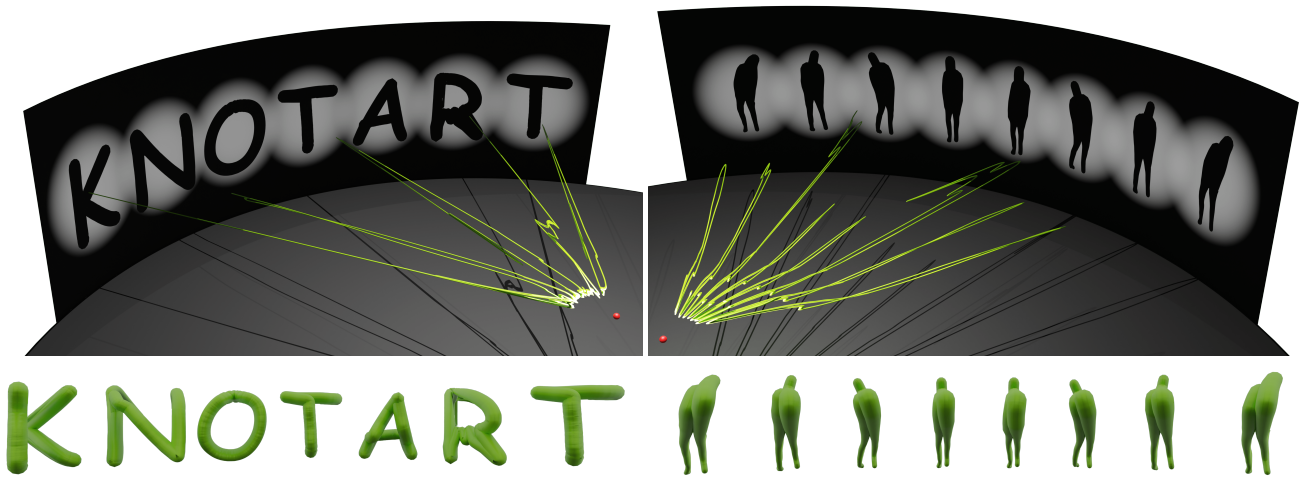
DOI: 10.1111/cgf.15138
Eurographics Symposium on Geometry Processing 2024
R. Hu and S. Lefebvre
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 5

Search Me Knot, Render Me Knot: Embedding Search and Differentiable Rendering of Knots in 3D

Aalok Gangopadhyay, Paras Gupta, Tarun Sharma, Prajwal Singh and Shanmuganathan Raman

CVIG Lab, Indian Institute of Technology, Gandhinagar



(a) Alphabetical Images (b) Dancing Person

Figure 1: *Results obtained by our approach for (a) alphabetical images and (b) dancing person image sequence. The red dot indicates the viewer location as well as the light location. Our method searches for the embedding of the tubular-knots (shown in green), so that the projected images resemble the target images. In both the cases, the top row image depicts the shadow cast on a nearby gallery wall by a light source placed at the red dot and the bottom row depicts the image a viewer will perceive if viewing from the red dot.*

Abstract

We introduce the problem of knot-based inverse perceptual art. Given multiple target images and their corresponding viewing configurations, the objective is to find a 3D knot-based tubular structure whose appearance resembles the target images when viewed from the specified viewing configurations. To solve this problem, we first design a differentiable rendering algorithm for rendering tubular knots embedded in 3D for arbitrary perspective camera configurations. Utilizing this differentiable rendering algorithm, we search over the space of knot configurations to find the ideal knot embedding. We represent the knot embeddings via homeomorphisms of the desired template knot, where the weights of an invertible neural network parametrize the homeomorphisms. Our approach is fully differentiable, making it possible to find the ideal 3D tubular structure for the desired perceptual art using gradient-based optimization. We propose several loss functions that impose additional physical constraints, enforcing that the tube is free of self-intersection, lies within a predefined region in space, satisfies the physical bending limits of the tube material, and the material cost is within a specified budget. We demonstrate through results that our knot representation is highly expressive and gives impressive results even for challenging target images in both single-view and multiple-view constraints. Through extensive ablation study, we show that each proposed loss function effectively ensures physical realizability. We construct a real-world 3D-printed object to demonstrate the practical utility of our approach.

CCS Concepts

• ***Computing methodologies*** → ***Computer graphics; Rendering;***

1. Introduction

Suppose an artist wishes to create a perceptual artwork in three-dimensional space by creating a structure that appears meaningful only when seen from specific viewpoints in the scene and appears arbitrary or meaningless from other viewpoints. With such a goal set, the artist has in mind the target visual images that should be perceived from specific viewpoints. Figuring out the 3D structure that would resemble the target images is an inverse process, which is very challenging and requires human ingenuity. Consider another scenario where light sources are placed at specific locations, such that when the light gets cast onto the 3D structure, a shadow is formed on a nearby wall. The artist would desire the shadow formed to match the target image in this scenario. In the first case, the viewer is at a specific location; in the second, the light source is at a specific location. Despite the differences, these scenarios involve the same underlying principle of perspective projection, where the target 2D image is the known parameter, and the 3D art structure is the unknown parameter. As an example, consider Fig. 1, where given a specific viewpoint (red dot), the top image depicts the shadow cast of 3D structures (green tube) on a nearby wall cast by a light placed at the viewpoint, and the bottom image depicts the image perceived by a viewer situated at the viewpoint.

Drawing inspiration from the wire-sculpture shadow art of Larry Kagan [Lar23; Med23], here we consider a specific form of perceptual art, where the 3D structure is made of tube-based shapes like wires, ropes or strings. A tube here refers to the volume sweep of a sphere along a curve embedded in 3D space. For the curves, we specifically consider knots or loops, which are closed curves in 3D. Knots have a rich historical and cultural significance in various traditions and arts [MK96; Evi24]. They are of deep interest in both arts and mathematics. They naturally lend themselves to intriguing shapes, resulting in visually appealing patterns. In this work, we are interested in exploring the intricate shadow projections that arise out of knot-based tubes.

Given a desired target silhouette image and the tube thickness, we search for the embedding or the configuration of the knot in 3D so that the knot-based tube, when viewed from a given camera viewpoint, resembles the target image. Moreover, while creating such a 3D structure, the artist would also need to ensure that the structure is physically realizable; i.e.., it should be free of self-intersections, and given the material that will be used for the tube, the structure should respect the physical constraints, for instance, the maximum allowed bending. Since such a structure is to be realised in the real world, the artist might want the tubular structure to fit inside a predefined region in space in the scene. Also, considering the material cost, the artist might want to minimize the usage of the material, which in this case would be directly proportional to the length of the knot.

We have proposed an end-to-end differentiable optimization framework to guide an artist in finding a 3D structure corresponding to the desired target image(s). We search for the 3D art structure by searching for knot embeddings in 3D, where we define the search space through a parametrized family of homeomorphisms represented using an Invertible Neural Network (INN). We have designed a differentiable rendering algorithm that, given a knot embedding, the thickness of the tube, and the information about the perspective viewing camera placed at a specific viewpoint, generates a silhouette rendering of the tube that would be perceived from that specific viewpoint. Our approach has the theoretical guarantee of the knot embedding being free of self-intersection. We have developed an appropriate loss function to avoid self-intersections in the tube. We also penalize high curvature regions, thereby restricting the tube bending. Material cost constraint is satisfied by adding a penalty whenever the knot length exceeds the allowed limit. In contrast, space constraint is satisfied by penalizing segments of the tube that move outside the predefined region in the scene.

