Team 4: Restoration of damaged statues with inpainting model

Alejandro Campayo Fernández
7058536

Aishwarya Kshirsagar
7056998

Sai Suresh Macharla Vasu
7047009

1. Task and Motivation

This project is inspired by the controversy surrounding the digital scanning of the Parthenon Marbles at the British Museum [2]. The Institute for Digital Archaeology (IDA) 3D scanned these sculptures to create detailed models for study and preservation, aiming to enhance accessibility to this cultural heritage. Now, the IDA plans to add color to these 3D models to envision the marbles in their original painted state, which has faded over time.

The IDA project [5] aims to provide an accurate representation of how historical sculptures originally appeared. However, after reviewing their objectives, we recognized the necessity for prior reconstruction of some sculptures, due to their deteriorated conditions, often marked by cracks or missing limbs.
Current methods for such reconstructions, like those of the Venus de Milo [7] or the statues of Aesculapius and Hygeia [3], have required many resources and time.
Our contribution aims to reduce the costs and time for these restorations by approaching the problem as an inpainting task. This task will involve training an AI model (GANs) to take an image of a damaged sculpture as input and generate a reconstruction of how it might have appeared before damage occurred.

We will draw inspiration from the paper titled "Restoration of Damaged Face Statues Using Deep Generative Inpainting Model" [11] to guide our project. Our goal is to expand upon this approach to include the restoration of extremities and heads on damaged statues.

Additionally, we plan to explore using existing models to generate 3D reconstructions from single image [12]. This approach won't involve training but will help us evaluate and visualize our results effectively. This aligns our work closely with the IDA's focus on working with 3D data, facilitating connections between our research and their efforts.

2. Goals

Many inpainting models face issues with blurring and not being able to recreate the intricate details of the image. So, this will be our primary challenge.
To get good results, we estimate a necessity of 10000 good quality images. However, our limited expertise in this area means this estimate could be inaccurate.
We are unsure whether the datasets available online or via web scraping will be sufficient. Therefore, we plan to use data augmentation techniques, similar to those in [11], to address potential shortages of images.
In [11], a total of 4000 training images and 750 test images were used for the task of inpainting facial features. However, our scenario may need a larger dataset, given that our task involves not only generating an arm but also accurately placing it in a realistic position. This requires the model to implicitly learn about pose estimation. Specifically, the model must estimate the orientations of the arms, legs, and head given a body orientation.
The second major challenge will be the amount of time and compute required for the task. In [11] they use a single TITAN X 12GB GPU. We are aware that we will have a GPU available for the project but we do not know its specifications yet. Finally, as we might need to train multiple times and then tweak our parameters according to the results.
For the midterm we would like to have a dataset containing approx 10.000 images for training and a model ready to be used even if it has not been trained yet. Moreover we would like to test the model for 3D reconstruction from a single image using Dust3R, that does not require any scene or camera information and can generate 3D reconstructions with a minimal of one image [12], since this is complementary just for our results to match better with the work of IDA.

3. Methods

In our project, we will employ two models to achieve accurate inpainting and 3D reconstruction of statues with missing parts.
The first step involves obtaining an image of a statue that has missing parts. This image will then be processed using our deep learning-based inpainting algorithm to reconstruct the missing parts of the statue.
Following the inpainting, a 3D reconstruction model will be applied to the inpainted image to generate a complete 3D

model of the statue. We will use Dust3R for this task [12]. To enhance the quality of the input images for the inpainting algorithm, we might use denoising and other preprocessing techniques. These steps will ensure that the input image is clean and suitable for the inpainting process, leading to better results.
For the task of inpainting, we will experiment with various methods to determine the most effective approach. So far, the architecture that has caught our attention the most is U-Net [8]. U-Net employs an encoder-decoder structure with skip connections to retain high-resolution details while reconstructing the image. However, the problem with using U-Net with a standard loss function is that the model often learns to return blurred image reconstructions. This happens because the model minimizes the loss by producing safe, average predictions rather than taking the risk of generating sharp but potentially incorrect edges. For this reason, we are considering using Generative Adversarial Networks (GANs) [4] with a U-Net generator. In GANs, the generator fills in missing regions of an image while the discriminator distinguishes between real and generated images. GANs have shown promising results in image inpainting tasks. By using GANs, we expect the model to improve its ability to handle complex image reconstruction tasks while preserving details and edges effectively. GANs, particularly tailored for image generation and completion, offer high-quality, realistic outputs by learning the underlying distribution of the data.
Traditional methods like Partial Differential Equation (PDE)-based image inpainting and Exemplar-based inpainting, while useful in certain contexts, have limitations that make them less suitable for our use case. For this reason we have decided not to use them. Exemplar-based inpainting can be useful in cases like removing objects from background. These traditional methods often struggle with reconstructing large missing areas and fail to capture the intricate details and textures present in the original images.

