Semantic Motif Segmentation of Archaeological Fresco Fragments

Aref Enayati* Luca Palmieri* Sebastiano Vascon Marcello Pelillo Sinem Aslan
DAIS, Ca' Foscari University of Venice, Italy
882341@stud.unive.it, {luca.palmieri, sebastiano.vascon, pelillo, sinem.aslan}@unive.it

Abstract

Archaeological fragment processing is crucial to support the analysis of pictorial contents of broken artifacts. In this paper, we focus on the unexplored task of semantic segmentation of fresco fragments. This task enables the extraction of semantic information from a fragment, facilitating subsequent tasks like fragment classification or reassembly. We introduce a semantic segmentation dataset of fresco fragments acquired at the Pompeii Archeological Site, accompanied by baseline models. Additionally, we introduce a supplementary task of fragment cleaning, providing a dataset with the detection of manual annotations of archaeological marks that require restoration before further analysis. Our experiments, using standard metrics and state-of-the-art baselines, demonstrate that semantic segmentation of fresco fragments is feasible, paving the way toward more complex activities that require a semantic understanding of fragmented artifacts. Dataset with annotations, and code will be released at `https://repairproject.github.io/fragment-restoration/`

1. Introduction

Archaeological fragment analysis plays a crucial role in understanding and reconstructing ancient artifacts that have been damaged or fragmented over time. Among various artifacts, ancient wall paintings, a.k.a., frescoes, emerge as a rich source of pictorial content. They often display architectural elements, depictions of humans and animals, landscapes, and mythological scenes, as well as floral and geometrical patterns, creating recurring visual motifs. Archaeologists rely on these essential components for the reconstruction and understanding of these fragmented artworks.

This paper addresses the semantic segmentation of images of fragments from two ceiling paintings found at the Pompeii Archaeological Site, which have been damaged by the eruption of 79 AD and World War II bombings. In particular, we tackle semantic fragment segmentation through two distinct scenarios.

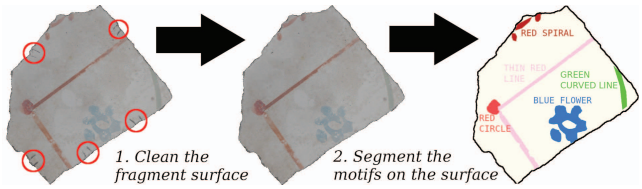

Figure 1. The proposed pipeline introduces a new dataset and the two baselines for the following tasks: cleaning of the fragment surface (from manual annotations) and segmenting the motifs to extract a clear representation of the pictorial content of the fragment.

In *Scenario 1*, we explored a 3-class semantic segmentation approach to differentiate between the *image background*, the unadorned fragment region (which we call as *fragment background*), and the *motifs* present on the fragment surface. This prepares the fragments for reassembly and further analysis of pictorial content, identified through the segmented motif class. The successful distinction of these diverse regions enhances the creation of a clean, high-quality dataset, forming a solid basis for subsequent learning-based models that aim to understand the visual details within the fragments.

In *Scenario 2*, the focus shifts to the semantically segmenting motifs into 12 distinct classes, excluding background regions. This approach allows for more extensive exploration of the diverse artistic motifs adorning fresco surfaces. Our dataset comprises fragment images sourced from two ceiling frescoes, encompassing a total of 12 distinct motif categories, identified by archaeological experts. However, the dataset can be expanded by introducing new motif classes from various frescoes.

In addition to semantic segmentation, we addressed the need for preprocessing the dataset for restoring manual annotations on the fragments made by the archaeologists.To mitigate potential bias in subsequent computational tasks, we systematically eliminated these annotations using a blind inpainting technique. This process ensures a fair evaluation of different approaches when working with the fragments. An overview of the proposed pipeline is illustrated in Fig. 1.

*Equal contribution.

In addition to the goal of reassembly, semantic segmentation of motifs have broader applications in the field of art restoration and computer vision. They serve as valuable resources for various computational tasks, such as *i)* fragment recognition, style classification and clustering, where motif types can provide high-level representation of fragments; *ii)* the segmented motifs can be used to explore recurring themes and symbolic elements present in the frescoes for artistic content analysis; *iii)* semantic segmentation of motifs allows for targeted inpainting and restoration of damaged or missing portions of frescoes, helping to preserve and reconstruct their original appearance.

Our contributions can be summarized as:

- The creation of high-quality rendered images of the painted surface of real fragments of broken frescoes from the Pompeii Archaeological Site,
- Introducing two fragment processing tasks (fragment cleaning and motif extraction) along with annotations in the form of bounding boxes and pixel-wise segmentation maps to aid these tasks.
- Presenting baseline methods for both tasks and conducting a comprehensive comparative analysis of their performance.

2. Related works

The archaeological fragment analysis has primarily been employed for fragment reassembly, known as anastylosis, in the literature [8, 9, 7]. However, it has also been applied to diverse domains, including the recognition of artistic styles on fragments [5], fragment retrieval [4, 16], and fragment classification [20].