The following are the major contributions of this work:

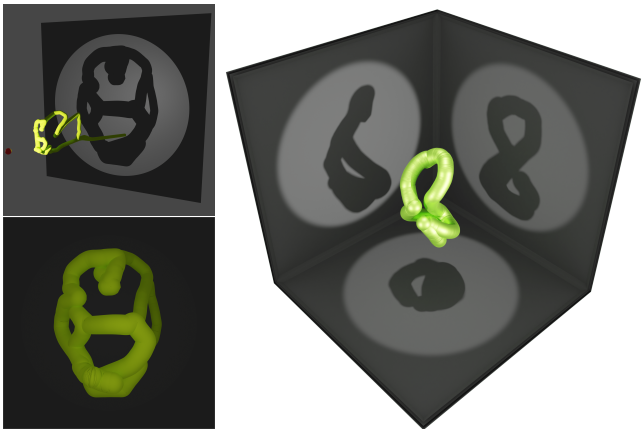
1. We develop an end-to-end differentiable framework for finding a 3D art structure based on tubular knots, which is perceived as the desired target image upon viewing from a particular viewpoint. Our approach also works in the multi-view setting, where we have multiple target images with corresponding viewpoint configurations.
2. We design an efficient differentiable rendering algorithm that, given an arbitrary camera configuration, renders the silhouette image of a tube of desired thickness along a given knot embedding in 3D, as viewed from the camera.
3. We propose a differentiable neural representation of knot embeddings through Invertible Neural Networks (INN), facilitating effective knot search using gradient-based optimization.
4. We propose several loss functions to ensure the 3D art structure is physically realizable.

To the best of our knowledge, we are the first to formalize the problem of knot-based inverse perceptual art in the optimization setting and propose a fully differentiable inverse rendering-based solution.

2. Related Work

Shadow Art. The inverse process of reconstructing or learning 3D structures that cast target shadows under specific lighting conditions has been previously explored. [MP09] deform the input target images to find a consistent shadow hull while minimizing the induced distortions, utilising volumetric representation. [STR22] use a differentiable rendering optimization framework for finding the 3D structure, using mesh representation. Our work, in contrast, focuses on knot-based tubular 3D structures and utilizes neural representations to represent the knot. One of the drawbacks in the previous works is that the predicted 3D structure is not guaranteed to be physically realisable, limiting its practical applicability. In contrast, our work emphasises physical realizability, making it relevant for real-world applications. In previous works on shadow art, the 3D structures are generally complex and visually difficult to grasp. However, using knot-based tubes ensures that the final shape is intuitive and easier to understand. Our approach introduces a novel form of shadow art, where both the shadow and the 3D structure are appealing and intuitive. Using knots thus makes the artwork more accessible and engaging for viewers, enriching the overall aesthetic experience.

Inverse Rendering and Differentiable Rendering. Several existing inverse rendering methods recover the scene parameters from images using supervised learning [BXS*20], [YKM*20]



(a) Single Viewer (b) Multiple Viewer

Figure 2: *(a) Result on Iron Man Image. The top image shows the shadow cast on the wall, whereas the bottom image shows the view from the red dot. (b) Given the orthographic projections of Viviani's curve as multiple target images, our method learns a knot embedding whose perspective projections approximate the target images when viewed from the respective viewing locations.*

that require multi-view or single-view images either during training or inference. Recently, unsupervised and weakly supervised methods [NMOG20], [HCLZ20], [WMW*21], [WWR22] for inverse rendering have been proposed that focus on de-rendering either general or specific object types. Differentiable Rendering algorithms depend largely on the underlying 3D data representation, either explicit ones, such as, voxels [YYY*16], [TEM18], meshes [LB14], [LLCL19] and point clouds [ID18], [CHLZ21], [YSW*19], [LKL18], [HCLZ20], [BXS*20] or implicit ones [LZP*20], [NMOG20], [YKM*20] that utilize generative models which extract the occupancy probabilities, distances, and transparencies, with respect to the surface. Our work uses unsupervised learning to focus on the inverse rendering of knot-based 3D tubes from ground truth silhouette images.

Knot Rendering and Optimization. Intersection algorithms for ray tracing with curves [Van85], [NO02], [Res17] have been studied previously with a focus on parametric curves. J.J. Van Wijk [Van85] formally defined and described the ray tracing intersection with the resulting shape when a sphere of changing radius is swept along a 3D trajectory. Rendering algorithms for knots have been proposed in KnotPlot [Sch98], which is a knot visualization and manipulation tool. However, these rendering methods are not differentiable. We have developed a differentiable knot renderer that facilitates search in the knot space to optimize an objective defined over the rendered image. Several existing knot energies in the literature for knot optimization, such as Möbius Energy [KS98] and Tangent-Point function [YSC21], [BO95] ensure the validity of the knot by penalizing self-intersections. Our approach instead has the guarantee that the represented knot is free of self-intersection, enabling optimization without the need for a penalty term.

Template-based Shape Deformation. Graph-based convolution networks [WZL*18], MLPs [TKG20], [GFK*18], and Neural ODEs [PKGF21], [Gup20] have been employed before to deform a simple genus-zero shape like a sphere or an ellipsoid into a shape of arbitrary complexity. The use of Invertible Neural Networks (INNs) [DKB14] that utilize diffeomorphic [Gup20] or homeomorphic [PKGF21] deformations have shown remarkable results. In our work, we utilize INNs to represent homeomorphic deformations of a template knot. Preserving the topology guarantees that there are no self-intersections in the generated knot, given that the template is self-intersection-free.

3. Proposed Approach

3.1. Problem Statement

Let $\mathcal{C} = (t, \theta, f, z_-, z_+, W, H)$ represent a pinhole camera model. Here $t = (t_x, t_y, t_z)$, $\theta = (\theta_x, \theta_y, \theta_z)$ and $f = (f_x, f_y, f_z)$ denote the camera's location, orientation, and focal lengths, respectively, with respect to the world coordinate frame. The distance of the near-clipping plane and the far-clipping plane from the camera location is denoted by z_- and z_+, respectively. W and H denote the width and height of the image rendered by the camera. From here on, unless specified otherwise, we assume that all the points are defined with respect to the camera coordinate frame.

Let $\mathcal{D} = \{(x,y,z) \mid z = f_z, |x| \le f_x, |y| \le f_y\}$ denote the image plane and $\mathcal{F} = \{(x,y,z) \mid \frac{f_z}{z}(x,y,z) \in \mathcal{D},\ z_- \le z \le z_+\}$ denote the viewing frustum . The image pixel grid is denoted by $G = \mathcal{Z}_W \times \mathcal{Z}_H$, where $\mathcal{Z}_N = \{z \in \mathbb{Z} \mid 0 \le z \le N-1\}$, then $\mathcal{Q} = \{(f_x \cdot (\frac{2i}{W-1} - 1), f_y \cdot (\frac{2j}{H-1} - 1)) \mid (i,j) \in G\}$, represents the coordinates of the pixels on the image plane.

A knot is defined as a topological embedding of the circle in $\mathbb{R}^3$. Let's say we choose a template knot having parametric representation $\mathcal{K} : [0,1] \to \mathbb{R}^3$, with $\mathcal{K}(0) = \mathcal{K}(1)$. Let $\mathcal{H} : \mathbb{R}^3 \to \mathbb{R}^3$ be a homeomorphism. Then $\hat{\mathcal{K}} = \mathcal{H} \circ \mathcal{K}$ describes a smooth deformation of the template knot. Let $\mathcal{H}_\phi$ be a family of homeomorphisms, parametrized by $\phi \in \Phi$. Then, $\mathcal{K}_\phi = \mathcal{H}_\phi \circ \mathcal{K}$ represents a family of knots, parametrized by $\phi \in \Phi$. Given $0 \le s_1 \le s_2 \le 1$, the arc-length of the knot segment of $\mathcal{K}_\phi$ between s_1 and s_2 is given by $\ell_{\mathcal{K}_\phi}(s_1, s_2) = \int_{s_1}^{s_2} |\mathcal{K}'_\phi(s)| ds$. Let the total length of the knot be denoted as $L_{\mathcal{K}_\phi}$ given by $L_{\mathcal{K}_\phi} = \ell_{\mathcal{K}_\phi}(0,1)$. Let $\mathcal{T}(\mathcal{K}_\phi, r) = \bigcup_{s \in [0,1]} \mathcal{B}_r(\mathcal{K}_\phi(s))$ represent a tube having thickness r, which is obtained by sweeping a ball of radius r along the knot $\mathcal{K}_\phi$. Here $\mathcal{B}_r(p)$ denotes a ball of radius r centered at point p.

