

Shadow Art Revisited: A Differentiable Rendering Based Approach

Kaustubh Sadekar
CVIG Lab, IIT Gandhinagar
sadelkar.k@iitgn.ac.in

Ashish Tiwari
CVIG Lab, IIT Gandhinagar
ashish.tiwari@iitgn.ac.in

Shanmuganathan Raman
CVIG Lab, IIT Gandhinagar
shanmuga@iitgn.ac.in

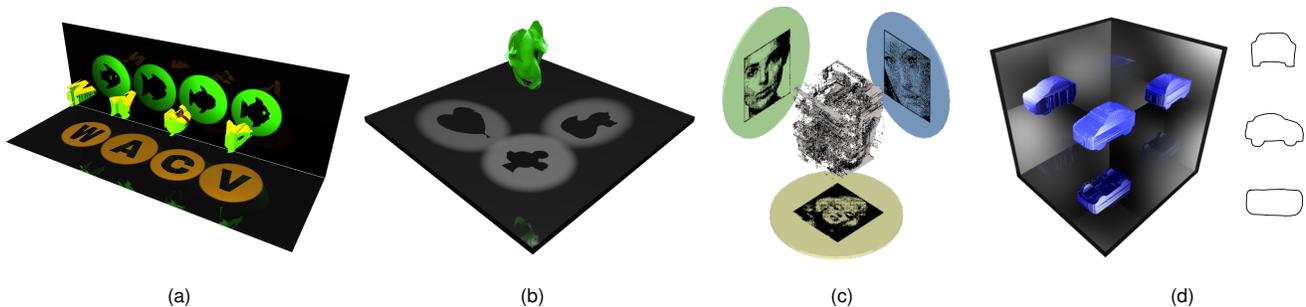


Figure 1: Shadow art sculptures generated using differentiable rendering casting the shadows of (a) WACV acronym on one plane and fishes on the other resembling an aquarium of floating objects, (b) dropping *Heart*, *Duck*, and *Mickey* (all on the same plane), and (c) face sketches using half-toned images. (d) 3D reconstruction of a car from hand drawn sketches.

Abstract

While recent learning-based methods have been observed to be superior for several vision-related applications, their potential in generating artistic effects has not been explored much. One such exciting application is *Shadow Art* - a unique form of sculptural art that produces artistic effects through 2D shadows cast by a 3D sculpture. In this work, we revisit shadow art using differentiable rendering-based optimization frameworks to obtain the 3D sculpture from a set of shadow (binary) images and their corresponding projection information. Specifically, we discuss shape optimization through voxel as well as mesh-based differentiable renderers. Our choice of using differentiable rendering for generating shadow art sculptures can be attributed to its ability to learn the underlying 3D geometry solely from image data, thus reducing the dependence on 3D ground truth. The qualitative and quantitative results demonstrate the potential of the proposed framework in generating complex 3D sculptures that transcend the ones seen in contemporary art pieces using just a set of shadow images as input. Further, we demonstrate the generation of 3D sculptures to cast shadows of faces, animated movie characters, and the applicability of the proposed framework to sketch-based 3D reconstruction of the underlying shapes.

1. Introduction

The very art of painting originates from trailing the edge of shadow, said Pliny the Elder, an ancient Roman author. Shadow art showcases the author’s imaginative and technical skill in playing with shadows. Many of us have seen or at least heard of “someone” making “something” interesting out of shadows. Remember creating shadows of rabbits or horses on the wall by playing with our fingers around a lamp. In this work, we show how differentiable rendering can generate some amazing 3D sculptures that cast mind-boggling shadows when lit from different directions.

Figure 2 (a) shows the cover of the book *Gödel, Escher, Bach* by Douglas Hofstadter that features blocks casting shadows of different letters when seen from different sides. Kumi Yamashita - one of the most prominent contemporary artists - demonstrated that simple objects arranged in a certain manner cast amazing shadows or silhouettes when lit from an appropriate direction. An exclamation sign transforms into a question mark (Figure 2 (b)) and a bunch of aluminum numbers form an image of a girl looking down the building (Figure 2 (c)). These and several other art pieces by Kumi Yamashita not only please our eyes, but also inspire emotion and pose intriguing questions. *Tim*

¹This work is supported by SERB IMPRINT-2 grant. Code, data, and results are available at kaustubh-sadekar.github.io/ShadowArt-Revisited/

Noble and *Sue Webster* have presented such artworks since 1997, projecting shadows of people in different positions (Figure 2 (d)). This specifically arranged ensemble shows how readily available objects can cast the clearest of illusions of clearly recognizable scenes (Figure 2(e)). Figure 2 (f) shows the aquarium of floating characters by *Shigeo Fukuda* where the shadows of the fish reveal their names in kanji characters. Even after such fascinating effects, the current state of shadow art seems to be well described by Louisa May Alcott, who says “*Some people seemed to get all sunshine, and some all shadow...*”. Mitra and Pauly [10] first introduced shadow art to the vision and graphics community by formally addressing the problem in an optimization framework. Since then, no significant progress has been observed in this direction.

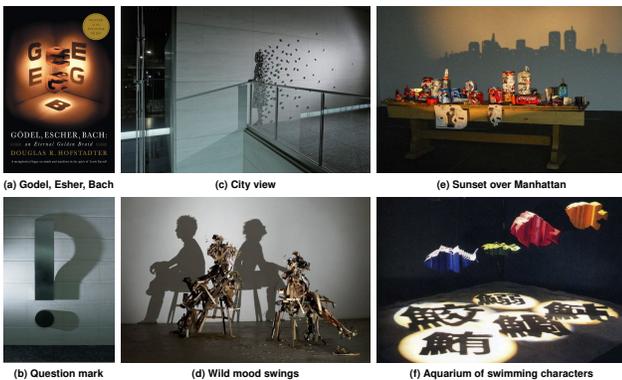


Figure 2: Examples of shadow art sculptures by (a) *Douglas Hofstadter*, (b, c) *Kumi Yamashita*, (d, e) *Tim Noble and Sue Webster*, and (f) *Shigeo Fukuda*.

