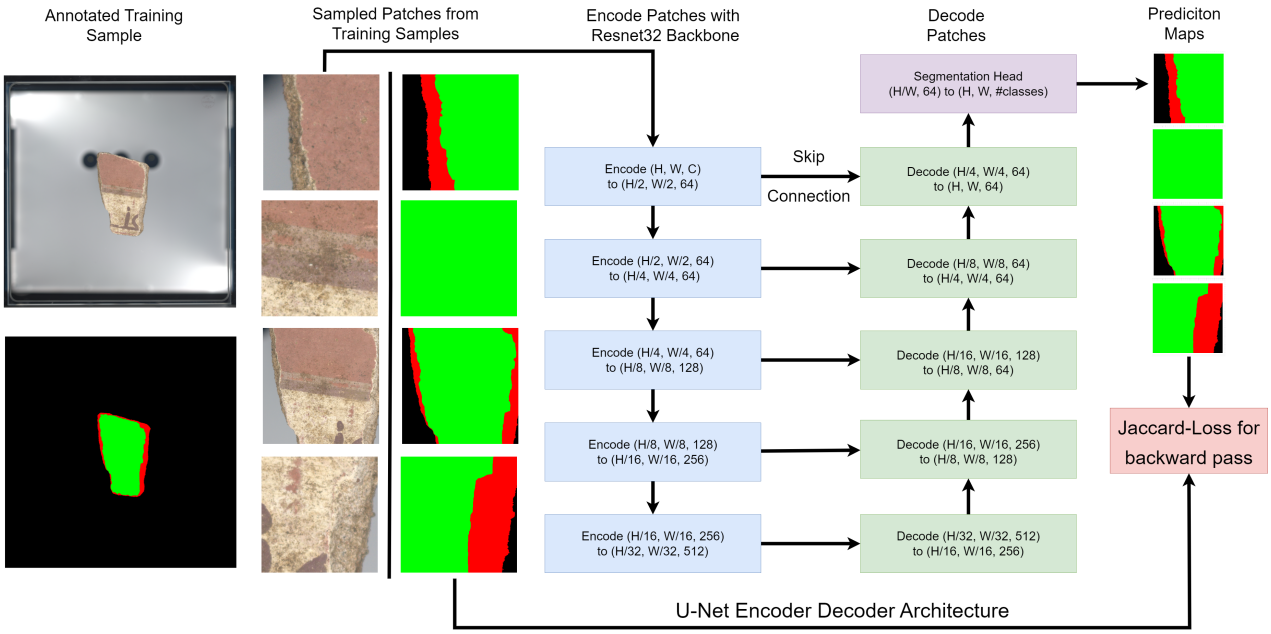

Figure 10: *Pipeline to train the U-Net Architecture with a ResNet32 encoder backbone (blue tiles) with patches drawn from the original high resolution training samples. The segmentation head (purple tile) projects generated segmentation maps to the 3 output classes. The green tiles indicate the decoder path of the U-Net architecture. The input (H, W, C) denotes the input patch Height, Width, and number of channels.*

tavern, of the European Cultural Park Bliesbruck-Reinheim (EKP). Since then the system has been operated under non-lab conditions and independently by the EKP staff on selected days in the summer months (the park is usually closed in winter). At present, the EKP is focusing on digitizing the fragile Roman wall fragments in a manner that ensures material preservation.

For digitization, the fragments are placed with the motif side down on a scratch-resistant glass plate mounted in the scanner (as already shown in figures 2). Acquisition is initiated by the scanner's control software. In an automatic process (see also the flow chart in figure 4), three images are captured in succession: a 2D colour image of the front and back of the fragment, and a 3D scan of the back of the fragment. After image acquisition, the three scans are automatically converted into a single digital representation by the control software. To do this, the 2D raw scans are first rectified and the fragments are extracted from the raw scans according to their smallest surrounding rectangle. Then the fragment and motif areas are segmented pixel wise to generate a mask image. In parallel, a height map is calculated from the 3D scan of the backside, which is finally scaled to the native resolution of the 2D scans of nearly 400 ppi. The entire acquisition process for one fragment takes just under one minute, including post-processing and image storage. The fragments do not have to be moved during the scanning process. In addition to the pure image acquisition, meta-information about the found objects can also be recorded and stored in the database of the assistance system. This can include information about the location or situation of the find, as well as information about the classification or content description of the objects to be digitised.