Let $\mathcal{R} : \Phi \times \mathbb{R}_{\ge 0} \times \mathcal{C}^* \to \mathcal{I}^*$ denote the rendering function, where $\mathcal{C}^*$ denotes the space of all pinhole cameras and $\mathcal{I}^* = [0,1]^G$ denotes the space of all grayscale images defined on grid $\mathcal{G}$. Given $\phi \in \Phi$, $r \ge 0$ and a pinhole camera $\mathcal{C} \in \mathcal{C}^*$, $\hat{\mathcal{I}} = \mathcal{R}(\phi, r, \mathcal{C})$ represents the rendered image of tube $\mathcal{T}(\mathcal{K}_\phi, r)$ as observed by the camera $\mathcal{C}$. Given a target image $\mathcal{I} \in \mathcal{I}^*$, the objective is defined as

$$\begin{aligned} \min_{\phi \in \Phi} \quad & ||\mathcal{I} - \mathcal{R}(\phi, r, \mathcal{C})||_2^2 \quad \text{s.t.} \\ & \Lambda(\mathcal{T}(\mathcal{K}_\phi, r)) = \emptyset \\ & B_{\mathcal{K}_\phi} \le B_0 \\ & \mathcal{T}(\mathcal{K}_\phi, r) \subseteq \Omega_0 \\ & L_{\mathcal{K}_\phi} \le L_0 \end{aligned} \tag{1}$$

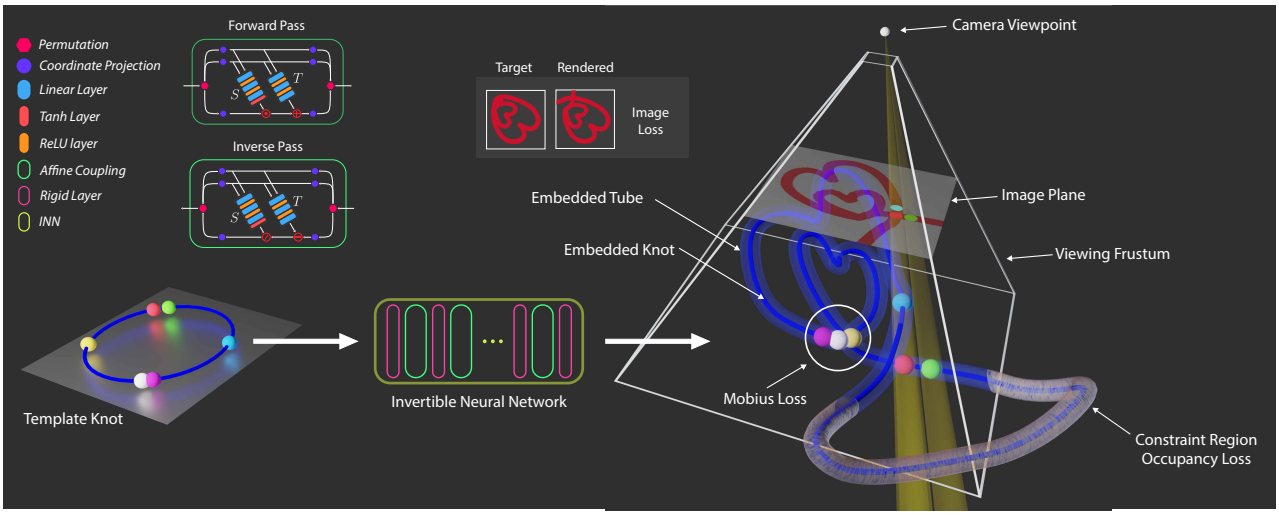


Figure 3: *Illustration showing the pipeline of our proposed framework. We forward pass a template knot through an INN consisting of an alternating sequence of affine coupling blocks and rigid transformation layers to obtain the knot representation. Given the camera viewpoint, we propose a differentiable renderer that generates the image of the knot (shown on right). Using gradient-based optimization, we find the ideal knot whose rendered image is closest to the target image.*

Here, $\Lambda(\mathcal{T}(\mathcal{K}_\phi, r))$ denotes the self-intersections in the tube, that is, all points in the tube that arise from multiple parameters, when the parametric representation of the tube is considered in the Frenet-Serret frame. $B_{\mathcal{K}_\phi}$ denotes the bending in the knot which is represented as the total integral of the curvature along the knot. $L_{\mathcal{K}_\phi}$ denotes the total length of the knot. Ω_0 represents the constraint region in space inside which the tube is supposed to lie, B_0 is the maximum allowed bending, and L_0 is the maximum allowed length of the knot. Note that our approach does not solve Eq. 1 as a constrained optimization problem. Instead, each constraint is incorporated as an individual term in the loss function, and thus, the constraints are not enforced strictly.

3.2. Parametric Family of Knots

$\mathcal{H} : \mathbb{R}^3 \to \mathbb{R}^3$ is said to be a homeomorphism, if $\mathcal{H}$ is bijective and bicontinuous. We use Invertible Neural Networks (INN) to represent $\mathcal{H}_\phi$, where ϕ denotes the INN parameters. We design an INN consisting of an alternating sequence of Affine Coupling Layers [DSB16] and Rigid transformation layers. The INN is invertible by design, guaranteeing bijectivity of $\mathcal{H}_\phi$ and the INN being a compositional function of continuous layers guarantees the bicontinuity of $\mathcal{H}_\phi$. The INN is explained in detail in Sec. 3.4.

Using an INN-based parametric family of homeomorphisms to search in the space of knots has several advantages. The INN being differentiable facilitates efficient gradient-based optimization for searching the optimal knot. By choosing a template knot $\mathcal{K}$, the use of homeomorphism guarantees that the search space is constrained to knots having the same knot type as that of $\mathcal{K}$. Thus the desired knot type can be fixed by appropriately choosing the template knot. An embedded knot, by definition, is free of self-intersection. If it is ensured that the template used is a valid knot, then it is guaranteed that $\mathcal{K}_\phi$ is a valid knot for any choice of $\phi \in \Phi$. Thus every $\mathcal{K}_\phi$ is free of self-intersection since a homeomorphism preserves the topological properties. Thus, our approach does not require any additional projection step or the use of loss functions to ensure the validity of the knots.

Our approach works for any choice of template knot. It is general in the sense that if we choose open curves instead of closed knots, then we can search in the space of open curves and render open tube structures. However, here we restrict our focus only to knots, specifically, the trivial knot, whose template is represented as $\mathcal{K} : [0,1] \to \mathbb{R}^3, \mathcal{K}(s) = (\cos s, \sin s, 0)$. We demonstrate that even a trivial knot is powerful enough to represent complex images. In Fig. 3, the INN and the affine coupling blocks are illustrated.

3.3. Differentiable Silhouette Rendering of Knots

We propose two algorithms for the rendering function $\mathcal{R}(\phi, r, \mathcal{C})$, both of which being differentiable, facilitate the end-to-end optimization of the objective defined in Eq. 1.