Regarding motif segmentation on ancient fresco images, there is limited prior work. To our knowledge, only one study focuses on motif segmentation within this context [4], which initially employs color thresholding to segment a group of fragments captured together during data acquisition. For each segmented fragment, color and shape features are computed to aid in fragment retrieval tasks. Specifically, the motifs' shapes are represented by extracting contour information after segmenting fragment regions using a mean-shift-based color clustering method. Additionally, their fragment retrieval technique is enhanced by incorporating spatiograms as color features. Another work [5] focused on classifying synthetically created fragments based on their artistic style. They utilized off-the-shelf CNN features and various low-level color, texture, and shape features, including color histograms, Hu moments, and Gray Level Co-Occurrence Matrix (GLCM) fed into classification algorithms. For fragmented image reconstruction, a template-matching approach based on computed SIFT features on the fragments was proposed [1] for finding the correct positions of fragments in the fresco plane. Funkhouser et al. [9] utilized 3D models of real-world archaeological fragments to compute a set of geometrical features for their pairwise alignment. In a recent study on archaeological puzzle-solving [7], color clustering was used as a preprocessing step to extract accurate gradient information on the fragments' borders. They also employed color and shape features for fragment reassembly.

[16] introduced software tools to aid in fragment reconstruction. Firstly, a content-based database allows virtual manipulation and retrieval of annotated fragments to identify the best combination before physical restoration, if required. Secondly, once the manual reconstruction is concluded, an inpainting module fills in the gaps and virtually restores the craquelure. Another work [20] proposed a new classification framework for 3D Terracotta Warrior fragments. The core of the framework is a bi-modal neural network that incorporates both geospatial and texture information to classify each fragment into four classes (arm, body, head, and legs). Geospatial information is directly extracted from the point cloud, while a method based on the 3D mesh model and an improved Canny edge detection algorithm is used to extract texture information. The fragment classification eases the subsequent reconstruction.

Although various image segmentation methods have been adopted, semantic segmentation of archaeological fragments is an unexplored task, despite its fundamental importance in understanding the content of the fragments.

3. Case studies

In this work, we focus on semantic segmentation of fragments obtained from fractured frescoes found in the Pompeii Archeological Site, which were damaged by the eruption of 79AD and WW2 bombardments. Initially, a larger collection of fresco fragments was digitized in 3D as part of the RePAIR project[1] which has the ultimate goal of achieving fresco anastylosis. From this extensive collection, we carefully selected a subset of high-quality 3D reconstructed fragments. To achieve photorealistic results, we used Blender[2], for rendering 2D images of these fragments from their corresponding 3D models. During this process, we extracted the flat intact surface containing the pictorial information by aligning it perpendicularly to a virtual camera for optimal rendering. This setup ensured optimal rendering conditions with uniform scale, ideal lighting, and no photographic distortion.

The final objective of the fragment datasets is the reassembly, with motifs playing a critical role in this process. However, the presence of manual markings and the deterio-

[1] 3D models of the fragments will be available at the end of the project. For more information, please visit `https://www.repairproject.eu/`.

[2] `https://www.blender.org/`

ration on the real-world fragments necessitates preprocessing. Collaborating with archaeological experts in this direction, we identified two key tasks: *i)* cleaning the manual annotations and *ii)* extracting the motifs present on the painted surface. For each of the aforementioned tasks, we created separate datasets to support the evaluation of different approaches and provide a baseline for comparison. Indeed, beyond the ultimate goal of reassembly, these cleaned and annotated datasets have broader applications in the field of art restoration and computer vision. They serve as valuable resource for various computational tasks, including fragment recognition, style classification and clustering, artistic content analysis, and digital restoration.

Black-Annotations on the Fresco Fragments (BoFF): Archaeologists make temporary markings on the intact surface of fresco fragments as reminders during the fresco anastylosis task, particularly when dealing with a large number of fragments. These markings include black marks on the borders, indicating neighboring relationships between fragments, and arrows showing the direction of construction lines typically visible only on the backside of the fragments. While useful for archaeologists, these markings can be misleading and introduce bias in computational tasks, such as image recognition, motif classification, style analysis, and automatic puzzle-solving. Thus, cleaning the dataset is crucial to ensure unbiased evaluations of different approaches.

The BoFF dataset is specifically designed for the automatic detection of manual markings in bounding boxes to facilitate their removal through inpainting. It contains 115 fragment images with 405 annotations of bounding boxes covering manual markings on the fragment images. The annotations were prepared using the Roboflow[3] platform. Examples of fragment images in BoFF dataset with manual marks highlighted in red rectangles are shown in Fig. 2

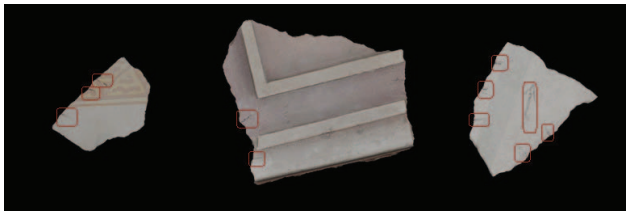

Figure 2. Example fragment images from the BoFF dataset, manual annotations indicated within red boxes.

Motifs on the Fresco Fragments (MoFF): The MoFF dataset is curated for motif extraction and categorization from fragmented frescoes, with a particular focus on Roman ceiling paintings at the Pompeii archaeological site.