Currently, just over 2.900 fragments have been digitized (as of April 2023). Since this is only about a quarter of the expected total, no targeted repositioning or reconstruction work is currently taking place. However, despite the incompleteness of the digitized fragments, various partial reconstructions have already been identified and digitally glued together. The scanner is designed to be mobile so that the "DigiGlue" system can also be used at other excavation sites in the future. After dismounting the cameras and other hardware components, the scanner can be disassembled into two parts, transported and reassembled with reasonable effort.

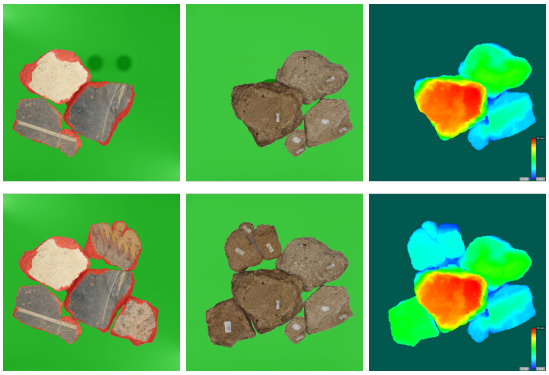

Figure 13: *Final results for the inner and outer contour segmentation of the front surface (left), the segmentation of the back surface (middle), and the 3D information of the back surface (right).*

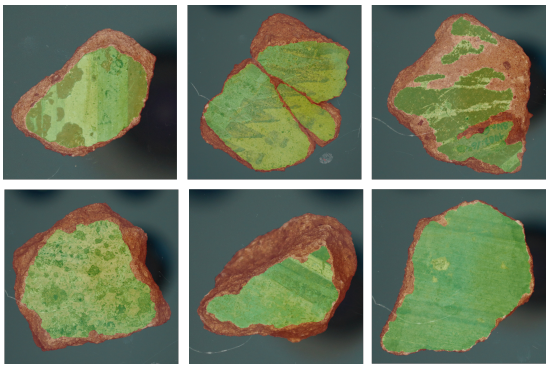

Figure 14: *Exemplar pixel wise segmentation validation results. The inner contour (wall painting) is highlighted in green, while the outer contour (raw fragment) is highlighted in red. The fragments are cropped after segmentation for better visualisation.*

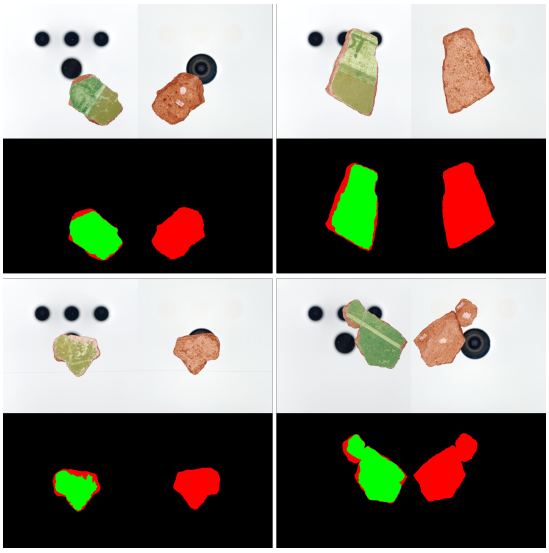

Figure 15: *Pixel wise segmentation validation results. The raw pixel wise segmentation results are overlayed on top of the input images and presented on black background.*

4.1. Pixel Wise Segmentation Results

In figure 16 we present the training loss of the Pixel Wise Segmentation. Here we report both the training performance on the randomly sampled 512×512 pixel patches (see section 3.4), as well as validation performance on full resolution images. The validation data is 20% of the available data and consists equally of unseen images of known (trained) fragments and unseen images of unknown fragments. Respectively we achieve a maximum train and validation average IoU of $0,970$ and $0,975$. When directly comparing the train and validation performance, one can see that the model is capable to generalize the pixel wise segmentation problem from small patches towards the full resolution input. Note that the IoU metric balances between classes, so the large amount of background during validation does not affect the score as it would with an accuracy or precision metric. Moreover, 50% of the validation data are images of unknown (trained) fragments. This allows the model to generalise to other fragments rather than remembering the training patch structures presented. In figure 15 and figure 14 we visualize some validation exemplars for qualitative measure. In figure 15 it is also noticeable that the model is not confused by the large area of background, while maintaining high accuracy as shown in figure 14. In figure 13 we show a test instance of our complete pipeline, including the segmentation on the left. Here one can see that the segmentation model is even capable to process multiple fragments at once.