Ellipse Renderer ($\mathcal{R}_\mathcal{E}$). We first randomly sample N points on the template unknot $\mathcal{K}$ to obtain the set $\mathcal{P}_\mathcal{K}$. We forward pass the template points through the INN $\mathcal{H}_\phi$ to obtain points on $\mathcal{K}_\phi$, denoted as the set $\mathcal{P}_{\mathcal{K}_\phi} = \mathcal{H}_\phi(\mathcal{P}_\mathcal{K})$. The sphere of radius r centred at a point $p = (l, m, n) \in \mathcal{P}_{\mathcal{K}_\phi}$ will become an ellipse, $\mathcal{E}_p$, on the image plane. Let $\mathcal{N}_p$ be the cone induced by p such that its apex is at the origin, height is $\|p\|_2$ and radius is r. The projected ellipse $\mathcal{E}_p$ is the intersection of the cone $\mathcal{N}_p$ with the image plane. The equation of $\mathcal{E}_p$ is given by $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, where $A = l^2 - k^2$, $B = 2lm$, $C = m^2 - k^2$, $D = 2f_z ln$, $E = 2f_z mn$, $F = f_z^2(n^2 - k^2)$, and $k^2 = \frac{(l^2+m^2+n^2)^2}{l^2+m^2+n^2+r^2}$. Projecting ellipses about each point in $\mathcal{P}_\mathcal{K}$ on the image plane gives a good approximation of the $\mathcal{T}(\mathcal{K}_\phi, r)$ when N is sufficiently large. However, if N is very small or the distance between consecutive sample points is large, it might result in an undesirable discontinuity in the rendered image. To avoid this, we propose another renderer which has an additional computational cost but is more accurate.

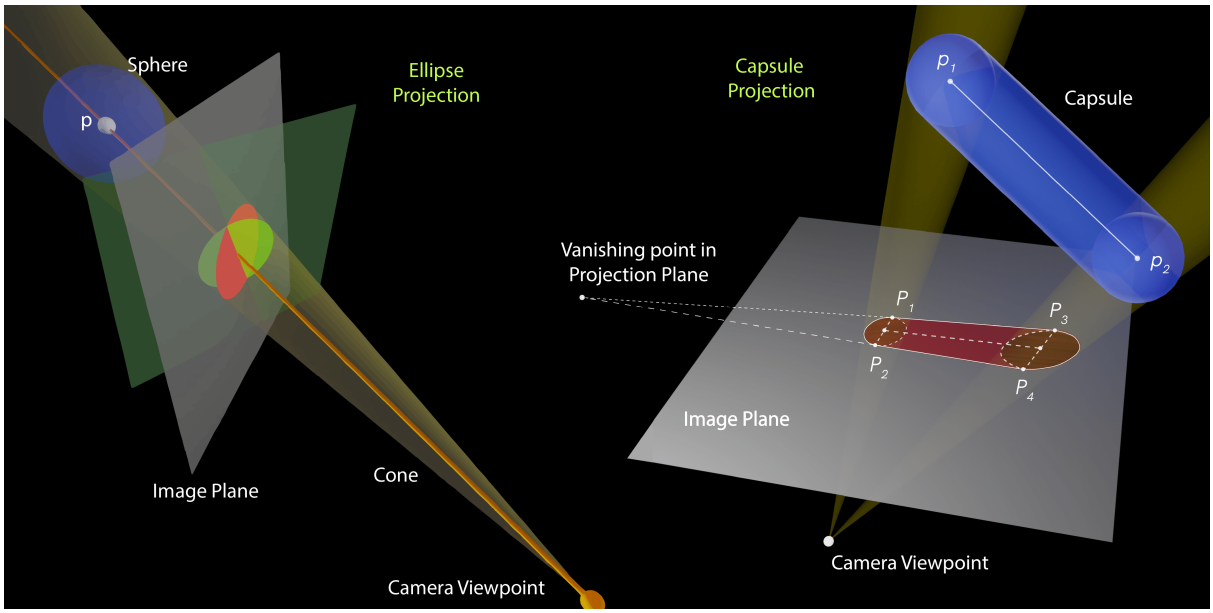


Figure 4: *Illustration of the Ellipse Renderer ($\mathcal{R}_\mathcal{E}$) and the Capsule Renderer ($\mathcal{R}_\mathcal{C}$).*

Capsule Renderer ($\mathcal{R}_\mathcal{C}$). We consider a cylindrical capsule between consecutive points on the knot $\mathcal{K}_\phi$, having spherical ends formed by spheres of radius r centred at those points. Let the capsule have a projection $\mathcal{U}_p$ on the image plane. $\mathcal{U}_p$ is the union of the projections of the two spheres at the ends of the capsule and the projection of the central cylinder plane joining them. The analytical equations and derivations of the projected ellipse and projected capsule have been included in the supplementary material. Let $S_{\mathcal{U}_p} : \mathbb{R}^2 \to \mathbb{R}$ be the signed distance function of the projected capsule $\mathcal{U}_p$. The occupancy function of $\mathcal{U}_p$ on the image plane can then be defined as $\mathcal{O}_{\mathcal{U}_p}(q) = \sigma(\tau \cdot S_{\mathcal{U}_p}(q))$, where σ is the sigmoid function. The parameter τ is the hardness factor, which controls how rapidly the occupancy function changes near the boundary. Let us consider a pixel $q \in \mathcal{Q}$ having grid coordinates (i, j). The pixel value of q, denoted as $\hat{\mathcal{I}}(i, j)$ can be computed as the maximum of the occupancy values across all the projected ellipses, given by, $\hat{\mathcal{I}}(i, j) = \max_{p \in \mathcal{P}_\mathcal{F}} \mathcal{O}_{\mathcal{U}_p}(q)$. The differentiable rendering process is illustrated in Fig. 4.

3.4. Network Architecture

We represent the family of homeomorphisms $\mathcal{H}_\phi$ using an Invertible Neural Network (INN), where ϕ denotes the INN parameters. The network architecture of the INN is illustrated in Fig. 3. The INN consists of an alternating sequence of Rigid Transformation layers and Affine Coupling Layers [DSB16].

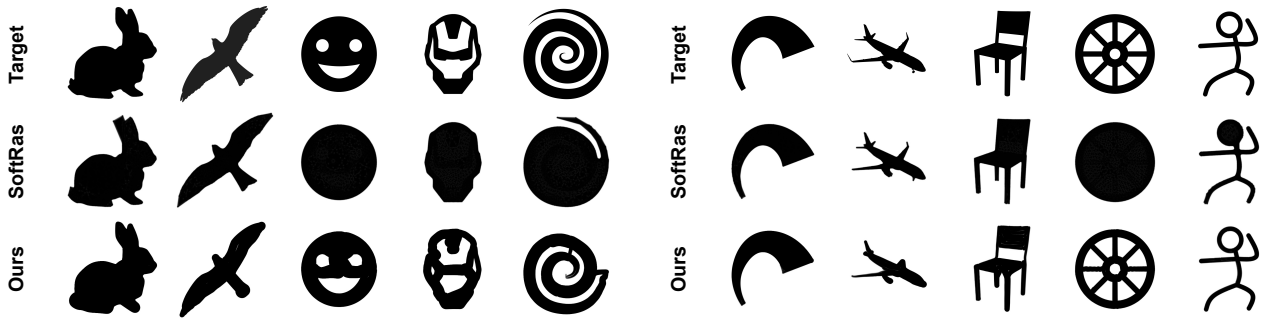
Rigid Transformation layer. This layer consists of a rotation in $\mathbb{R}^3$ along with a translation in $\mathbb{R}^3$, which gets updated during each iteration of the optimization process. Both rotation and translation are invertible transformation. The inverse of a rotation is yet another rotation, and the inverse of a translation is yet another translation. This makes the Rigid Transformation layer an invertible layer with explicitly available inverse.