The question: Can we develop a method that learns to create or optimize 3D sculptures that can generate such artistic effects through their shadows? In this work, we attempt to answer this question through the use of *Differentiable Rendering*. Here, instead of trying to render a scene of our creation realistically, we try to reconstruct a representation of a scene from one or more images of a scene [2]. Our work is mostly inspired by examples of shadow art shown in Figure 2. Specifically, our objective is to generate 3D shadow art sculptures that cast different shadows (of some recognizable objects) when lit from different directions using a differentiable renderer.

Why differentiable rendering? Most learning-based methods for 3D reconstruction require supervision through precise 3D ground truths for training. However, all we have is a set of desired shadow images in our case. Differentiable rendering-based methods estimate the 3D shape under 2D supervision using single or multi-view images, thus, eliminating the need for any 3D data collection and annotation.

Contributions. The following are the major contributions of this work.

- We introduce a differentiable rendering-based framework to create 3D shadow art sculptures that cast different shadows when lit from different directions using just the input shadow images and the corresponding projection information.
- We demonstrate the efficacy of deploying a differentiable rendering pipeline over voxel and mesh-based representations to generate shadow art sculptures.
- We show that the proposed framework can create artistic effects beyond the ones seen in contemporary art forms by generating 3D sculptures using half-toned face images and sketches drawn from multiple viewpoints.
- To the best of our knowledge, ours is the first work to address shadow art using differentiable rendering.

2. Related Work

Shadows provide essential cues in the way we perceive the world around us and have been central in capturing the imagination of many artists, including stage performers. Several artists have typically used manual and trial-and-error style approaches to create 3D shadow sculptures. However, with the advent of digital design technology, the need for an automated framework is inevitable.

Shadow Art. Shadows in many computer graphics and computer vision applications have been studied from both perceptual (artist’s) and mathematical (programmer’s) points of view. It started by understanding how shadows effect the perception of spatial relationships in computer-generated images [22, 23]. Pellacini *et al.* developed an interactive user interface for cinematic shadow design. By imposing constraints on the desired shadows, it allows the users to modify the positions of light sources and shadow blockers [17]. Then evolved the idea of shadow volume - a visual hull used for 3D reconstruction [4]. Sinha and Pollefeys [19] used min-cuts and strict silhouette constraints to study the reconstruction of continuous surfaces from multiple images.

Relation with the state-of-the-art method. The work closest to ours is by Mitra *et al.* [10]. They described shadow art more formally by introducing a voxel-based optimization framework to recover the 3D shape from arbitrary input (shadow) images by deforming them and handling inherent image inconsistencies. This work demonstrates the potential of differentiable rendering in generating 3D shadow sculptures all from arbitrary shadow images without any explicit input image deformation. Although the associated 3D object might not exist in the real world, the method still creates shadow sculptures that transcend the ones seen in contemporary art forms casting the physically realizable shadows when lit from appropriate directions.

Differentiable Rendering. We briefly review methods that learn the 3D geometry via differentiable rendering. These methods are categorized based on the underlying 3D representation: point clouds, voxels, meshes, or implicit neural representation. In this work, we primarily focus on voxel and mesh-based representations. Several methods operate on voxel grids [8, 13, 16, 21]. Paschalidou *et al.* [16] and Tulsiani *et al.* [21] propose a strong probabilistic ray potential formulation. However, it requires the intermediate evaluations to be saved for backpropagation and is limited to smaller resolution voxel grids. On one hand, Sitzmann *et al.* [20] used an LSTM-based differentiable renderer to infer implicit scene representations from RGB images and Liu *et al.* [7] perform max-pooling over the ray intersections from multi-view silhouettes. On the other hand, [14] demonstrate that volumetric rendering is inherently differentiable for implicit representations and does not require saving the intermediate results for the backward pass. OpenDR [9] roughly approximates the backward pass of mesh-based graphics pipelines. Liu *et al.* [6] proposed Soft Rasterizer to make the rasterization step differentiable. It uses a deformable template mesh for training and yields compelling results in reconstruction tasks. We deploy this in our mesh-based differentiable rendering pipeline for rasterization.

This work describes the differentiable rendering optimization frameworks for both voxel and mesh-based representations and discuss their strengths and weaknesses.

3. Method

3.1. Problem Formulation

The key idea of our work is to generate an artistic 3D sculpture \mathcal{S} that casts N different shadows when lit from N different directions using differentiable rendering based optimization pipeline. The prime focus here is to create interesting shadow art effects using the 3D sculpture \mathcal{S} . The input to the pipeline is a set $\mathcal{X} = \{X_1, X_2, \dots, X_N\}$ of shadow configuration $X_i = (I_i, P_i)$. I_i represents the target shadow image and P_i is the corresponding projection information. The shadow of an object can be regarded as its projection on a planar surface. Assuming directional lighting, this projection is an *orthographic projection* when the surface is perpendicular to the lighting direction and a *perspective projection*, otherwise [1]. Obtaining the shadow of an object is equivalent to finding the corresponding silhouette captured by a camera pointing in the same direction as the light source. Therefore, I_i the shadow image is essentially a silhouette. From here on, we shall use the term silhouette images and shadow images interchangeably.

The shadow art problem is similar to a multi-view 3D reconstruction problem [5, 11], where we try to estimate the 3D structure of an object given its N silhouette views.

