NeuralFur: Animal Fur Reconstruction From Multi-View Images

Vanessa Sklyarova[1,2*]   Berna Kabadayi[1,4*]   Anastasios Yiannakidis[1]   Giorgio Becherini[1]

Michael J. Black[1]   Justus Thies[1,3]

[1]Max Planck Institute for Intelligent Systems   [2]ETH Zürich
[3]Technical University of Darmstadt   [4]University of Tübingen

Figure 1. From multi-view images, *NeuralFur* reconstructs detailed geometries of animals with a mesh-based body and strand-based fur. The reconstructions can be integrated in computer graphics frameworks for simulation and rendering with artist-defined colors.

Abstract

Reconstructing realistic animal fur geometry from images is a challenging task due to the fine-scale details, self-occlusion, and view-dependent appearance of fur. In contrast to human hairstyle reconstruction, there are also no datasets that can be leveraged to learn a fur prior for different animals. In this work, we present a first multi-view-based method for high-fidelity 3D fur modeling of animals using a strand-based representation, leveraging the general knowledge of a vision language model. Given multi-view RGB images, we first reconstruct a coarse surface geometry using traditional multi-view stereo techniques. We then use a vision language model (VLM) system to retrieve information about the realistic length structure of the fur for each part of the body. We use this knowledge to construct the animal's furless geometry and grow strands atop it. The fur reconstruction is supervised with both geometric and photometric losses computed from multi-view images. To mitigate orientation ambiguities stemming from the Gabor filters that are applied to the input images, we additionally utilize the VLM to guide the strands' growth direction and their relation to the gravity vector that we

incorporate as a loss. With this new schema of using a VLM to guide 3D reconstruction from multi-view inputs, we show generalization across a variety of animals with different fur types. For additional results and code, please refer to https://neuralfur.is.tue.mpg.de.

1. Introduction

Animals frequently appear as characters in animated films. The realism of these characters relies on strand-based models of the fur, which are ideal for physics simulation and provide high realism. Unfortunately, the manual creation of fur grooms for animals is a time-consuming process for experienced artists [47]. In this work, we study the automatic generation of fur groom geometry, which can be directly applied to background characters or used as an initialization for hero-level assets. While the automated creation of strand-based human hair grooms from text [51], scans [22, 49], images [14, 20, 42, 43, 52, 54, 76] or videos [50, 58, 74] has been widely studied, there is no equivalent work on animal fur.

Unlike human hair, animal fur tends to cover the entire body, varies across different animals, and differs in length, density, and orientation even within the same animal. Another difficulty is that there are no large datasets

---

of animals with fur that can be leveraged to learn a prior for the inner fur structure and the global fur style to constrain optimization-based reconstruction from a set of images. Thus, we have to ask how can we get information about how hair varies across animals and over the body of an animal? Our key idea is to leverage both given multi-view images, where we can extract orientations and silhouettes, and an vision language model (VLM) that is used to resolve ambiguities from the image signal and to provide general knowledge about fur of the specific animal. This idea leads to our method called *NeuralFur*, which enables automatic reconstruction of a variety of 3D animals with detailed strand-based fur from given multi-view images. *NeuralFur* is the first method that automatically reconstructs animal fur that can be imported into a standard computer graphics frameworks for simulation and rendering.

While recent methods [4, 32, 34, 46] show interesting results for reconstructing 3D animals from images and videos using the mesh-based, parametric SMAL [78] model, they are lacking geometrical details like fur. In contrast, we aim for a realistic digital replica of the geometry of an animal with the following properties: (1) The fur is represented with strands, which are compatible with existing rendering and simulation pipelines. (2) The underlying geometry reflects the true geometry of an animal rather than a coarse approximation of it. (3) Animal body parts are taken into account during 3D reconstruction as different parts of the animals differ in strand length. To this end, we propose a method that leverages recent advances in human hair reconstruction and general knowledge from vision language models to reconstruct detailed 3D animal models with fur. Specifically, we build our method based on the state-of-the-art strand-based human-hair reconstruction technique Gaussian Haircut [74], where hair is represented with connected Gaussian primitives. Based on multi-view images, we reconstruct a fine-grained geometry using NeuS [56]. We use a vision language model to reason about fur thickness, growth direction, and length. This information is used to refine the NeuS-based reconstruction, and to guide the 3DGS-based fur strand reconstruction.