Affine Coupling layer. Each Affine Coupling layer consists of two Multi-Layer Perceptrons (MLP) : S for scaling and T for translation. It also contains a permutation function σ. Both S and T take two values as input and output a single value, representing the scaling factor and the translational shift, respectively. During the forward pass, given an input $p = (x, y, z)$, it is first permuted by σ to obtain $p' = (u, v, w) = \sigma(p)$. Then p' is modified to get $q' = (u', v', w')$, where $u' = u$, $v' = v$ and $w' = w \cdot S(u, v) + T(u, v)$. Only w is modified using an affine transformation whose scaling and translation factor depend on u and v, whereas u and v are left unmodified. The final output q obtained at the end of the forward pass is given by $q = \sigma^{-1}(q')$. During the inverse pass, given an input $p = (x, y, z)$, it is first permuted by σ^{-1} to obtain $p' = (u, v, w) = \sigma^{-1}(p)$. Then p' is modified to get $q' = (u', v', w')$, where $u' = u$, $v' = v$ and $w' = \frac{w - T(u,v)}{S(u,v)}$. The final output q obtained at the end of the inverse pass is given by $q = \sigma(q')$. It can be seen that the inverse pass is indeed the inverse operation of the forward pass due to the invertibility of the affine function. The permutation σ for each Affine Coupling layer is randomly chosen during initialization and is fixed after that during the entire optimization process.

The Rigid Transformation layers and Affine Coupling Layers are both invertible layers. Since the composition of invertible functions is invertible, our overall network is invertible. Moreover, each component in our network is continuous, and the composition of continuous functions is continuous, making our overall network continuous. Our network $\mathcal{H}_\phi$ is thus bijective and bicontinuous and represents a homeomorphism for each value of ϕ.

3.5. Embedding Search of Knots

Given the objective defined in Eq.1, the rendering function and the differentiable representation of knots enable us to use gradient descent-based methods for optimization. The objective



(a) Ours w/ constraints, SoftRas w/o constraints (b) Ours w/o constraints, SoftRas w/o constraints

Figure 5: *Rendering of the 3D tube embeddings obtained using our method (bottom row) and the rendering of the sphere mesh deformation using SoftRas [LLCL19] (middle row) for the corresponding ground truth silhouettes (top row). In (a), the physical realisability constraints are applied only on our method and not on SoftRas method. In (b), results of both methods are without the constraints.*

and the constraints are represented in terms of minimized loss functions. First, we define some preliminary concepts needed to describe the loss functions. Let $p_1 = \mathcal{K}_\phi(s_1)$ and $p_2 = \mathcal{K}_\phi(s_2)$ denote two points on the knot, where $0 \leq s_1 \leq s_2 \leq 1$. Let $d(p_1,p_2) = \|p_1 - p_2\|_2$ denote the Euclidean distance between the two points. Let $g(p_1,p_2) = \min(\ell_{\mathcal{K}_\phi}(s_1,s_2),\ L_{\mathcal{K}_\phi} - \ell_{\mathcal{K}_\phi}(s_1,s_2))$ denote the geodesic distance between the two points. The geodesic distances in our approach is computed using polyline approximation. The loss functions are illustrated in Fig. 3.

Image Loss ($\mathcal{L}_I$): This loss is used to make the rendered image $\hat{\mathcal{I}}$ similar to the target image $\mathcal{I}$ and is given by

$$\mathcal{L}_I = \left\| \mathcal{I} - \hat{\mathcal{I}} \right\|_2^2 \tag{2}$$

Length Loss ($\mathcal{L}_L$): As the radius of the knot is constant throughout, the material cost of the tube is proportional to the length of the knot. The material cost budget corresponds to a maximum allowed length L_0. The length loss adds a penalty whenever knot length exceeds L_0, given by

$$\mathcal{L}_L = \max(L_{\mathcal{K}_\phi} - L_0, 0) \tag{3}$$

The set $\mathcal{P}_\mathcal{K}$ is obtained by sampling a random ring-graph from the circle, which is then forward passed through the INN $\mathcal{H}_\phi$ to obtain $\mathcal{P}_{\mathcal{K}_\phi}$. The length of edges in the deformed ring is added up to obtain the knot length $L_{\mathcal{K}_\phi}$.

Möbius Loss ($\mathcal{L}_M$): Even though the knot $\mathcal{K}_\phi$ is guaranteed to have no self-intersection in our representation, the tube $\mathcal{T}(\mathcal{K}_\phi, r)$ can have self-intersections. In order to penalize tube self-intersections, we define a loss based on the Möbius Energy [KS98]. The Möbius Energy between two knot points is defined as $\mathcal{M}(u,v) = \frac{1}{d(u,v)^2} - \frac{1}{g(u,v)^2}$. We define Möbius loss on a randomly sampled set $\mathcal{B}_M = \{(u,v) \mid u,v \in \mathcal{P}_{\mathcal{K}_\phi}, u \neq v\}$ as

$$\mathcal{L}_M = \frac{1}{|\mathcal{B}_M|} \left(\sum_{(u,v)\in\mathcal{B}_M} \mathcal{M}(u,v) \cdot \max(2r - d(u,v), 0) \right) \tag{4}$$

This loss penalizes those pairs of points for which the geodesic distance is much larger than the Euclidean distance, and the Euclidean distance is less than twice the radius (implying tube self-intersection).

Occupancy Loss ($\mathcal{L}_R$): An artist or an architect, while realizing such a tube structure might have space constraints, would like the structure to be bounded within a predefined constraint region Ω_0. Let $S_{\Omega_0} : \mathbb{R}^3 \to \mathbb{R}$ be the signed distance function of Ω_0, which is positive for the inside points and negative for the outside points. The occupancy Loss penalizes the tube points that go outside the constraint region Ω_0 and is defined as

$$\mathcal{L}_R = \sum_{p\in\mathcal{P}_{\mathcal{K}_\phi}} \max(r - S_{\Omega_0}(p), 0) \tag{5}$$

This loss adds a penalty whenever a sphere of radius r about a knot point does not completely lie within Ω_0.

Bending Loss ($\mathcal{L}_R$): Based on the material to be used for the tube, there might be restrictions on the extent to which the tube can physically bend. Let B_0 be the maximum bending that a tube can physically attain at any point. The bending loss penalizes those points on the knot whose squared curvature, given by $B_{\mathcal{K}_\phi}(p)$ exceeds the maximum allowed bending B_0 and is defined as

$$\mathcal{L}_B = \sum_{p\in\mathcal{P}_{\mathcal{K}_\phi}} \max(B_{\mathcal{K}_\phi}(p) - B_0, 0) \tag{6}$$

The total loss function is a weighted sum given as $\mathcal{L} = w_{img} \cdot \mathcal{L}_I + w_{len} \cdot \mathcal{L}_L + w_{mob} \cdot \mathcal{L}_M + w_{occ} \cdot \mathcal{L}_R + w_{bend} \cdot \mathcal{L}_B$. The hyperparameters used in our framework, the computing infrastructure used to run the experiments and the details of the INN architecture are included in the supplementary material.

4. Results and Discussion

We create a test bed of target silhouette images and demonstrate in Sec. 4.1 that our method obtains impressive results. We also conduct experiments (Sec. 4.2 and 4.3) which indicate the effectiveness of the proposed loss functions and the renderer. A real-world demonstration is shown in Sec. 4.4.