However, the following are the key differences in shadow art. (i) The N views can correspond to arbitrary silhouettes (not necessarily of the same object). (ii) The learned 3D sculpture may bear no resemblance with any real-world object and be an abstract art that casts the desired shadows when lit from appropriate directions. Undoubtedly, there exist multiple 3D shapes that can cast the same set of shadows. However, our concern is to learn one such 3D sculpture that can create the desired artistic effects through its shadows.

3.2. System Overview

By providing shadow configuration $\mathcal{X} = \{X_i = (I_i, P_i) | i = 1, 2, \dots, N\}$ as input to the pipeline, the objective is to learn the underlying 3D sculpture \mathcal{S} , as described earlier. The projection information P_i corresponds to the camera position (and hence, the light source position) associated with i^{th} shadow image I_i such that $P_i = (\mathbf{R}_i, \mathbf{t}_i)$. Here, \mathbf{R}_i and \mathbf{t}_i are the 3D rotation and translation of the camera, respectively. We start by initialising \mathcal{S} with a standard geometry which is further optimized by minimizing image-based losses, such that the rendered silhouette images $\tilde{I}_i = I_i$ for all $i = 1, 2, \dots, N$. The prime reason for using differentiable rendering is that it allows gradient flow directly from images back to parameters of \mathcal{S} to optimize it in an iterative fashion. In other words, it does not require any explicit 3D supervision and optimizes the 3D shape solely from image based losses. For further simplicity, let the set of target shadow images and the associated projection information be denoted as \mathcal{I} and \mathcal{P} , respectively, such that $\mathcal{I} = \{I_1, I_2, \dots, I_N\}$ and $\mathcal{P} = \{P_1, P_2, \dots, P_N\}$. Further, let $\tilde{\mathcal{I}} = \{\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_N\}$ be the set of shadow images obtained from learned 3D sculpture \mathcal{S} as per projections \mathcal{P} .

In this work, we consider two common representations for 3D shapes i.e. *voxel* and *mesh* based representations and discuss which representation fits the best for the under different scenarios. In the following section, we elaborate on the optimization pipelines for the voxel and mesh-based representations of the 3D object to create visually plausible shadow art using differentiable rendering.

3.3. Voxel Based Optimization

This section looks at a differentiable rendering pipeline that uses voxels to represent the 3D geometry. A voxel is a unit cube representation of a 3D space. The 3D space is quantized to a grid of such unit cubes and is parameterized by an N -dimensional vector signifying the volume occupied in 3D space. Additionally, it encodes occupancy, transparency, color, and material information. Even though occupancy and transparency probabilities (in the range $[0, 1]$) are different, they are considered to be the same to maintain differentiability during the ray marching [2]. A typical rendering process involves collecting and aggregating the

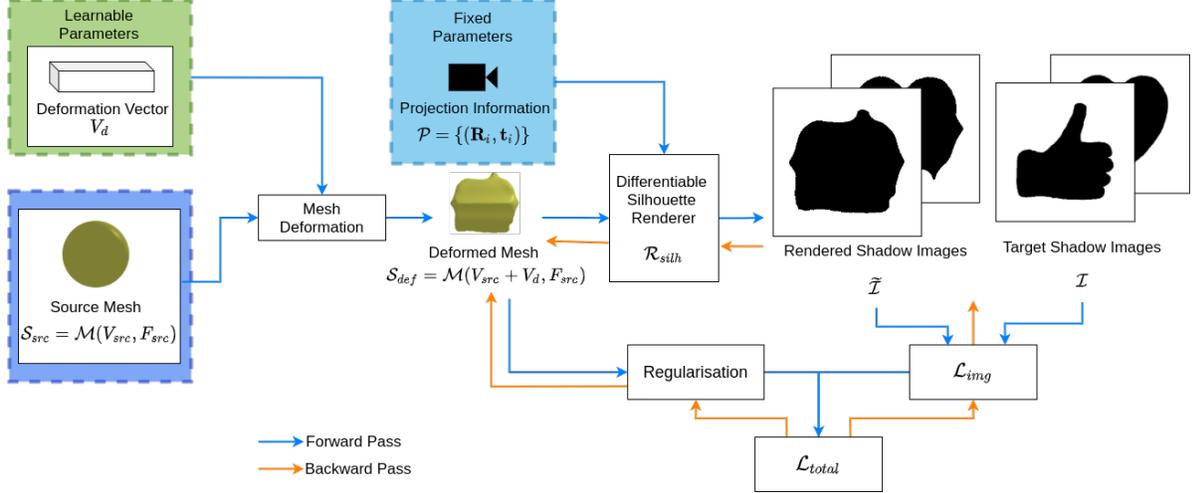


Figure 3: Information flow in the proposed mesh-based differentiable rendering pipeline.

voxels located along a ray and assigning a specific color to each pixel based on the transparency or the density value. While rendering a pixel, all the voxels located along a ray projecting to a pixel are considered. However, our objective is to do the inverse, i.e., to find the 3D geometry associated with silhouettes corresponding to different directions.