With this scheme, we address the following key challenges: (i) the NeuS reconstruction results in a geometry that includes the fur. To obtain the actual geometry of the underlying body requires considering the thickness / volume of the fur which can vary largely depending on the body part region (e.g., fur is typically very short in the face region, whereas there could be very long hair in other regions like at the mane of a lion). To localize the body parts of an animal, we fit the SMAL [78] model to the NeuS reconstruction and transfer general body part labels from SMAL to the NeuS reconstruction. Given an image of an animal and the body part labels, we leverage a VLM to retrieve information about the local fur such that we can

remove the fur from the geometry (defurring). This step allows us to model the actual body of the animal, where we can place the roots for our fur strands. (ii) animal fur grows in different directions within different animal body parts. As orientation maps extracted from the input images based on Gabor-filters are undirected, we leverage a vision language model to extract a general guiding signal for the growing direction to resolve the ambiguity. To regularize fur-strand geometries during optimization, we employ a strand-level prior learned from human hair, as no fur-specific datasets are available. Although not ideal, this approach proves sufficient for our purposes.

As illustrated in Figure 1, our method, for the first time, is effective in reconstructing entire geometries of animals with fur from multi-view images. *NeuralFur* generalizes across different animal types, and takes their region-dependent fur properties into account. The fur is compatible with standard computer graphics pipelines, which enables downstream applications with physics-based rendering and simulation. Note that we do not target appearance reconstruction; colors and shader attributes can be assigned to our reconstructed fur strand geometries in a downstream rendering engine, as shown in Figure 1. Extensive qualitative comparisons with prior animal- and hair-reconstruction methods demonstrate that, unlike previous approaches, our method more accurately reconstructs detailed animal fur geometry.

In summary, we present the first strand-based fur reconstruction method for animals from multi-view images. It is enabled by the following contributions:

- the hybrid usage of images and general knowledge from a VLM model, where the external knowledge is used to resolve ambiguities from the input to estimate the underlying animal body and to guide the growing direction and fur length. A key aspect is the localization of the retrieved VLM information to specific regions of the animal.
- an implicit fur representation network based on a multi-layer perceptron (MLP), which enables the prediction of spatially varying fur strands.

2. Related works

We propose a method for reconstructing detailed 3D animal geometry with strand-based fur from multi-view images, guided by a vision-language model. Our approach unifies classical mesh-based animal reconstruction with strand-based human hair techniques.

Animal reconstruction. Creating a 3D digital replica of animals presents numerous challenges. These include recovering the pose [38, 66, 72], shape [3, 62], and appearance of an animal from images [44, 59] or videos [46, 60, 65, 67–69]. Despite recent progress, relatively few
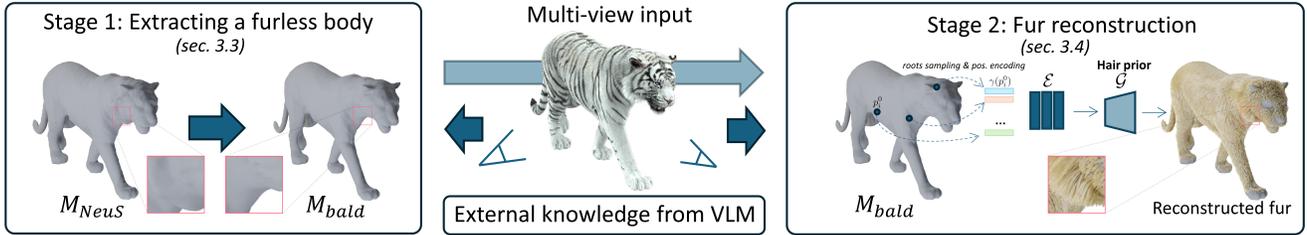
Figure 2. Our method, *NeuralFur*, consists of two stages: (i) extracting a furless mesh geometry by intelligently shrinking the full mesh reconstructed from multi-view images, and (ii) reconstructing strand-based fur by initializing roots from the furless mesh. For both stages, external knowledge from a VLM is leveraged. Based on the depicted animal, the VLM provides information about fur thickness, length, and orientation. This guidance is then used to train a neural fur strand representation (MLP), which can be queried at any mesh surface location to generate fur strands suitable for rendering.