[3]https://roboflow.com/

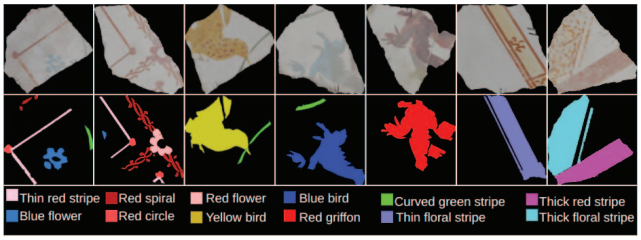

Figure 3. 12 motif categories in the MoFF dataset

These frescoes feature recurring geometric colored patterns known as motifs. The dataset consists of 405 high-resolution images sourced from two distinct ceiling frescoes. Precise pixel-wise segmentation masks, denoting motifs from 12 different classes (see Fig. 3), identified by archaeologists, are provided. These pixel-wise annotations are available in two setups: a 3-class annotation, comprising image background, fragment background, and motif class, and a detailed 12-class annotation, exclusively focusing on distinct motif types (See Fig. 4). Specifically, the motifs belonging to the categories of *Thick Red Stripe*, *Thick* and *Thin Floral Stripe* are exclusively found in one fresco, while *Thin Red Stripe* category is shared by both frescoes. The remaining categories are present only in the other fresco.

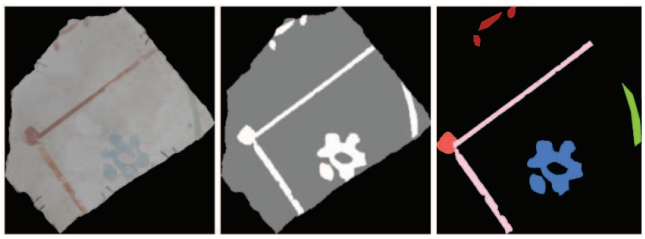

Figure 4. Illustration of pixel-wise semantic annotations for the MoFF dataset. The input image is on the left, followed by 3-class annotations for Scenario 1 in the middle, and motif-wise annotations for Scenario 2 on the right (different colors represent distinct pixel classes).

The MoFF dataset presents several challenges for semantic segmentation. While the image and fragment background classes dominate in terms of pixel count, the motifs of interest are sparsely distributed across the dataset. Furthermore, the dataset exhibits an imbalance problem among the different motif classes as well (See Fig. 5). This imbalance can significantly impact the performance of segmentation models, leading to a bias towards the majority classes and potentially causing difficulties in accurately delineating rare motifs. Moreover, due to the damage and aging of the ancient fresco fragments, the motifs present diverse color and texture variations, making the task of accurate segmentation more challenging. Addressing such challenges and developing robust segmentation algorithms capable of handling such scenarios is still an open question within the computer vision research community [14, 21].

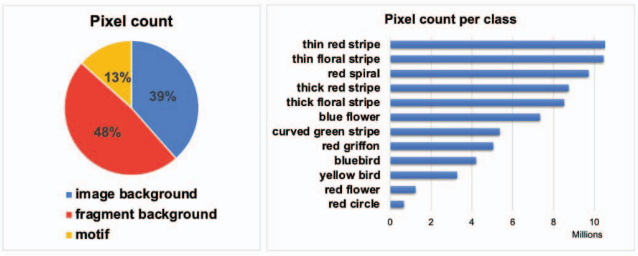

Figure 5. Distribution of pixel count per class for Scenario 1 (left) and Scenario 2 (right)

4. Methodology

Our methodology centers around two distinct scenarios for the segmentation of ancient fresco fragments. In *Scenario 1*, we aim to semantically segment the images into three classes (image and fragment background, and motif class), which eventually yields to detect also the fragment shape, while in *Scenario 2*, our focus is on segmenting the motifs into 12 individual classes. Herein, we assume that fragment regions are already segmented during data acquisition, which is the case in our work. Before delving into these scenarios, we address the challenge posed by human annotations on the fragments. These markings create occlusion and hinder the visibility of motifs and also would potentially affect tasks like fragment reassembly and puzzle solving. To overcome this, our approach involves automatically removing these manual markings, ensuring the fragments are free from such annotations, while simultaneously facilitating the subsequent segmentation tasks.

4.1. Restoration of manual annotations

We employed a blind inpainting approach to automatically detect and remove the manually drawn annotations without the need for pre-defined masks for each fragment.

Detection of manual annotations: The primary goal is to identify and precisely localize the manually drawn annotations on the fragments, thus generating a mask that guides the inpainting algorithm to focus only on the regions requiring restoration. To achieve this, we examined two different approaches: *traditional image processing operations* and *YOLOv5*[12], a state-of-the-art object detection model. For the first approach, adaptive thresholding highlighted the annotation locations, followed by dilation to expand the detected regions and connected component analysis to filter out noise. Despite YOLO was not specifically designed for image inpainting, it was used in a number of works [13, 2] for identifying unwanted objects in an image before their inpainting. In a similar context, we trained YOLOv5 on the BoFF dataset curated specifically for this task, to detect the annotations using bounding boxes. It is worth noting that the performance of the traditional approach heavily depends on the optimal adjustment of input parameters for each image. Even with optimal parameters, large areas of the fragments' surface were highlighted in the resulting masks, risking the degradation of fresco motifs during inpainting. In contrast, YOLOv5 identified more precise regions without any parameter adjustments, which would preserve the artistic integrity of the motifs during inpainting. As a result, we opted to proceed with YOLOv5 for the subsequent restoration steps.