We assume that the 3D object \mathcal{S} is enclosed in a 3D cube of known size centered at the origin. Hence, \mathcal{S} can be defined by a learnable 3D tensor V that stores the density values for each voxel. We initialize V with all ones. The color value for each voxel is set to 1. It is kept fixed in the form of a color tensor C . Next, we render \mathcal{S} using a differentiable volumetric rendering method described in [18]. To restrict the voxel density values to the range $[0, 1]$, V is passed through a sigmoid activation function (σ) to obtain \tilde{V} , as described in Equation 1.

$$\tilde{V} = \sigma(V) \quad (1)$$

We then pass \tilde{V} through the differentiable volume renderer \mathcal{R}_{vol} along with the fixed color tensor C and the associated projection information \mathcal{P} to obtain the set of corresponding rendered images $\tilde{\mathcal{I}}$, as described in Equation 2.

$$\tilde{\mathcal{I}} = \mathcal{R}_{vol}(\tilde{V}, C, \mathcal{P}) \quad (2)$$

The voxel densities V are optimized by minimizing the image level loss between a set of rendered shadow images $\tilde{\mathcal{I}}$ and the corresponding target shadows in \mathcal{I} . The image-level loss \mathcal{L}_{img} is a weighted combination of L_1 and L_2 losses, as described in Equation 3.

$$\mathcal{L}_{img} = \lambda_1 \mathcal{L}_{L_1} + \lambda_2 \mathcal{L}_{L_2} \quad (3)$$

Here, $\lambda_1 = 10.0$ and $\lambda_2 = 10.0$ are the weights associated with L_1 and L_2 losses, respectively. The resulting voxel-

based representation of \mathcal{S} can finally be converted to a 3D mesh making it suitable for 3D printing (see Figure 9). One simple way to achieve this is by creating faces around each voxel having a density greater than a certain threshold value (as described in [18]).

3.4. Mesh Based Optimization

In this section, we also propose to use mesh-based differentiable rendering to meet our objective. The entire workflow is described in Figure 3. The 3D object \mathcal{S} can be represented as a mesh $\mathcal{M}(V, F)$. Here, V is a set of vertices connected by a set of triangular faces F that define the surface of \mathcal{S} . We start by initializing a source mesh $\mathcal{S}_{src} = \mathcal{M}(V_{src}, F_{src})$ with an icosphere consisting of $|V_{src}|$ vertices and $|F_{src}|$ faces. The idea is to learn the per-vertex displacements V_d to deform \mathcal{S}_{src} to the final desired mesh that casts desired shadows (silhouettes), when lit from appropriate directions. This is achieved by rendering the deformed mesh $\mathcal{S}_{def} = \mathcal{M}(V_{def}, F_{def})$ through a mesh-based differentiable silhouette renderer \mathcal{R}_{silh} (as described in [18]) from the associated projection \mathcal{P} such that,

$$\begin{aligned} V_{def} &= V_{src} + V_d \\ F_{def} &= F_{src} \\ \tilde{\mathcal{I}} &= \mathcal{R}_{silh}(\mathcal{S}_{def}, \mathcal{P}) \end{aligned} \quad (4)$$

3.4.1 Loss Function

The source mesh is optimized by minimizing image-level loss \mathcal{L}_{img} (described in Equation 3), normal consistency loss, and imposing Laplacian and edge length regularisation.

Normal consistency. We use normal consistency loss

to ensure smoothness in the resulting 3D sculpture. For a mesh $\mathcal{M}(V, F)$, let $e = (\mathbf{v}_x, \mathbf{v}_y)$ be the connecting edge of two neighboring faces $f_x = (\mathbf{v}_x, \mathbf{v}_y, \mathbf{a})$ and $f_y = (\mathbf{v}_x, \mathbf{v}_y, \mathbf{b})$, such that $f_x, f_y \in F$ with normal vectors \mathbf{n}_x and \mathbf{n}_y , respectively. If $\tilde{\mathcal{E}}$ is the set of all such connecting edges e and $|F|$ is the total number of faces in mesh, the normal consistency over all such neighbouring faces f_x and f_y is given as per Equation 5.

$$\mathcal{L}_{norm} = \frac{1}{|F|} \sum_{e \in \tilde{\mathcal{E}}} (1 - \cos(\mathbf{n}_x, \mathbf{n}_y)) \quad (5)$$

where,

$$\begin{aligned} \mathbf{n}_x &= (\mathbf{v}_y - \mathbf{v}_x) \times (\mathbf{a} - \mathbf{v}_x) \\ \mathbf{n}_y &= (\mathbf{b} - \mathbf{v}_x) \times (\mathbf{v}_y - \mathbf{v}_x). \end{aligned}$$

Laplacian regularisation. To prevent the model from generating large deformations, we impose uniform Laplacian smoothing [12], as described by Equation 6.

$$\mathcal{L}_{lap} = \frac{1}{|V|} \sum_{i=1}^{|V|} \left(\left\| \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} w_{ij} \mathbf{v}_j - \mathbf{v}_i \right\|_1 \right) \quad (6)$$

Here, $|V|$ is the number of vertices in the mesh \mathcal{M} and $\mathcal{N}(\mathbf{v}_i)$ is the neighbourhood of vertex \mathbf{v}_i .

$$w_{ij} = \frac{\omega_{ij}}{\sum_{k \in \mathcal{N}(i)} \omega_{ik}}$$

For uniform Laplacian smoothing, $\omega_{ij} = 1$, if $(\mathbf{v}_i, \mathbf{v}_j)$ form an edge, $\omega_{ij} = -1$ if $i = j$, and $\omega_{ij} = 0$, otherwise.

Edge length regularisation. Edge-length regularisation is included to prevent the model from generating flying vertices and is given by Equation 7.

$$\mathcal{L}_{edge} = \sum_{i=1}^{|V|} \sum_{\mathbf{v}_j \in \mathcal{N}(\mathbf{v}_i)} \|\mathbf{v}_i - \mathbf{v}_j\|_2^2 \quad (7)$$

Finally, the overall loss function is as described in Equation 8.

$$\mathcal{L}_{total} = \lambda_a \mathcal{L}_{img} + \lambda_b \mathcal{L}_{norm} + \lambda_c \mathcal{L}_{lap} + \lambda_d \mathcal{L}_{edge} \quad (8)$$

Here, $\lambda_a = 1.6$, $\lambda_b = 2.1$, $\lambda_c = 0.9$, and $\lambda_d = 1.8$ are the weights associated with the losses \mathcal{L}_{img} , \mathcal{L}_{norm} , \mathcal{L}_{lap} , and \mathcal{L}_{edge} , respectively.