works address 3D animal reconstruction and appearance [2, 29, 63, 64], primarily due to the scarcity of 3D animal datasets [62, 72] and the inherent difficulty of capturing static and dynamic animals [29]. Early work focuses on category-specific reconstruction, such as dogs [4, 18, 45], horses [23, 79], and birds [17]. Cashman and Fitzgibbon [5] learn a parametric dolphin model from images, LASSIE [71] learns articulated animal shape via part discovery. Kanazawa et al. [16] deform template meshes for cats from correspondences, and [18] predict canine pose. Zuffi et al. [78] introduced SMAL, a parametric model for quadruped animals. GenZoo [34] uses SMAL [78] and trains a regressor on synthetic data to estimate pose and shape from a single image. AniMer [32] estimates animal pose and shape using a family-aware transformer, enhancing the reconstruction accuracy of diverse quadrupedal families. AWOL [77] generates animals and trees using language guidance. RAW [21] reconstructs animals and their surroundings. Both model-based approaches [32, 34, 62] (i.e., methods that rely on parametric models for animal representation) and model-free approaches [57, 70] typically rely on coarse geometry to represent animals, providing a low-dimensional representation. However, animals often exhibit complex geometries, including fur and fine surface details. In this work we propose a *layered representation* consisting of a mesh for body with a strand-based fur layer that integrates seamlessly with classical graphics pipelines for simulation, rendering, and editing

Strand-based reconstruction. Strands are the standard representation for high-fidelity 3D hair modeling in research [39, 49, 73] and production [8, 12, 14]. This parametrization offers advantages over volumes or meshes in terms of physics simulation [10, 13, 15] and geometric control [48, 51, 61, 75]. However, manual modeling remains labor-intensive due to the high strand count and geometric complexity. To automate this process, image-based methods estimate strand orientations from photographs [36]

and optimize 3D strands to align with 2D projections [6, 7, 31, 33, 37, 52]. Yet, these approaches struggle with occlusions and typically model strands as polylines without thickness, separating geometry and appearance [42, 50, 54, 58]. Recent works incorporate learned priors [42, 50, 58] or anisotropic 3D Gaussians for richer geometry and appearance representation [30, 74, 76]. While sharing detail-preservation challenges with hair reconstruction methods, strand-based fur capture must additionally address regional fur length modeling, part segmentation, and furless geometry estimation.

Large vision-language models. Vision–language models (VLMs) combine visual perception with language understanding, enabling open-ended reasoning about images. Large-scale pretraining approaches, such as CLIP [40], show strong generalization across diverse visual tasks, as do more recent models like BLIP [24], LLaVA [25, 26], and Flamingo [1]. More recently, commercial systems such as GPT-5, Gemini, and Claude have demonstrated fast and efficient multi-modal reasoning, including the ability to process multiple images simultaneously. In this work, we use ChatGPT [35] to support accurate modeling of realistic animal fur geometry.

3. Method

Our method consists of two stages: (i) extracting a furless mesh geometry of the animal from multi-view images and (ii) strand-based fur reconstruction, see Figure 2. We initialize our method with NeuS [56] which results in a mesh that models the entire animal, including fur. To decompose this mesh into a body mesh and a fur layer, we take advantage of the external knowledge of a VLM that we localize to the different regions of the animal (Section 3.2). Specifically, we extract fur thickness for "de-furring" (Section 3.3) as well as fur length information for fur growing from this model (Section 3.4).
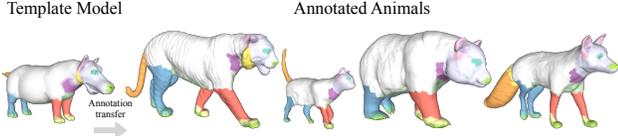
Figure 3. **Annotation.** Automatic part annotation for each animal obtained from the fitted SMAL [78] model.

3.1. Shape initialization

We reconstruct the coarse geometry of an animal using NeuS [56] which leverages a signed distance function (SDF). Specifically, it uses Multi-layer Perceptrons (MLPs) for color and signed distance prediction by applying a volumetric raymarching algorithm and supervising it with silhouette constraints and an Eikonal penalty function. From the reconstructed SDF the visible geometry of the animal is extracted with Marching Cubes [27, 28] resulting in our initial mesh M_{NeuS}.

3.2. Localizing external knowledge

Reconstructing fur solely from multi-view images without additional constraints is insufficient as we do not know underlying body parts or fur length. Therefore, we incorporate external knowledge from a VLM (ChatGPT [35]) to annotate body regions with fur lengths in centimeters. To obtain these annotations, we prompt it with several frontal and side-view images of the animal and request fur length, thickness, and growth direction for each region. However, the answers from ChatGPT need to be structured and attached to specific regions (resulting in localized knowledge). To this end, we leverage the parametric animal model SMAL [78]. SMAL is a morphable model which can represent a variety of four-legged animals. It offers a fixed topology which is used for region annotations that is shared across all animals. We define the following regions on SMAL: *leg (front,rear), paw pads, paws, belly, neck, face, ears, inner ear canal, under tail, eyes, tail, nose tip, body and mane if available*.