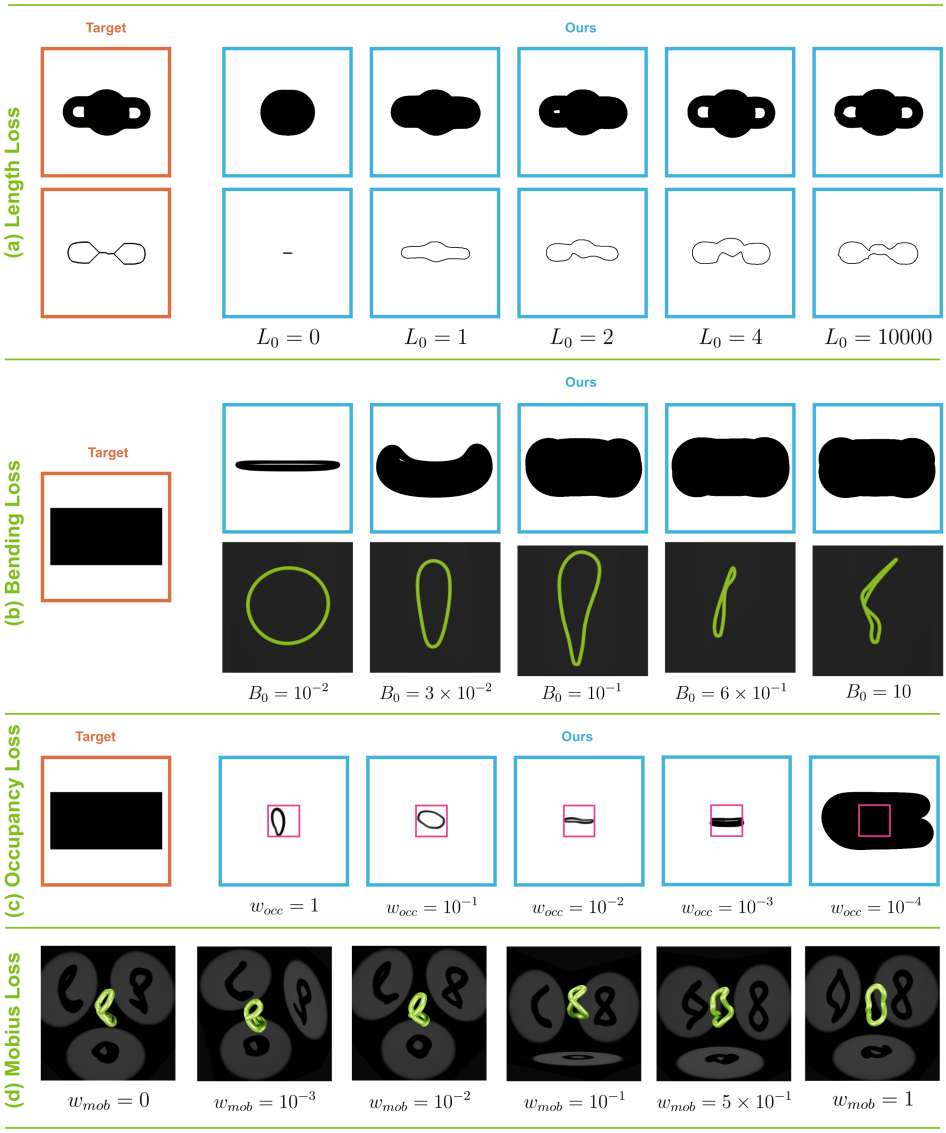


Figure 6: *Ablation experiments of proposed loss functions.*

4.1. Inverse perceptual art results

The proposed framework was tested on silhouette images having varying levels of complexity, under different settings.

Single Viewer, Single Knot. In this problem setting, there is a single viewer configuration with a single target image. The objective is to find the embedding of a single knot whose projected image resembles the target image as illustrated in Fig. 2(a). To the best of our knowledge, no existing works in the literature address the problem of knot-based inverse perceptual art (mentioned in Eq. 1). This prohibits us from performing exhaustive comparative analysis with other methods.

However, the work of SoftRas [LLCL19], which, instead of tube embedding, uses sphere mesh deformation, closely resembles our work, with whom we compare our results, as shown in Fig. 5. In SoftRas, due to the deformation of a sphere mesh, there is difficulty in forming image shapes with several holes. In contrast our method is able to generate complex shapes with complex topologies. When the physical constraints imposed are too strong, obtaining complex shadow images could be physically impossible. This is not a limitation of our approach but an inherent physical limitation. To show that our method is highly expressive, in Fig. 5(b), we remove the physical constraints and demonstrate that complex shadow shapes can be generated through complicated embedding of the knot.

Single Viewer, Multiple Knots. In this problem setting, there is a single viewing location and multiple knots. The objective is to find the embedding of all the knots so that the projected image resembles the target image. Fig. 1(a) shows the result of a single light source casting shadows from multiple knots. And when a viewer is placed at the location, all the knots together form a single image.

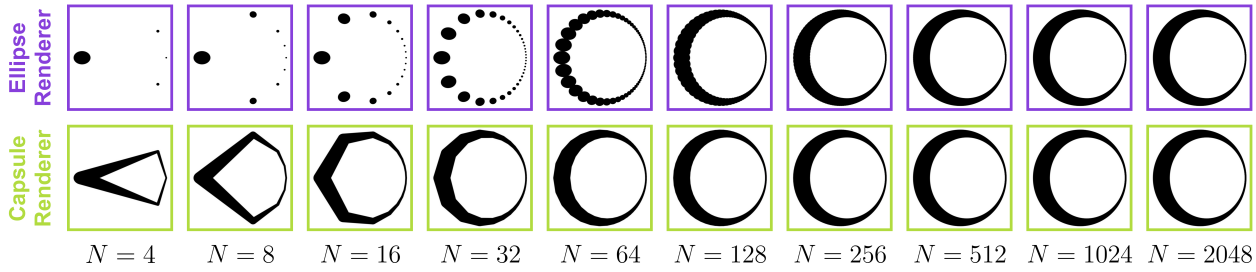


Figure 7: *Renderings for varying number of N.*

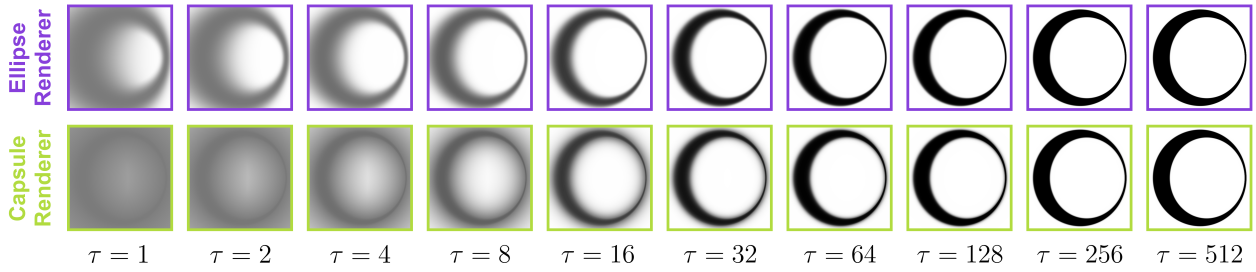


Figure 8: *Renderings for varying value of* τ.

Multiple Viewer, Single Knots. In this case, there are multiple viewing configurations and a target image corresponding to each view. The embedding of a single knot needs to be searched, whose projected image from each view matches the respective target image. As demonstrated in Fig. 2(b) and Fig. 6(d), there are three target silhouette images representing projections on three perpendicular image planes. These three images are the orthographic projections of Viviani's curve. Viviani's curve has a self-intersection in the middle and is thus not a knot. Instead, our method learns a knot embedding that is physically realizable and free of self-intersection, whose perspective projections on the image plane resemble the orthographic projections of Viviani's curve.

Single Viewer (single target video), Spatio-temporal Knots. In this scenario, given a temporal sequence of target images, the objective is to find the temporal sequence of embeddings of a single knot, so that at a given time instant, the rendering of the knot embedding is similar to the target image at the corresponding time instant. In this case, additional regularization is added to ensure smooth spatio-temporal deformation of the knot. In Fig. 1(b), all the knots are constrained to be smooth temporal deformations of each other. The animation depicting this temporal deformation of the knots creates the perception of a person dancing. Thus, in this case, perceptual art also has a temporal aspect. The animation is available in the supplementary material.