3.5. Implementation Details

The aforementioned differentiable rendering pipelines are implemented using Pytorch3D [18]. For initializing

the mesh, we use a level 4 icosphere composed of 2,562 vertices and 5,120 faces. For the voxel-based rendering pipeline, we assume that the object is inside a cube (a grid of $128 \times 128 \times 128$ voxels) centered at the origin with a side of length 1.7 world units. We train the optimization pipeline with custom silhouette images of size 128×128 for 2000 epochs. We keep the learning rate to 1×10^{-4} . We keep the learning rate to 1×10^{-2} and train the optimization pipeline for 500 epochs. The training is performed on NVIDIA Quadro RTX 5000 with 16 GB memory.

4. Experimental Analysis

In this section, we perform an extensive analysis of the results obtained using voxel and mesh-based differentiable rendering pipelines to create plausible shadow art effects. We discuss the evaluation metrics and perform ablation studies to understand the effect of various loss terms in the design.

4.1. Evaluation Metrics

Following our discussion in Section 3.1, we assess the quality of silhouettes (shadow images) obtained through the 3D sculpture \mathcal{S} as per projections \mathcal{P} . To compare the rendered silhouette images with the target silhouette images (representing shadows), we use Intersection over Union (IoU) and Dice Score (DS). Additionally, we need to quantify the quality of the 3D sculpture \mathcal{S} obtained after optimization. While we do not have any ground truth for 3D shape, and this is an optimization framework, we need a "no reference" quality metric. Therefore, we decided to use normal consistency evaluated over \mathcal{S} to assess the quality of the mesh.

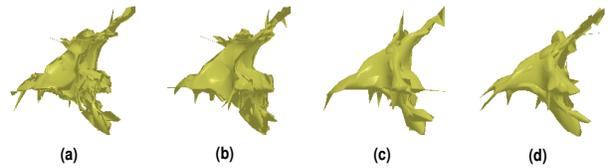


Figure 4: Qualitative analysis of effect of various loss terms. (a) \mathcal{L}_{img} , (b) $\mathcal{L}_{img} + \mathcal{L}_{edge}$, (c) $\mathcal{L}_{img} + \mathcal{L}_{edge} + \mathcal{L}_{lap}$, and (d) $\mathcal{L}_{img} + \mathcal{L}_{edge} + \mathcal{L}_{lap} + \mathcal{L}_{norm}$.

4.2. Ablation Study

Figure 4 depicts the qualitative effect of different loss terms used in the optimization pipeline. The underlying mesh in this figure corresponds to the arrangement shown in Figure 5 (c). The image-based loss \mathcal{L}_{img} alone is not sufficient for generating plausible 3D sculptures as they are expected to suffer from distortions due to flying vertices (spike-like structures in Figure 4 (a)) or large deformations.

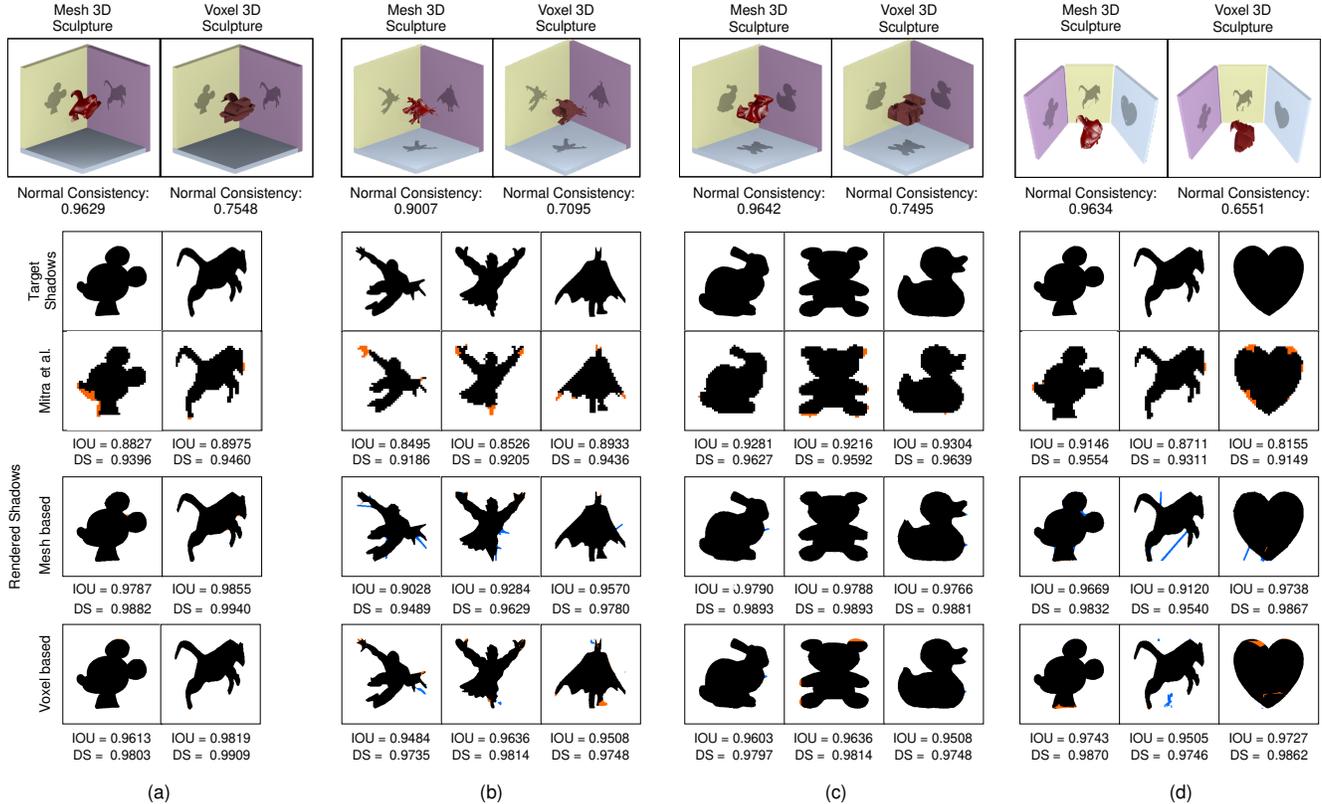


Figure 5: Qualitative and quantitative results on (a) two views (b,c) three orthogonal views, and (d, e) three non-orthogonal views using voxel and mesh-based rendering for shadow art.