Inpainting manual annotations: We explored multiple inpainting algorithms, including Telea [18], Biharmonic [3], and Criminisi [6], to remove the detected annotations from fragment surfaces. These regions are often located at the borders of the fragments, which can lead to unintended inpainting of the background color (black). To address this issue, we implemented two strategies: firstly, introducing a dent in the mask to restrict background inpainting to the foreground regions of the fresco, and secondly, employing a two-iteration inpainting process, creating a secondary mask for the black regions that emerged on the fragment after the initial iteration. Fig. 6 presents a comparative qualitative analysis of the inpainting methods applied within this adopted strategy to the regions detected by YOLOv5. It is evident that the Telea and Biharmonic methods resulted in blurred inpaintings, making them less suitable for our fresco restoration task. Thus, we proceeded with the Criminisi inpainting method.

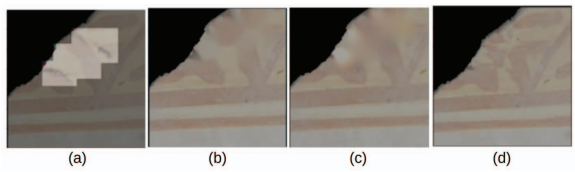

Figure 6. Qualitative analysis of Iterative Telea, Biharmonic and Criminisi inpainting on a randomly chosen fragment. (a) Zoomed input image, created inpainting mask is highlighted, (b) Telea [18], (c) Biharmonic [3], (d) Criminisi [6]

4.2. Segmentation Approaches for Ancient Fresco Fragments

In this section, we present two segmentation scenarios to extract the pictorial content from ancient fresco fragments.

4.2.1 Semantic Segmentation of Background and Motifs

For the first scenario, we adopted the widely recognized UNET model [17] for semantically segmenting the *image background*, *fragment background*, and *motifs* into a single unified motif class. Our approach involves a comprehensive exploration of various configurations for the UNET model, including different architectures, color spaces, and image enhancement operations, aiming to identify the best-

performing configuration for the three-class segmentation task. Challenges included dealing with imbalanced class distributions, particularly due to a large number of background pixels compared to the motif class. Additionally, a single unified motif class poses challenges with high intra-class variations due to different motifs exhibiting varying characteristics in terms of color and shape. This scenario assumes high inter-class variations, making the task comparatively easier than the one in Scenario 2.

Modified U-Net Architecture: The modification of the U-Net model was achieved by reducing the depth of the network. The original network contains four layers in the contracting path (encoder) and a corresponding four layers in the expanding path (decoder), as well as a bottleneck layer. In contrast, the simplified version contains two layers in the contracting path, two layers in the expanding path, and a bottleneck layer. Together with using relatively small input image sizes, the simplification process provided several advantages, including reduced overfitting and eventually enhanced performance on the small dataset, which is the case in this work, and a shorter training time. Moreover, we explored the effect of using weighted loss during training to deal with the class imbalance problem.

Different color spaces: Ancient fresco fragments present unique challenges. These challenges include variations in the color and texture of motifs due to the aging process, regions with faded motifs, and potential deterioration or damage to the fragments, leading to missing or unclear motifs. By exploring different color spaces, we aim to enhance the model's ability to handle these challenges effectively. Certain color spaces might better capture the subtle differences in motif color and texture, allowing for more accurate segmentation even in cases of fading or deterioration. Moreover, utilizing specific color spaces might enhance the visibility of faded motifs, leading to improved segmentation results. In this context, in addition to conventional RGB, we explored HSV and YCrCb. HSV separates color information from brightness, and YCrCb separates the luminance information from chrominance information, which can help in better isolating motifs from the background, especially in cases where variations in brightness occur due to the presence of dust and fading.

Image Enhancement Techniques: In order to enhance the visibility of motifs by increasing the contrast of the image, we explored the effect of image enhancement methods on motif segmentation performance. In particular, inspired by the reported improvements in [15], we employed Contrast Limited Adaptive Histogram Equalization (CLAHE), Histogram Equalization, and gamma correction. For achieving contrast enhancement in the brightness component of the image while maintaining the original color balance and avoiding undesirable color shifts, we applied histogram equalization, gamma correction, and CLAHE to V and Y channels of HSV and YCrCb, respectively. RGB images were converted to HSV, and after applying such operations on the V channel, they were transformed back to the RGB color space.

4.2.2 Semantic Segmentation of Motif Regions

In this scenario, we address the more challenging task of semantically segmenting motifs into 12 distinct classes. Several challenges arise in this scenario, including the class imbalance problem, high intra-class variation caused by broken fragments resulting in different-sized and shaped partitions of the same motif type, and low inter-class variation where some motifs may share similar characteristics, leading to potential misclassifications during manual annotation or automatic segmentation. To tackle these challenges, we noted the need for higher-resolution input images and more sophisticated models. Consequently, we adopted 512-pixel-sized images and utilized the original U-NET and YOLOv8-Seg [11] models, selecting the color space that performed well in the previous scenario.