To leverage these annotations, we fit SMAL to the M_{NeuS} and transfer the labels based on nearest neighbors, see Figure 3. Note that M_{NeuS} offers more geometrical detail than SMAL. The fitting operates in 3D space using three stages: (1) in the first stage, we optimize for global translation $\boldsymbol{\gamma}$ and global rotation $\boldsymbol{\theta_0}$ of the SMAL model. (2) in the second stage, we additionally optimize for shape parameters $\boldsymbol{\beta}$ and joint pose parameters $\boldsymbol{\theta}$. (3) in the last stage, we optimize for per-vertex deformations, which helps to obtain finer details. In each optimization step, we sample 20,000 points on our initial mesh reconstruction and, similar to WLDO [4], optimize the respective parameters of the different stages with respect to Chamfer distance, Laplacian smoothness regularization as well as normal consistency.

Figure 4. From left to right: input image, initial shape M_{NeuS}, furless geometry M_{bald}, and their overlay.

3.3. De-furring – Extracting a furless geometry

As the initial geometry reconstruction M_{NeuS} represents the entire body including fur as a single continuous surface, it cannot directly be used to place roots of fur strands. Thus, we first have to derive the geometry of the "bald" animal body M_{bald}. Unfortunately, no datasets exist that contain paired examples of shaved and unshaved animals. Therefore, we leverage the localized external knowledge from the VLM. We query information about the effective thickness of the fur and "erode" the geometry correspondingly to retrieve the furless animal by converting M_{NeuS} into an SDF and applying a spatially varying shrinkage that offsets the surface inward:

$$\text{SDF}_{\text{defur}}(x) = \text{SDF}(x) + s(x), \tag{1}$$

where $s(x)$ is the locally defined shrinkage term based on the thickness value defined for the nearest vertex. From this SDF, we extract the furless mesh M_{bald} using Marching cubes [27, 28], see Figure 4.

3.4. Fur modeling and reconstruction

Neural fur representation. Similar to hair reconstruction methods, we parametrize the fur geometry of animals using a strand-based representation with N individual strands. The i-th fur strand is represented as a 3D polyline with L points: $\mathbf{S}_i = \{\mathbf{p}_i^l\}_{l=1}^L$. Based on this polyline representation, we define directions between nearest points as $\mathbf{d}_i^l = \mathbf{p}_i^{l+1} - \mathbf{p}_i^l$ and normalized directions as $\mathbf{b}_i^l = \mathbf{d}_i^l/\|\mathbf{d}_i^l\|_2$. As polylines have a high degree of freedom in 3D space, we model them in the latent space $\mathbf{z} \in \mathbb{R}^{64}$ of the synthetic hairstyles prior model $\mathcal{G}$ from NeuralHaircut [50]. We

find that the strand-level hair prior also serves as a good approximation for fur when scaled appropriately (see below). The strands are defined in a local coordinate system which is typically derived from a tangent-bitangent-normal (TBN) basis constructed via texture space. In our scenario, we are operating on the defurred mesh M_{bald} which does not come with a unified texture space as for human heads [42, 50, 74]. However, it is important that the local coordinate systems are smoothly aligned over the entire defurred mesh. We, therefore, compute a directional face field [55] with further sign consistency resolved using parallel transport [53] across shared edges. The bitangent vector is defined as a cross-product between the normal and the tangent.

Instead of modeling a finite set of fur strands rooting at the defurred mesh, we define a continuous function that can be sampled at the surface and outputs the latent code of the strand. Specifically, we use a multilayer perceptron (MLP) $\mathcal{E}$ that can be queried at a surface point $\mathbf{p}_i^0$ which defines the strand root on the defurred mesh M_{bald}:

$$\mathbf{z}_i = \mathcal{E}(\gamma(\mathbf{p}_i^0)), \tag{2}$$

where γ is positional encoding [41] and $\mathbf{z}_i$ is the latent strand code in the local TBN at the surface point $\mathbf{p}_i^0$. To obtain the polyline from the latent fur $\mathbf{z}_i$ we use the pretrained hair decoder $\mathcal{G}$ from NeuralHaircut [50]. Unlike prior works on hair modeling, our goal is to achieve precise, centimeter-level control over fur length for each specific body region. To this end, we normalize the output strand from the prior model and multiply by the length ℓ_j that we retrieve from the VLM during optimization:

$$\mathbf{p}_i' = \mathcal{G}(\mathbf{z}_i), \ \mathbf{p}_i = \ell_j \frac{\mathbf{p}_i'}{\|\mathbf{p}_i'\|_2}, \ \mathbf{p}_i \in \mathbb{R}^{L \times 3}. \tag{3}$$