Our approach learns knot embeddings whose projected images closely resemble the target silhouette images in all the scenarios, proving the effectiveness of our approach. The knot-based tube embeddings when rendered in 3D, were also found to satisfy the physical constraints specified. More results and the time evolution of the 3D knot embeddings during the optimization process, are available in the supplementary material.

4.2. Ablation Study of Loss Functions

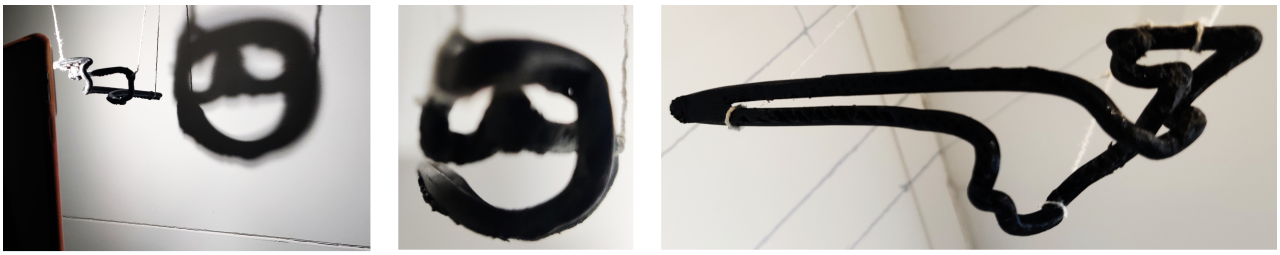
We conduct experiments to investigate the importance of each loss function proposed in our approach. Since image loss is the primary objective, keeping it weight fixed at $w_{img} = 1$, we vary the other loss function's hyperparameters and observe the effects.

Length Loss. Fig. 6(a) shows the rendered image and its skeleton corresponding to a given ground truth for different considered values of the maximum allowed length, L_0. For smaller values of L_0, the proposed algorithm tries to render an image as close to the ground truth, while respecting the budget constraint set by L_0. When large values of L_0 are considered, 3D knot embedding is given more freedom to deform in space and successfully generates the desired image.

Bending Loss. In Fig. 6(b), given the target image of a rectangle, the maximum allowed bending B_0 is varied. For low values of B_0, the bending loss forces the knot to have the least amount of curvature, resulting in the symmetric circular shape. Only when a higher value of bending is allowed, the knot starts to deform, and eventually for a sufficiently large value of B_0, deforms to result in an image that closely resembles a rectangle.

Occupancy Loss. In Fig. 6(c), the target image is a rectangle, and the pink sub-region within the image indicates the constraint region. When the occupancy loss weight w_{occ} is higher, the knot is forced to stay confined with the constrained region. Only when w_{occ} is made sufficiently small, do the other losses start to dominate and defy the region constraints. This indicated that the occupancy loss successfully forces the knot to stay within a user-defined region.

Möbius Loss. The effect of Möbius loss is investigated in Fig. 6(d), in the multi-view, multi-target setting. The target images are chosen in such a way that the optimal solution necessitates a self-intersection in the tube. For smaller values of w_{mob}, we see that



(a) (b) (c)

Figure 9: *Real-world demonstration of our approach.*

the Möbius loss is not given significance and the resulting configuration of tube indeed has self-intersections. Only when the weight w_{mob} is increased sufficiently, does the Möbius loss component become strong enough to avoid self-intersections. This demonstrates that Möbius loss indeed avoids self-intersection in the tube.

4.3. Differentiable Renderer Experiments

We have proposed two methods for rendering the tube: Ellipse Renderer ($\mathcal{R}_{\mathcal{E}}$) and Capsule Renderer ($\mathcal{R}_{\mathcal{C}}$). In this section, we discuss the scenario of when one would be preferred over the other. In Fig. 7, for a fixed circle in 3D, we vary the number of sample points N and see its effect on $\mathcal{R}_{\mathcal{E}}$ and $\mathcal{R}_{\mathcal{C}}$. When N is large, both produce the desired results. However, when N is small, such as $N = 64$, $\mathcal{R}_{\mathcal{E}}$ generates a poor approximation, whereas $\mathcal{R}_{\mathcal{C}}$ produces a much better approximation of the actual image. This demonstrates that $\mathcal{R}_{\mathcal{C}}$ is superior than $\mathcal{R}_{\mathcal{E}}$, when N is small. Thus $\mathcal{R}_{\mathcal{C}}$ is particularly advantageous if the INN $\mathcal{H}_\phi$ has many layers, in which case querying points from the knot (forward pass) is computationally expensive. If the forward pass through INN $\mathcal{H}_\phi$ is computationally cheap, then querying points from the knot is inexpensive, making the computations involved in capsule rendering the main bottleneck. In such a case, $\mathcal{R}_{\mathcal{E}}$ would be preferred over $\mathcal{R}_{\mathcal{C}}$.

In Fig. 8, for both $\mathcal{R}_{\mathcal{E}}$ and $\mathcal{R}_{\mathcal{C}}$, we vary the hardness factor τ. Making τ large generates a rendering close to the desired image. But large τ also results in exploding gradients making the optimization process numerically unstable. For τ extremely small, the optimization is stable, but the rendered image is less sharp, producing a knot configuration whose actual image is different from the desired target. In our experiments, we use the value of $\tau = 100$ and observe that it gives good results, while also being numerically stable.

4.4. Real World Demonstration

Given the target image of a smiley (third column in Fig. 5), our approach generates a tube structure, which we fabricate using a 3D printer, shown in Fig. 9(c). We cast light on this structure (Fig. 9(a)) whose shadow resembles a smiley and when viewed from the specified location (Fig. 9(b)) it appears like a smiley.

4.5. Limitations

Ideally, for fabrication, we would take a deformable tube with a fixed radius, and then deform and arrange it in 3D space, and close both ends. The arrangement in 3D space would be guided by our algorithm. Then, the shadow cast by the 3D structure would resemble the target image. There could be other similar fabrication methods to realize the shadow art mentioned in our work. Note that, in Sec. 4.4, we have used 3D printing as the fabrication process, which is not the ideal choice of fabrication, given our problem statement and the proposed loss functions. The sole purpose behind using 3D printing was to demonstrate that the 3D object generated by our approach indeed casts a shadow that matches the target image while also satisfying the physical constraints.

5. Conclusion and Future Work

In this work, we have proposed an end-to-end differentiable optimization framework for searching 3D knot-based tube embedding that, when projected onto the image plane of a user-specified camera, forms a silhouette image similar to the input target image. In order to represent the knot embeddings, we have used invertible neural networks (INNs) as a parametrized family of homeomorphisms and created a differentiable silhouette renderer that renders the tube. We also present several loss functions to ensure that the generated tubular structure is physically realizable. We present the power of a simple unknot being used as the template knot and how its deformation through homeomorphism can be derendered even in the case of complex images. The inverse knot rendering from single-view and multiple-view sets of silhouette images is represented for any given camera attribute. Even though we have used the unknot as the template shape, our approach is general enough and is compatible with shapes other than an unknot. Exploring other template shapes, such as different knot types and embedded tree graphs, would be interesting. In future work, we aim to experiment with other formulations of inverse perceptual art. For instance, the viewer configuration and radius are fixed for the desired target image in the current setting. A variant in which the viewer configurations are left unspecified would be an interesting problem. In such a variant, the proposed approach will also find the viewer configuration along with the knot embedding and the thickness of the tube. As future work, other types of differentiable silhouette rendering techniques could be explored, such as discretizing the sphere and cylinder into meshes and utilizing a highly efficient differentiable mesh rendering framework. The current differentiable renderer is silhouette-based. We propose to develop realistic shading-based renderers for knots, which can help derender real images of knots, as a future work.

6. Acknowledgement

This work is supported by Prime Minister Research Fellowship (PMRF2122-2557) and Jibaben Patel Chair in Artificial Intelligence.