5. Experiments

5.1. Evaluation metrics

To evaluate the performance of YOLO-v8 and U-Net models, various metrics are used depending on the task they are designed for. For the semantic segmentation tasks, we used *Intersection-over-Union* (*IoU*) (or *Jaccard Index*) and *Pixel Accuracy* (*PA*) averaged over the k pixel classes ($k = 3$ and $k = 13$ for Scenarios 1 and 2, respectively) present in the predicted segmentation masks. The formula for calculating either averaged IoU and PA, and the IoU and PA for a single class i is given in Eq. 1 and Eq. 2, where TP_i, FN_i, and FP_i represent True Positives, False Negatives, and False Positives computed for class i, respectively.

$$IoU_{average} = \frac{1}{k} \sum_{i=1:k} IoU_i = \frac{1}{k} \sum_{i=1:k} \frac{TP_i}{(TP_i + FN_i + FP_i)} \quad (1)$$

$$PA_{average} = \frac{1}{k} \sum_{i=1:k} PA_i = \frac{1}{k} \sum_{i=1:k} \frac{TP_i}{TP_i + FP_i} \quad (2)$$

For the detection of the manual annotation task, we evaluated the performance of YOLOv5 both in terms of *Precision* (which has the same formula with $PA_{average}$ shown in Eq. 2) and *mAP50* (mean average precision at IoU threshold 0.5). In this task, Precision takes priority over Recall, as inpainting regions of false positive detections (i.e., regions without manual marks) can potentially cause the loss of valuable information in the fresco image. By maximizing precision, we ensure that the inpainting algorithm is applied only to regions with actual manual marks, minimizing the risk of losing important fresco details.

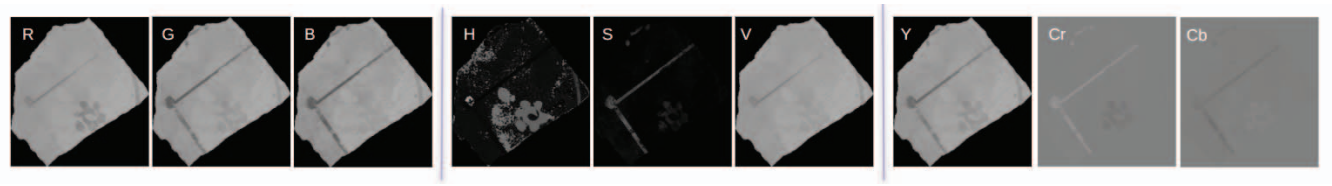

Figure 7. Representation of different channels of each color space

5.2. Restoration of manual annotations

We evaluated YOLOv5 model as a baseline for the BoFF dataset for detecting manual annotations on the fragments. The dataset, composed of 115 images, split randomly into training, validation, and test sets in an 80/10/10 ratio. This led to 91, 12, and 12 images in each set, respectively, along with 324, 42, and 39 annotated boxes. The images were resized to 416×416 pixels to suit the model's requirements. We used a pre-trained model from Yolov5[4] for model initialization. In training, rotation-based data augmentation expanded the train set to 273 images with 972 box annotations, boosting model performance. Default hyperparameters, including 300 epochs, the SGD optimizer, and a batch size of 32 were used during training. To optimize the model's performance, we experimented with various layer freezing configurations and found that freezing the first three layers (layers 0 to 2) yielded the best results on the validation set. Early stopping was employed, with the patience parameter set to 100, to prevent overfitting and enhance the model's generalization capability.

We evaluated the trained YOLOv5 model on the test set in terms of Precision and mAP50. In our experiments, the trained model achieved **Precision** and **mAp0.5** scores of **0.741** and **0.596**, respectively, on the test set. YOLOv5 model was able to identify and locate manual annotations, even in manual marks appear in low-contrast, while accounting for their different shapes, sizes, and intensities, despite displaying a degree of limitation when faced with non-monochromatic backgrounds within the detected bounding boxes. More specifically, we obtained 28 True Positives, 3 False Positives, and 11 False Negatives in 12 test images.

The qualitative detection results are shown in Fig. 8. We have carefully chosen two sample fragment images from BoFF dataset to encompass various fragment characteristics, including instances of elongated arrows on the fragment surface and fragments displaying textured patterns. Corresponding inpainting masks and the resulting inpainting outcomes, achieved using the method outlined in Section 4, are also presented.

5.3. Segmentation of Fresco Fragments

Prior to motif segmentation on the MoFF dataset, we used the YOLOv5 model trained on the BoFF training set to detect manual annotations on the MoFF images and inpaint them. Then, we cropped them to focus only on the fragment region, effectively removing any unnecessary background information. These preprocessing steps ensure that the subsequent motif segmentation tasks, i.e., Scenarios 1 and 2, can be carried out on cleaned fragment surfaces.