4. Datasets

As mentioned earlier, in [11], they used 4,000 training images and 750 test images for the task of inpainting facial traits. Given the complexity of our task, where we not only need to estimate body parts but also their varying positions across images, we anticipate requiring more data. We estimate approximately 10,000 images will be necessary.
To acquire this data we intend to use online sources (scraping) as well as data from publicly available datasets.

4.1. Pose Estimation Datasets

Despite the limited literature on statue images, we think this task is also a pose estimation problem. There are many publicly available datasets for pose estimation [9] [6] [1] where instead of sculptures, pictures of people are provided. The biggest advantage of these datasets is that most of them have annotations on where the joints locations are. With this information we could decide to mask specific parts of the body in our training data. We anticipate that training our model on human datasets will enable it to generalize effectively to our scenario involving statues.

4.2. Greek Sculptures Dataset

Dataset [10] offers 1600 pictures of art of 8 Greek gods (200 pictures for each). However, some of these images will have to be filtered out because of being paintings/drawings. We estimate we can keep aproximately 3/4 of the data. Moreover, this data has a limitation which is that no joints information is provided. This hinders masking specific parts of the image during training.

4.3. Web Scraping

Our final recourse would be to scrape images from the internet, either of sculptures or, if necessary, of real people if statue images are insufficient. This approach faces similar challenges as discussed in Section 4.2, particularly in lacking joint information. Our strategy for training the GAN will involve masking specific parts of the images, predominantly limbs, to guide the model in predicting the unmasked portions accurately as ground-truth.

5. Evaluation

5.1. Quantitative Evaluation

To evaluate the performance of our inpainting model, we will employ several widely-used quantitative metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Fréchet Inception Distance (FID). We can use the loss functions like Mean Squared Error(MSE) or Mean Absolute Error(MAE) for this task.

5.2. Qualitative Evaluation

For the qualitative evaluation, we will begin with an image of a complete statue, one with no missing parts. We will then artificially degrade this image by removing certain parts, such as the head or arms, to create a partially incomplete image. The degraded image will be provided as input to our inpainting model. The inpainted image generated by our model will be compared to the original, undamaged image to assess the model's ability to accurately reconstruct the missing parts.

References

[1] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2

[2] Mark Brown. British museum facing legal action over parthenon marbles 3d scan refusal. *The Guardian*, 2022. 1

[3] Leonarda Fazio, Mauro Lo Brutto, Sara Gonizzi Barsanti, and Saverio Giulio Malatesta. The virtual reconstruction of the aesculapius and hygeia statues from the sanctuary of isis in lilybaeum: Methods and tools for ancient sculptures' enhancement. *Applied Sciences*, 12(7), 2022. 1

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. 2

[5] Institute for Digital Archaeology. Our purpose, 2024. Accessed: 2024-06-29. 1

[6] Sam Johnson and Mark Everingham. Leeds sports pose dataset. `https://paperswithcode.com/dataset/lsp`, 2010. 2

[7] Rachel Kousser. Creating the past: The vénus de milo and the hellenistic reception of classical greece. *American Journal of Archaeology*, 109(2):227–250, 2005. 1

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 2

[9] Leonid Sigal, Alexandru O. Balan, and Michael J. Black. Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. `http://humaneva.is.tue.mpg.de/datasets_human_1`, 2006. 2

[10] Thatgeeman. Sculptures of greek olympians dataset, 2024. Accessed: 2024-06-29. 2

[11] Abraham Theodorus. Restoration of damaged face statues using deep generative inpainting model, August 2020. 1, 2

[12] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy, 2023. 1, 2