Since we do not have any ground truth for explicit 3D supervision, we examine the effect of including regularisation in the objective function. Figure 4 (b) shows that the spikes are reduced by introducing edge-length regularisation. Further, as shown in Figure 4 (c), Laplacian smoothing prevents the sculpture from experiencing super large deformations. Finally, normal consistency loss ensures further smoothness in the optimized surface. Figure 4 (d) shows the result obtained by applying all the regularization as mentioned above terms along with the image-based loss. The resulting quality of the mesh validates our choice of loss terms.

4.3. Qualitative and Quantitative Analysis

In this section, we perform the qualitative and quantitative evaluation on a wide variety of shadow images, including those used in [10] to illustrate the versatility of our approach in generating 3D shadow art sculptures represented using both voxels and mesh. For every result in Figure 5 (a)-(d), we show the learned 3D sculptures (voxel and mesh-based) along with the respective shadows cast from different specified directions. We could not include the optimized 3D sculpture from [10] as the associated object file was not downloadable through their optimization tool. We

have been able to incorporate both orthogonal (Figure 5 (a, b, c)) and non-orthogonal views (Figure 5 (d) and Figure 1 (b)) to obtain the shadows that are consistent with the desired target shadow images. For a quantitative comparison, we also report IoU and Dice score. As depicted in Figure 5, the IoU and Dice Score are comparable for both voxel and mesh-based renderings. However, the corresponding voxel-based 3D sculptures are not that smooth (low normal consistency value) when compared to those of mesh-based 3D sculptures. It is important to note that the underlying voxel representation has been converted to a mesh representation to compute normal consistency values. While [10] have focused only on foreground inconsistencies (marked in orange color), we also show the background inconsistencies (marked in blue color) that appear in some of the rendered shadow images. Ours is an end-to-end optimization approach without any additional editing tool to prune the generated 3D sculpture. In some cases, the mesh-based approach is found to produce certain discontinuities near non-convex regions (Figure 5 (b,d)) for at least one view. This is mainly attributed to the inability of icosphere to handle sharp discontinuities in the desired shape, especially when regularisation has been imposed (Equation