[4]https://pytorch.org/hub/ultralytics_yolov5/

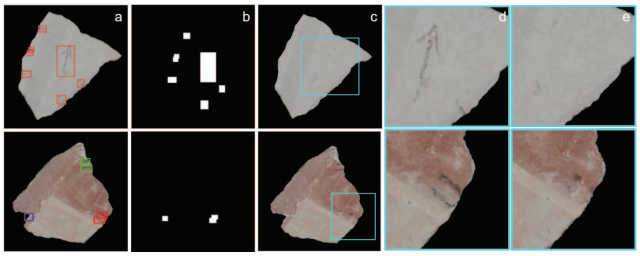

Figure 8. Restoration of manual annotations on two example fragments. a) YOLOv5 model detections (True Positives, False Positives, and False Negatives by YOLOv5 are highlighted in red, blue, and green boxes, respectively); b) generated inpainting masks1; c) inpainting results; (d) and (e) are detailed views of (a) and (c).

In all experiments, the MoFF dataset was split into train, validation, and test sets with an 80/10/10 ratio, resulting in 324, 40, and 39 images, respectively. Hyperparameter tuning was performed on the validation set, exploring batch sizes of 16, 32, and 64, as well as learning rates of 0.01, 0.001, 0.0001, and 0.00001.

5.3.1 Semantic Segmentation of Background and Motifs

In the experiments for this scenario, to optimize computational efficiency, we resized the restored and cropped images to 256×256 pixels for the comparative performance analysis of U-NET architectures, color space, and image enhancement. For Scenario 1, in addition to the averaged IoU and PA for the three-pixel classes (image background, fragment background, and motif), we also calculated them specifically for the motif class. This allowed us to evaluate the model's overall segmentation accuracy across all classes and its ability to precisely capture the motifs on the fresco fragments.

U-NET architecture: Both the original and modified U-NET architectures performed best with a batch size of 32 and the learning rate scheduler starting from 0.001, where the first 25 epochs maintained the same rate, followed by a reduction by a factor of 0.1. For training, we used conven-

tional RGB images. Early stopping was employed with the patience parameter set to 10 to prevent overfitting. To address the class imbalance problem, we experimented with the weighted cross-entropy loss [10]. It favors less occurring pixels classes in the training set with higher weights, during training in addition to the initial experiment adopting unweighted loss.

Performance results are presented in Table 1. It is seen that adopting weighted loss significantly improves performance in terms of PA_{motif}. This improvement is expected due to the increased weight assigned to the least represented class, which happens to be the motif class. This adjustment penalizes prediction errors within this class more heavily, consequently leading to enhanced detection performance.

Table 1. Performance evaluation of different UNET architectures, using *unweighted and †weighted loss for imbalanced distribution of classes. RGB images are used in the experiments. No augmentation was applied.

Architecture	$IoU_{average}$	IoU_{motif}	$PA_{average}$	PA_{motif}
Original UNET*	0.74	0.01	0.83	0.003
Modified UNET*	0.75	0.07	0.84	0.07
Original UNET†	0.76	0.19	0.86	0.21
Modified UNET†	**0.82**	**0.32**	**0.89**	**0.39**

Different color spaces and image enhancement methods: Using the simplified UNET architecture which performed best in Table 1, we explored segmentation performance using input images represented in different color spaces and various image enhancement techniques applied to them. Training is accomplished using 50 epochs using the weighted loss in these experiments. Performance results are shown in Table 2, where we highlight the first, second, and third best-performing configurations using **red**, blue, and *green* fonts, respectively.

It is observed that the RGB color space achieves relatively high overall segmentation scores in terms of $IoU_{average}$ and $PA_{average}$, while its motif-specific segmentation scores are significantly lower than in overall segmentation. HSV color space outperforms RGB in both overall segmentation and motif-specific segmentation, showing higher IoU and PA scores. This demonstrates that HSV better captures the fine details in motif color and texture, leading to improved motif segmentation results. YCrCb color space also performs reasonably well, with competitive scores compared to RGB, but slightly lower than HSV in both overall segmentation and motif-specific segmentation.

Image enhancement operations, except Gamma correction, had a negative impact on the RGB color space significantly. In contrast, for the HSV color space, most image enhancement methods (i.e., HistEq, and Gamma) had a positive impact, resulting in higher IoU and PA scores for both overall segmentation and motif-specific segmentation. This highlights HSV's superiority in capturing fine motif details and benefiting from image enhancements. For the YCrCb color space, CLAHE and HistEq yielded improvements in overall segmentation.

These results indicate that the choice of color space, coupled with specific image enhancement techniques, can influence segmentation performance, and HSV with appropriate enhancement stands out as the most effective approach for accurate motif segmentation on the MoFF dataset.

Table 2. Segmentation performance of different color spaces and image enhancement techniques on MoFF dataset. The first, second, and third best-performing configurations are shown using **red**, blue, and *green* fonts, respectively.