5. Conclusion and Outlook

In this contribution we presented our system for the easy and fast digitisation of wall paintings. The fragments are acquired from both sides at the same time and in highest image quality. The back sides of the fragments are also scanned in 3D and fused with the 2D images. All sensors are registered to each other and the system can be used by non-experts. After the image acquisition, the relevant image information is automatically segmented. The outer contour of the front and back sides and the inner contour of the front side of the fragments are segmented. This will greatly assist the subsequent process of digitally and physically reconstructing the wall paintings. The segmentation is done by our high-performance ML pipeline, which is able to process image data in a high-resolution in a very short time. To further support the training of our ML pipeline with relatively few examples, we also used a laser line, which highlights the inner contour particularly well.

Once the fragments have been digitised by the system described here, a specially developed software tool is used to assemble the virtual mural puzzle.

Acknowledgements

The development of the “DigiGlue” scanner was realized together with MFB MusterFabrik Berlin GmbH and we thank them for their collaboration.

References

[ABB*23] Altenhöner R., Berger A., Bracht C., Klimpel P., Meyer S., Neuburger A., Stäcker T., Stein R.: *DFG-Praxisregeln "Digitalisierung". Aktualisierte Fassung 2022*. Zenodo, 2023. URL: https://zenodo.org/record/7435724, doi:10.5281/zenodo.7435724. 2

[BTFN*08] Brown B. J., Toler-Franklin C., Nehab D., Burns M., Dobkin D., Vlachopoulos A., Doumas C., Rusinkiewicz S., Weyrich T.: A system for high-volume acquisition and matching of fresco fragments: Reassembling theran wall paintings. *ACM Trans. Graph. 27*, 3 (aug 2008), 1–9. URL: https://doi.org/10.1145/1360612.1360683, doi:10.1145/1360612.1360683. 3

[COR*16] Cordts M., Omran M., Ramos S., Rehfeld T., Enzweiler M., Benenson R., Franke U., Roth S., Schiele B.: The cityscapes dataset for semantic urban scene understanding. *CoRR abs/1604.01685* (2016). URL: http://arxiv.org/abs/1604.01685, arXiv:1604.01685. 3, 5

[DDS*09] Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.: Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (2009), Ieee, pp. 248–255. 6

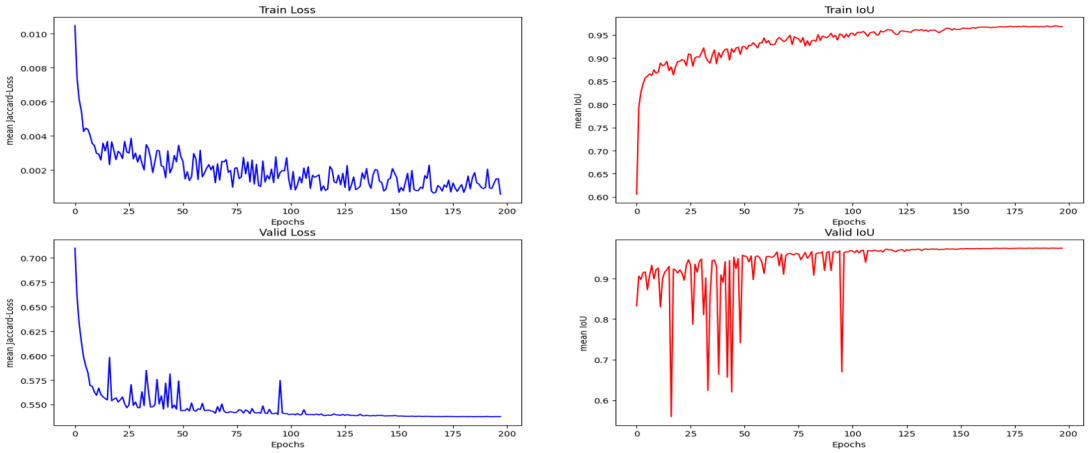

Figure 16: *Training logs of the pixel wise segmentation. Training is done with* 512×512 *pixel patches, while validation was conducted on full resolution images. Max. train and validation mean IoU is* $0,970$ *and* $0,975$*, respectively. Thus, the model architect can generalize from low resolution training patches towards high resolution inputs during inference.*

[DQX*17] DAI J., QI H., XIONG Y., LI Y., ZHANG G., HU H., WEI Y.: Deformable convolutional networks. *CoRR abs/1703.06211* (2017). URL: http://arxiv.org/abs/1703.06211, arXiv:1703.06211. 3

[GLU12] GEIGER A., LENZ P., URTASUN R.: Are we ready for autonomous driving? the kitti vision benchmark suite. pp. 3354–3361. doi:10.1109/CVPR.2012.6248074. 3