Fur reconstruction. We optimize the animal fur based on this neural fur representation using geometry and photometric losses obtained through differentiable rasterization of strands using Gaussian Splatting [19]. Specifically, in each optimization step, we sample a set of strand roots $\{\mathbf{p}_i^0\}_{i=1}^N$ and compute the corresponding polylines by decoding the latent strand code retrieved by the MLP at the root location. Similar to Gaussian Haircut [74], we attach Gaussian primitives to each strand segment for rendering. Besides the sampling, every step is differentiable, and we can compute gradients to update the MLP weights with respect to:

$$\begin{aligned}
\mathcal{L}_{\text{rec}} = &\ \lambda_{\text{sil}}\,\mathcal{L}_{\text{sil}} + \lambda_{\text{dir}}\,\mathcal{L}_{\text{dir}} + \lambda_{\text{dir}}^{gpt}\,\mathcal{L}_{\text{dir}}^{gpt} \\
&+ \lambda_{\text{chm}}\,\mathcal{L}_{\text{chm}} + \lambda_{\text{penetr}}\,\mathcal{L}_{\text{penetr}} + \lambda_{\text{shape}}\,\mathcal{L}_{\text{shape}}.
\end{aligned} \tag{4}$$

Silhouette loss: We supervise using a silhouette loss $\mathcal{L}_{\text{sil}}$, which is L_1 distance between the rendered and estimated animal silhouette from the input images.

Orientation loss: We further supervise orientations using:

$$\mathcal{L}_{\text{dir}} = \sum_p \tau_p \min\{\mathrm{d}(\beta_p, \hat{\beta}_p), \mathrm{d}(\beta_p, \hat{\beta}_p) \pm \pi\} - \log \tau_p, \tag{5}$$

where d denotes the absolute angular difference between the directions, $\hat{\beta}_p$ denotes the direction in the pixel p obtained using a set of Gabor filters [36] and τ_p is a rendered confidence factor.

We mask the silhouette and direction losses using a non-baldness mask derived from M_{NeuS}. To construct this mask, we render the textured mesh, assigning a value of 1 to the texture of vertices with a nonzero length and 0 otherwise. In this way, bald regions (such as around the eyes, nose tip, and paw pads) are discarded from the loss.

Orientation consistency loss: As orientation maps derived from Gabor filters are undirected, we incorporate an auxiliary loss function that leverages prior knowledge of Chat-GPT on normalized hair growing direction for each part $\hat{\mathbf{g}}_i \in \mathbb{R}^3$ to enforce directional consistency:

$$\mathcal{L}_{\text{dir}}^{gpt} = \frac{1}{N(L-1)} \sum_{i=1}^N \sum_{l=1}^{L-1} w^l \max(0, -\cos(\mathbf{b}_i^l \cdot \hat{\mathbf{g}}_i)), \tag{6}$$

where $w^l = \frac{l-1}{L-1}$ is a linear weight that increases from the root to the tip.

Chamfer loss: For geometry supervision [50], we use the one-way Chamfer distance between the full geometry M_{NeuS} and the learned strands:

$$\mathcal{L}_{\text{chm}} = \sum_{k=1}^K \|\mathbf{x}_k - \mathbf{p}_k\|_2^2, \tag{7}$$

where $\mathbf{x}_k$ are K random points sampled from S and their nearest points on the strands $\mathbf{p}_k$.