Configuration	$IoU_{average}$	IoU_{motif}	$PA_{average}$	PA_{motif}
RGB	0.82	0.32	*0.89*	0.39
HSV	**0.87**	*0.65*	**0.92**	0.81
YcrCb	0.80	0.45	0.88	0.72
RGB&CLAHE	0.47	0.14	0.63	0.14
RGB&HistEq	0.82	0.43	*0.89*	0.48
RGB&Gamma	0.86	0.48	0.91	0.58
HSV&CLAHE	0.56	**0.69**	0.71	0.83
HSV&HistEq	0.80	0.57	*0.89*	**0.98**
HSV&Gamma	0.86	0.66	**0.92**	0.90
YCrCb&CLAHE	**0.87**	0.59	**0.92**	0.70
YCrCb&HistEq	*0.85*	0.60	**0.92**	*0.88*
YCrCb&Gamma	*0.85*	0.51	0.91	0.67

Fig. 9 presents qualitative results for two example fragments, one clean and the other deteriorated. The U-NET model mostly succeeds in segmenting the fragment regions; however, the motif segmentation in the degraded fragment exhibits some incorrect segmentations.

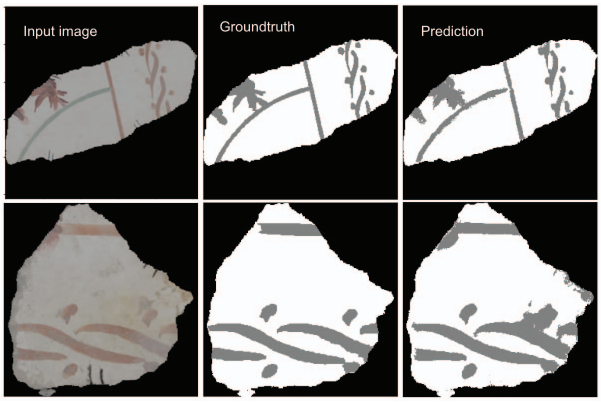

Figure 9. Example images and segmentation results of Modified U-NET for Scenario 1 (computations were done by HSV images).

5.3.2 Semantic Segmentation of Motif Regions

In the experiments for Scenario 2, restored and cropped fragment images in HSV color space were resized to 512×512 pixels to obtain a more detailed representation of the motifs.

Semantic motif segmentation using U-NET: We explored both UNET architectures - modified and original - for se-

mantic motif segmentation. Rotation is used as a data augmentation technique to enhance the model's robustness to unseen data. The goal is to segment the motif regions into 12 classes, corresponding to different motif types. To achieve this, we trained the U-NET models with ground truth masks containing 13 classes (12 motif classes + 1 background class), where the fragment background and image background classes were merged into a single background class. This approach simplifies the task, focusing only on motif segmentation without the need to distinguish between fragment and image backgrounds, which aligns with the primary objective. The U-NET was trained in a similar manner to the previous scenario, except that the batch size was decreased to 8 for the Original and 16 for the Modified U-NET. The rest of the parameters did not change. On-the-fly data augmentation was added in the form of random rotation and random flips, both horizontal and vertical.

Semantic motif segmentation using YOLOv8: The YOLOv8 architecture achieves motif segmentation through its instance segmentation capabilities. Although an official paper describing the details of its implementation is not yet available, it builds upon the success of the previous models [19] as a fast and reliable object detector. This makes it well-suited for our experiments, where detecting motifs is the primary focus while the background holds relatively less significance. The model was trained from scratch using ground truth polygonal masks[5] containing the 12 different motif classes. Training is performed using a batch size of 16 for 200 epochs.

Results: We present the results in Table 3, evaluating them using the same metrics as in our prior experiments. It is seen that U-NET, even in the original architecture, was unable to achieve high IoU_{motifs} or PA_{motifs} scores, while the YOLO approach significantly outperforms U-NET in terms of both metrics. In this assessment, the PA_{motifs} and IoU_{motifs} refer to the average across all 12 motif classes, excluding the background.

Table 3. YOLOv8 achieves the best results regarding the motif segmentation (PA_{motifs} includes all classes without background), while UNET wins when including the background in the evaluation (PA_{avg} refer to all classes including background, same for IoU).

Architecture	IoU_{motifs}	IoU_{avg}	PA_{motifs}	PA_{avg}
YOLOv8	**0.582**	0.538	**0.634**	**0.797**
Original U-NET	0.416	**0.606**	0.452	0.630
Modified U-NET	0.345	0.569	0.392	0.600

We also present the results for all 13 classes, including the background class. Interestingly, it is seen that the IoU_{avr} of YOLOv8 drops notably when including this measure, whereas the opposite is true for U-NET. U-NET's IoU_{avr} reaches its peak for both the original and modified architectures. This demonstrates that U-NET predicts background class better, which is actually not our primary goal. Future work could involve devising a custom loss function to adjust the weighting of background pixels. This adjustment would enable U-NET to focus more on motif segmentation, potentially leading to higher PA values. Eventually, it is worth noting that segmenting 12 different motif types, which encompass smaller regions compared to the overall background region, presents a significant challenge for future research.

Fig. 10 presents two qualitative results for the experimented architectures. It is seen that YOLOv8 localizes motifs better and predicts motif class with higher precision than two U-NET architectures.

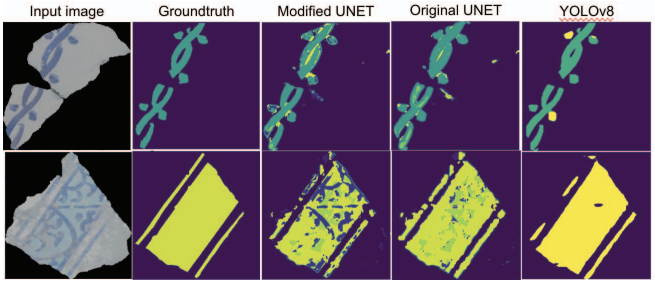

Figure 10. Semantic motif segmentation results of different architectures for Scenario 2.