References

[BO95] BUCK, GREGORY and ORLOFF, JEREMEY. "A simple energy function for knots". *Topology and its Applications* 61.3 (1995), 205–214 3.

[BXS*20] BI, SAI, XU, ZEXIANG, SUNKAVALLI, KALYAN, et al. "Deep 3d capture: Geometry and reflectance from sparse multi-view images". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 5960–5969 2, 3.

[CHLZ21] CHEN, CHAO, HAN, ZHIZHONG, LIU, YU-SHEN, and ZWICKER, MATTHIAS. "Unsupervised learning of fine structure generation for 3d point clouds by 2d projections matching". *Proceedings of the ieee/cvf international conference on computer vision*. 2021, 12466–12477 3.

[DKB14] DINH, LAURENT, KRUEGER, DAVID, and BENGIO, YOSHUA. "Nice: Non-linear independent components estimation". *arXiv preprint arXiv:1410.8516* (2014) 3.

[DSB16] DINH, LAURENT, SOHL-DICKSTEIN, JASCHA, and BENGIO, SAMY. "Density estimation using real nvp". *arXiv preprint arXiv:1605.08803* (2016) 4, 5.

[Evi24] EVILSIZER, DAN. *The Art of the Knot - The Art Of The Knot — theartoftheknot.com*. https://theartoftheknot.com/. [Accessed 03-06-2024]. 2024 2.

[GFK*18] GROUEIX, THIBAULT, FISHER, MATTHEW, KIM, VLADIMIR G, et al. "A papier-mâché approach to learning 3d surface generation". *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, 216–224 3.

[Gup20] GUPTA, KUNAL. *Neural mesh flow: 3d manifold mesh generation via diffeomorphic flows*. University of California, San Diego, 2020 3.

[HCLZ20] HAN, ZHIZHONG, CHEN, CHAO, LIU, YU-SHEN, and ZWICKER, MATTHIAS. "DRWR: A differentiable renderer without rendering for unsupervised 3D structure learning from silhouette images". *arXiv preprint arXiv:2007.06127* (2020) 3.

[ID18] INSAFUTDINOV, ELDAR and DOSOVITSKIY, ALEXEY. "Unsupervised learning of shape and pose with differentiable point clouds". *Advances in neural information processing systems* 31 (2018) 3.

[KS98] KUSNER, ROBERT B and SULLIVAN, JOHN M. "Möbius-invariant knot energies". *Ideal knots* 19 (1998), 315–352 3, 6.

[Lar23] LARRY KAGAN, LONSDALE GALLERY. *Larry Kagan's art exhibition at the Lonsdale Gallery*. https://lonsdalegallery.com/artists/larry-kagan/. Jan. 2023 2.

[LB14] LOPER, MATTHEW M and BLACK, MICHAEL J. "OpenDR: An approximate differentiable renderer". *European Conference on Computer Vision*. Springer. 2014, 154–169 3.

[LKL18] LIN, CHEN-HSUAN, KONG, CHEN, and LUCEY, SIMON. "Learning efficient point cloud generation for dense 3d object reconstruction". *proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1. 2018 3.

[LLCL19] LIU, SHICHEN, LI, TIANYE, CHEN, WEIKAI, and LI, HAO. "Soft Rasterizer: A Differentiable Renderer for Image-Based 3D Reasoning". *Proceedings of the IEEE/CVF ICCV*. Oct. 2019 3, 6, 7.

[LZP*20] LIU, SHAOHUI, ZHANG, YINDA, PENG, SONGYOU, et al. "Dist: Rendering deep implicit signed distance function with differentiable sphere tracing". *Proceedings of the IEEE/CVF Conference on CVPR*. 2020, 2019–2028 3.

[Med23] MEDIA, BLAUWEISS. *Larry Kagan: Shape & Shadow*. https://www.youtube.com/watch?v=dSdkrzkcu20. Jan. 2023 2.

[MK96] MURASUGI, KUNIO and KURPITA, BOHDAN. *Knot theory and its applications*. Springer, 1996 2.

[MP09] MITRA, NILOY J and PAULY, MARK. "Shadow art". *ACM Transactions on Graphics* 28.CONF (2009), 156–1 2.

[NMOG20] NIEMEYER, MICHAEL, MESCHEDER, LARS, OECHSLE, MICHAEL, and GEIGER, ANDREAS. "Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 3504–3515 3.

[NO02] NAKAMARU, KOJI and OHNO, YOSHIO. "Ray Tracing for Curves Primitive". *The 10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002, WSCG 2002, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 4-8, 2002*. 2002, 311–316. URL: http://wscg.zcu.cz/wscg2002/Papers%5C_2002/A83.pdf 3.

[PKGF21] PASCHALIDOU, DESPOINA, KATHAROPOULOS, ANGELOS, GEIGER, ANDREAS, and FIDLER, SANJA. "Neural parts: Learning expressive 3d shape abstractions with invertible neural networks". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 3204–3215 3.

[Res17] RESHETOV, ALEXANDER. "Exploiting Budan-Fourier and Vincent's theorems for ray tracing 3D Bézier curves". *Proceedings of High Performance Graphics*. Association for Computing Machinery, 2017, 1–11 3.

[Sch98] SCHAREIN, ROBERT GLENN. "Interactive topological drawing". PhD thesis. University of British Columbia, 1998 3.

[STR22] SADEKAR, KAUSTUBH, TIWARI, ASHISH, and RAMAN, SHANMUGANATHAN. "Shadow art revisited: a differentiable rendering based approach". *Proceedings of the IEEE/CVF WACV*. 2022, 29–37 2.

[TEM18] TULSIANI, SHUBHAM, EFROS, ALEXEI A, and MALIK, JITENDRA. "Multi-view consistency as supervisory signal for learning shape and pose prediction". *Proceedings of the IEEE Conference on CVPR*. 2018, 2897–2905 3.

[TKG20] TULSIANI, SHUBHAM, KULKARNI, NILESH, and GUPTA, ABHINAV. "Implicit mesh reconstruction from unannotated image collections". *arXiv preprint arXiv:2007.08504* (2020) 3.

[Van85] VAN WIJK, JARKE J. "Ray tracing objects defined by sweeping a sphere". *Computers & Graphics* 9.3 (1985), 283–290 3.

[WMW*21] WU, SHANGZHE, MAKADIA, AMEESH, WU, JIAJUN, et al. "De-rendering the World's Revolutionary Artefacts". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 6338–6347 3.

[WWR22] WIMBAUER, FELIX, WU, SHANGZHE, and RUPPRECHT, CHRISTIAN. "De-rendering 3D Objects in the Wild". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, 18490–18499 3.

[WZL*18] WANG, NANYANG, ZHANG, YINDA, LI, ZHUWEN, et al. "Pixel2mesh: Generating 3d mesh models from single rgb images". *Proceedings of the European conference on computer vision (ECCV)*. 2018, 52–67 3.

[YKM*20] YARIV, LIOR, KASTEN, YONI, MORAN, DROR, et al. "Multiview neural surface reconstruction by disentangling geometry and appearance". *Advances in Neural Information Processing Systems* 33 (2020), 2492–2502 2, 3.

[YSC21] YU, CHRIS, SCHUMACHER, HENRIK, and CRANE, KEENAN. "Repulsive curves". *ACM Transactions on Graphics (TOG)* 40.2 (2021), 1–21 3.

[YSW*19] YIFAN, WANG, SERENA, FELICE, WU, SHIHAO, et al. "Differentiable surface splatting for point-based geometry processing". *ACM Transactions on Graphics (TOG)* 38.6 (2019), 1–14 3.

[YYY*16] YAN, XINCHEN, YANG, JIMEI, YUMER, ERSIN, et al. "Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision". *Advances in neural information processing systems* 29 (2016) 3.