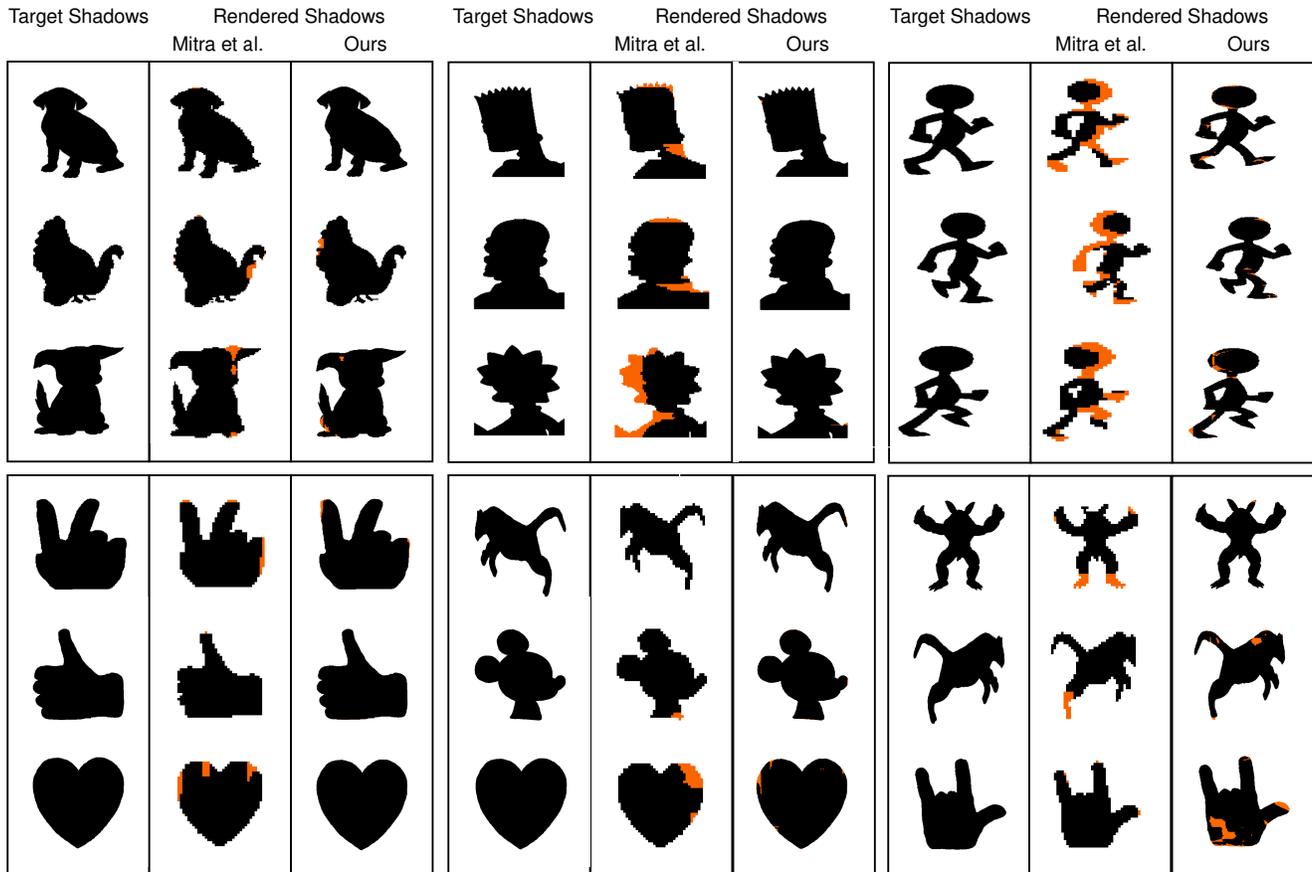


Figure 6: Qualitative evaluation of results obtained through (A) shadow art tool in [10] and (B) our voxel based rendering pipeline. The inconsistencies are highlighted in orange color.

8). The voxel-based approaches may contain a few outliers (voxels outside the desired 3D shape, as marked in blue in Figure 5 (d)) which is generally not the case with mesh-based approaches. However, the mesh-based differentiable rendering method lags in handling sharp discontinuities and holes present in the shadow images. While these shortcomings are handled effectively by voxel-based methods, they tend to generate discretized 3D sculptures and are often associated with high memory and computational requirements. Overall, the differentiable rendering-based optimization for both the approaches generates plausible 3D shadow art sculptures and outperforms [10] in handling shadow inconsistencies by a large extent without having to deform the desired shadow images explicitly.

4.4. Comparison with the State-of-the-art method

We show the qualitative comparison of the results obtained using our voxel-based differentiable rendering pipeline and the voxel-based optimization tool presented in [10] without any deformation to the target shadow image. In Figure 6, we observe that the shadows rendered using

the proposed pipeline are highly consistent with that of the desired target shadows when compared to those produced by [10]. The authors of [10] argue that it might be impossible to find a consistent configuration with a given choice of input images. Therefore, they introduce deformations in the input image to achieve consistency of the rendered shadow images with the desired ones. However, the differentiable rendering-based optimization can handle inconsistencies without causing any explicit change in the target shadow images.

5. Applications

In this section, we show some additional artistic shadow art creations and an extension to yet another application that can also benefit from our optimization approach. Figure 7 depicts the creation of faces of well-known scientists around the world and movie characters like *Minions* and *Ironman*, demonstrating the strength of differentiable rendering based optimization approach to handle complex objects or scenes with consistency. In addition to the binary silhouette images, half-toned images can also be used to generate 3D

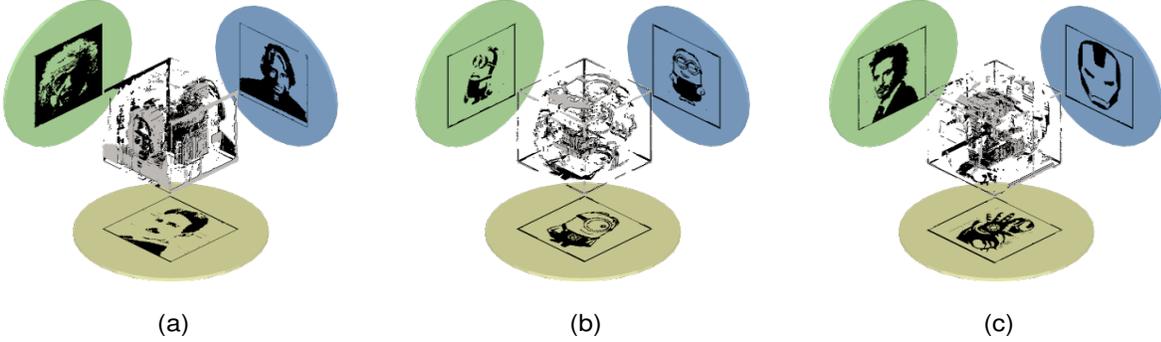


Figure 7: A seemingly random voxel soup creates three distinct shadow images of (a) *Albert Einstein, Nikola Tesla, and APJ Abdul Kalam*, (b) *Minions*, and (c) *Ironman*.

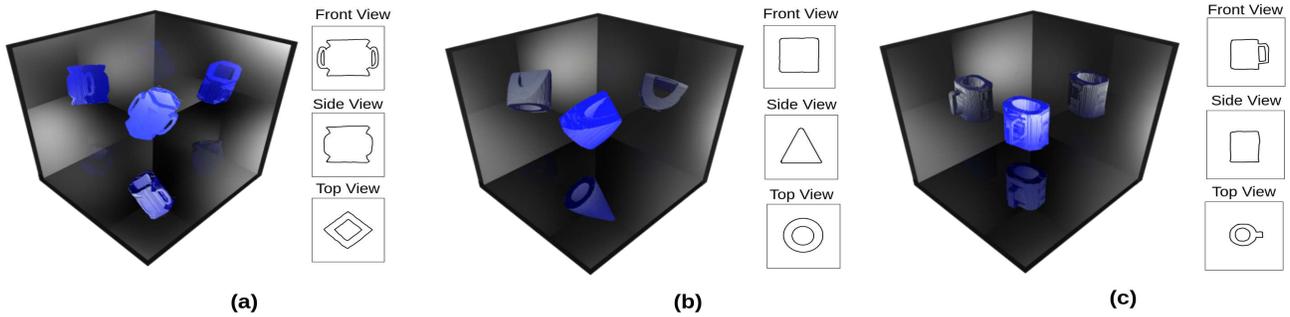


Figure 8: 3D reconstruction of (a) flower vase, (b) pen-stand, and (c) coffee mug using the associated hand drawn sketches from three different views.