6. Conclusions

This paper focuses on an unexplored aspect related to fresco fragments: semantic segmentation. This task involves extracting meaningful information from fragments, facilitating downstream activities like fragment classification and reassembly. We have introduced a dataset for semantic segmentation comprising real fresco fragments from the Pompeii Archaeological Site, along with baseline models.

Additionally, we have introduced an additional task concerning fragment cleaning. In this context, we have curated a dataset containing annotations of archaeological marks that necessitate restoration before fragment analysis. Our experiments have been conducted on both datasets using standard metrics, state-of-the-art baselines, and a comprehensive analysis of diverse color spaces.

The results demonstrate that semantic segmentation of fresco fragments is an achievable objective, opening up possibilities for more complex activities that necessitate a semantic understanding of fragmented artifacts.

Acknowledgements: This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 964854.

[5] YOLOv8 uses polygonal masks around the objects and does not make use of a pixel-wise segmentation mask.

References

[1] Paola Barra, Silvio Barra, Michele Nappi, and Fabio Narducci. Saffo: A sift based approach for digital anastylosis for fresco reconstruction. *Pattern Recognition Letters*, 138:123–129, 2020.

[2] Vojtěch Bartl, Jakub Špaňhel, and Adam Herout. Persongone: Image inpainting for automated checkout solution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3115–3123, 2022.

[3] Marcelo Bertalmio, Guillermo Sapiro, Vicent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000.

[4] Sonia Caggiano, Maria De Marsico, Riccardo Distasi, and Daniel Riccio. Mosaic: Multi-object segmentation for assisted image reconstruction. In *Pattern Recognition: Applications and Methods: 4th International Conference, ICPRAM 2015, Lisbon, Portugal, January 10-12, 2015, Revised Selected Papers 4*, pages 282–299. Springer, 2015.

[5] Lucia Cascone, Michele Nappi, Fabio Narducci, and Sara Linda Russo. Classification of fragments: recognition of artistic style. *Journal of Ambient Intelligence and Humanized Computing*, 14(4):4087–4097, 2023.

[6] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. In *IEEE Transactions on image processing*. IEEE, 2004.

[7] Niv Derech, Ayellet Tal, and Ilan Shimshoni. Solving archaeological puzzles. *Pattern Recognition*, 119:108065, 2021.

[8] Piercarlo Dondi, Luca Lombardi, and Alessandra Setti. Dafne: A dataset of fresco fragments for digital anastlylosis. *Pattern Recognition Letters*, 138:631–637, 2020.

[9] Thomas Funkhouser, Hijung Shin, Corey Toler-Franklin, Antonio García Castañeda, Benedict Brown, David Dobkin, Szymon Rusinkiewicz, and Tim Weyrich. Learning how to match fresco fragments. *Journal on Computing and Cultural Heritage (JOCCH)*, 4(2):1–13, 2011.

[10] Yaoshiang Ho and Samuel Wookey. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE access*, 8:4806–4813, 2019.

[11] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023.

[12] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, Kalen Michael, TaoXie, Jiacong Fang, imyhxy, Lorna, (Zeng Yifu), Colin Wong, Abhiram V, Diego Montes, Zhiqiang Wang, Cristi Fati, Jebastin Nadar, Laughing, UnglvKitDe, Victor Sonck, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Dhruv Nair, Max Strobel, and Mrinal Jain. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation, Nov. 2022.

[13] Chanran Kim, Younkyoung Lee, Jong-Il Park, and Jaeha Lee. Diminishing unwanted objects based on object detection using deep learning and image inpainting. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, pages 1–3. IEEE, 2018.

[14] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022.

[15] Tawsifur Rahman, Amith Khandakar, Yazan Qiblawey, Anas Tahir, Serkan Kiranyaz, Saad Bin Abul Kashem, Mohammad Tariqul Islam, Somaya Al Maadeed, Susu M Zughaier, Muhammad Salman Khan, et al. Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in biology and medicine*, 132:104319, 2021.

[16] Daniel Riccio, Sonia Caggiano, Maria De Marsico, Riccardo Distasi, and Michele Nappi. Mosaic+: Fragment retrieval and reconstruction enhancement for virtual restoration. *Journal of Visual Languages & Computing*, 31:139–149, 2015.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[18] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics, gpu, and game tools*, 9(1):23–34, 2004.

[19] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7464–7475, June 2023.

[20] Kang Yang, Xin Cao, Guohua Geng, Kang Li, and Mingquan Zhou. Classification of 3d terracotta warriors fragments based on geospatial and texture information. *Journal of Visualization*, 24:251–259, 2021.

[21] Ying Yu, Chunping Wang, Qiang Fu, Renke Kou, Fuyu Huang, Boxiong Yang, Tingting Yang, and Mingliang Gao. Techniques and challenges of image segmentation: A review. *Electronics*, 12(5), 2023.