[HB13] HOLZ D., BEHNKE S.: Fast range image segmentation and smoothing using approximate surface reconstruction and region growing. In *Advances in Intelligent Systems and Computing* (Berlin, Heidelberg, 2013), vol. 194 of *Advances in Intelligent Systems and Computing*, Springer Berlin Heidelberg, pp. 61–73. 5

[HH23] H. HAMEEUW A. VAN DER PERRE V. B.: *Interactive 2D and Multispectral Imaging on the Crossroads of Archaeology, Egyptology and Assyriology*. 2023, ch. 3, pp. 72–111. doi:10.1163/9789004527119_005. 3

[HvWLW20] HARDY H., VAN WALSUM M., LIVERMORE L., WALTON S.: Research and development in robotics with potential to automate handling of biological collections. *Research Ideas and Outcomes 6* (2020), 213–41. 3

[HZRS15] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. *CoRR abs/1512.03385* (2015). URL: http://arxiv.org/abs/1512.03385, arXiv:1512.03385. 6

[KF19] KATHAROPOULOS A., FLEURET F.: Processing megapixel images with deep attention-sampling models. *CoRR abs/1905.03711* (2019). URL: http://arxiv.org/abs/1905.03711, arXiv:1905.03711. 3

[LCC*22] LUXMAN R., CASTRO Y. E., CHATOUX H., NURIT M., SIATOU A., LE GOÏC G., BRAMBILLA L., DEGRIGNY C., MARZANI F., MANSOURI A.: Lightbot: A multi-light position robotic acquisition system for adaptive capturing of cultural heritage surfaces. *Journal of Imaging 8*, 5 (2022), 134. URL: https://www.mdpi.com/2313-433X/8/5/134, doi:10.3390/jimaging8050134. 3

[LHL*21] LIU Z., HU H., LIN Y., YAO Z., XIE Z., WEI Y., NING J., CAO Y., ZHANG Z., DONG L., WEI F., GUO B.: Swin transformer V2: scaling up capacity and resolution. *CoRR abs/2111.09883* (2021). URL: https://arxiv.org/abs/2111.09883, arXiv:2111.09883. 3

[LMB*14] LIN T., MAIRE M., BELONGIE S. J., BOURDEV L. D., GIRSHICK R. B., HAYS J., PERONA P., RAMANAN D., DOLLÁR P., ZITNICK C. L.: Microsoft COCO: common objects in context. *CoRR abs/1405.0312* (2014). URL: http://arxiv.org/abs/1405.0312, arXiv:1405.0312. 3

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. *CoRR abs/1505.04597* (2015). URL: http://arxiv.org/abs/1505.04597, arXiv:1505.04597. 3, 5, 6

[Rsi21] RSIPVISION I.: Computer vision news 2021 october, 2021. https://www.rsipvision.com/ComputerVisionNews-2021October/22/ [Accessed: (30.05.2023)]. 3

[Saa] Saarbruecker-zeitung_01juli2022_das römische puzzle in bliesbruck-reinheim. 1

[SRFF17] SANTOS P., RITZ M., FUHRMANN C., FELLNER D.: 3d mass digitization: a milestone for archeological documentation. *Virtual Archaeology Review 8*, 16 (2017), 1. doi:10.4995/var.2017.6321. 3

[SRT*16] SIRINUKUNWATTANA K., RAZA S. E. A., TSANG Y.-W., SNEAD D. R. J., CREE I. A., RAJPOOT N. M.: Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging 35 5* (2016), 1196–1206. 3

[Vit84] VITTER J.: Faster methods for random sampling. *Communications of the ACM 27*, 7 (1984), 703–718. 5

[WDC*23] WANG W., DAI J., CHEN Z., HUANG Z., LI Z., ZHU X., HU X., LU T., LU L., LI H., WANG X., QIAO Y.: Internimage: Exploring large-scale vision foundation models with deformable convolutions, 2023. arXiv:2211.05778. 6

[XWY*21] XIE E., WANG W., YU Z., ANANDKUMAR A., ALVAREZ J. M., LUO P.: Segformer: Simple and efficient design for semantic segmentation with transformers. *CoRR abs/2105.15203* (2021). URL: https://arxiv.org/abs/2105.15203, arXiv:2105.15203. 6

[Zha00] ZHANG Z.: A flexible new technique for camera calibration. *IEEE transactions on pattern analysis and machine intelligence 22*, 11 (2000), 1330–1334. 4

[ZZP*17] ZHOU B., ZHAO H., PUIG X., FIDLER S., BARRIUSO A., TORRALBA A.: Scene parsing through ade20k dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 5122–5130. doi:10.1109/CVPR.2017.544. 3