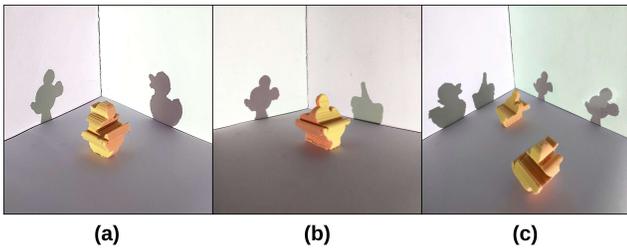


Figure 9: 3D printed sculptures of quality upto the 3D printer’s resolution.

shadow art sculptures, as shown in Figure 1. Another interesting extension is towards sketch-based modeling [15] where we use hand-drawn sketches of a shape from different viewpoints to automatically create the underlying 3D object. We demonstrate the creation of a flower vase (Figure 8 (a)), pen-stand (Figure 8 (b)), and a coffee mug (Figure 1 (c)) solely from hand-drawn sketches from three different views.

Figure 9 shows the 3D printed structures to demonstrate their physical realizability. Contouring [3] can be used to convert the voxel grids to a triangular mesh, and some steps of subdivision on meshes yields a smooth, printable surface. The structures shown in Figure 6 and 8 can be phys-

ically printed, as shown in Figure 9. The facial structures in Figure 7 and 1 (c) are difficult to print using the conventional 3D printing techniques due to missing connectivity in the learned voxel representation. Converting such representations to meshes is challenging and can potentially be explored in the future.

6. Conclusion

We have introduced an optimization framework for generating 3D shadow art sculptures from a set of shadow images and the associated projection information. The key idea is to explore the strength of differentiable rendering in creating visually plausible and consistent shadows of rigid objects, faces, and animated movie characters by generating the associated 3D sculpture. We have discussed voxel and mesh-based rendering pipelines and have identified the benefits of each of them for the task at hand. Additionally, we have demonstrated the applicability of the proposed framework in reconstructing 3D shapes using their sketches drawn from three different viewpoints. At present, we have primarily considered the shadows associated with static sculptures, and hence, they themselves are static in nature. Dynamic shadow art can also be explored in the near future.

References

- [1] W. Abbott. *The Projection of Shadows*, pages 336–342. Springer Netherlands, Dordrecht, 1971.
- [2] Hiroharu Kato, Deniz Beker, Mihai Morariu, Takahiro Ando, Toru Matsuoka, Wadim Kehl, and Adrien Gaidon. Differentiable rendering: A survey. *arXiv preprint arXiv:2006.12057*, 2020.
- [3] Michael Kazhdan, Allison Klein, Ketan Dalal, and Hugues Hoppe. Unconstrained isosurface extraction on arbitrary octrees. In *Symposium on Geometry Processing*, volume 7, 2007.
- [4] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on pattern analysis and machine intelligence*, 16(2):150–162, 1994.
- [5] Jinho Lee, Baback Moghaddam, Hanspeter Pfister, and Raghu Machiraju. Silhouette-based 3d face shape recovery. 2003.
- [6] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7708–7717, 2019.
- [7] Shichen Liu, Shunsuke Saito, Weikai Chen, and Hao Li. Learning to infer implicit surfaces without 3d supervision. *arXiv preprint arXiv:1911.00767*, 2019.
- [8] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *arXiv preprint arXiv:1906.07751*, 2019.
- [9] Matthew M Loper and Michael J Black. Opendr: An approximate differentiable renderer. In *European Conference on Computer Vision*, pages 154–169. Springer, 2014.
- [10] Niloy J Mitra and Mark Pauly. Shadow art. *ACM Transactions on Graphics*, 28(CONF):156–1, 2009.
- [11] Adem Yasar Mulayim, Ulas Yilmaz, and Volkan Atalay. Silhouette-based 3-d model reconstruction from multiple images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 33(4):582–591, 2003.
- [12] Andrew Nealen, Takeo Igarashi, Olga Sorkine, and Marc Alexa. Laplacian mesh optimization. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 381–389, 2006.
- [13] Thu Nguyen-Phuoc, Chuan Li, Stephen Balaban, and Yong-Liang Yang. Rendernet: A deep convolutional network for differentiable rendering from 3d shapes. *arXiv preprint arXiv:1806.06575*, 2018.
- [14] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3504–3515, 2020.
- [15] Luke Olsen, Faramarz F Samavati, Mario Costa Sousa, and Joaquim A Jorge. Sketch-based modeling: A survey. *Computers & Graphics*, 33(1):85–103, 2009.
- [16] Despoina Paschalidou, Ali Osman Ulusoy, and Andreas Geiger. Superquadrics revisited: Learning 3d shape parsing beyond cuboids. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10344–10353, 2019.
- [17] Fabio Pellacini, Parag Tole, and Donald P Greenberg. A user interface for interactive cinematic shadow design. *ACM Transactions on Graphics (TOG)*, 21(3):563–566, 2002.
- [18] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv:2007.08501*, 2020.
- [19] Sudipta N Sinha and Marc Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 349–356. IEEE, 2005.
- [20] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *arXiv preprint arXiv:1906.01618*, 2019.
- [21] Shubham Tulsiani, Tinghui Zhou, Alexei A Efros, and Jitendra Malik. Multi-view supervision for single-view reconstruction via differentiable ray consistency. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2626–2634, 2017.
- [22] Leonard Wanger. The effect of shadow quality on the perception of spatial relationships in computer generated imagery. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 39–42, 1992.
- [23] Leonard R Wanger, James A Ferwerda, Donald P Greenberg, et al. Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics and Applications*, 12(3):44–58, 1992.