Curvature consistency loss: We employ an additional regularization on the consistency of curvature profiles across the fur. For strand i, we compute bending angles θ_i^l between the nearest strand segments:

$$\theta_i^l = \arccos\left(\text{clamp}\left(\mathbf{b}_i^{l-1} \cdot \mathbf{b}_i^l, -1, 1\right)\right), \ l = 2, \ldots, L-1. \tag{8}$$

The curvature signature for each strand $\theta(S_i) = \{\theta_i^l\}_{l=2}^{L-1}$ and mean curvature among N strands is defined as $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \theta(S_i)$. With this information, we compute the curvature consistency loss:

$$\mathcal{L}_{\text{shape}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{L-2} \|\theta(S_i) - \bar{\theta}\|^2. \tag{9}$$

Penetration loss: To prevent interpenetrations into M_{bald}, we add an additional penetration loss:

$$\mathcal{L}_{\text{penetr}} = \frac{1}{N} \sum_{i=1}^N \max(0, -\text{SDF}_{\text{defur}}(x_i)). \tag{10}$$

3.5. Implementation details

For complete geometry reconstruction, we run NeuS [56] for 300,000 iterations. During training fur reconstruction, we sample 15,000 strands per iteration. At inference, we reconstruct 500,000 strands, with the ability to scale to larger counts, as we are using an MLP to represent the fur groom. To model fur in metric space, we first identify the eye keypoints in 3D geometric space and use VLM to estimate the distance between the eye centers in the image. This measured scale is then used to compute positional offsets, allowing the fur to be grown accurately in centimeters. Our method takes around 10 hours on a single A100.

4. Experiments

We train our method on the Artemis [29] dataset, which consists of 36 high-quality multi-view images captured for each animal. Since no prior work addresses this specific task of animal reconstruction with fur modeling, we compare our approach against reconstruction approaches as well as a strand-based hair reconstruction method. For quantitative evaluation, we use unsupervised metrics such as geometry consistency of the reconstructed fur and distance to the outer surface, and provide comparison results on an artist-created synthetic asset with ground-truth fur geometry.

4.1. Qualitative evaluation

We compare our method against popular 3D reconstruction approaches for animals, such as SMAL [78] and Gen-Zoo [34], general scene reconstructions, NeuS [56], as well as a method designed for strand-based human hair reconstruction, Gaussian Haircut [74]. As there are no prior works in the direction of fur modeling it serves as the closest baseline to ours. We present qualitative results of our method with baselines in Figure 5. Note that SMAL, NeuS, and GenZoo are only able to produce coarse animal geometry, which includes the full outer surface. While Gaussian Haircut (GH) generates realistic and high-quality reconstructions of strand-based hairstyles, it fails to reconstruct accurate fur geometry. In Figure 5, we show results after the second stage, as the third stage significantly degrades performance. A key limitation is the lack of explicit control over strand length, which is only optimized implicitly. In contrast, our method reconstructs realistic fur directly from images with high accuracy.

4.2. Quantitative evaluation

For quantitative comparison of our method with Gaussian Haircut [74] and ablation of design choices of our model, we use an artist-created synthetic asset of a tiger with available ground-truth strand-based geometry for fur. To do that, we realistically render the animal using Blender [9] from 90 views from a circular trajectory, and then use it as an input

	Thresholds: cm / degrees								
Method	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision			Recall			F-score		
Ours	26.22	39.32	48.05	20.58	34.08	45.69	23.06	36.51	46.84
GH (2nd stage)	16.24	25.51	32.34	23.51	36.04	45.87	19.21	29.87	37.93
GH (3rd stage)	5.78	10.44	15.43	21.39	32.82	42.36	9.10	15.84	22.62
w/ opt.length	3.05	6.26	9.82	16.63	28.27	38.45	5.16	10.25	15.65
w/ fix. length	17.18	26.65	33.73	22.24	35.08	45.44	19.38	30.29	38.72
w/ Unet	5.77	12.16	20.06	7.05	15.39	25.80	6.35	13.58	22.57

Table 1. Quantitative evaluation on a synthetic tiger model where ground truth information of the fur is given as reference.

Method	Fur length		Fur curvature			Fur orientation		Fur silhouette
	$\mu_L \pm \sigma_L$ (cm)	$\sigma_{loc}(L)\downarrow$	$\text{Var}_{glob}(\kappa)\downarrow$	$\text{Var}_{loc}(\kappa)\downarrow$	$\kappa_{max}\downarrow$	$\text{Var}_{loc}(\text{dir})\downarrow$	$\text{Var}_{loc}^{first}(\text{dir})\downarrow$	CD $\downarrow$
	"Panda"							
Ours	5.197 ± 1.38	0.283	0.0005	0.000069	1.80	0.031	0.041	0.000261
GH (2nd stage)	7.322 ± 8.06	3.615	0.0018	0.000789	3.10	0.317	0.302	0.001105
GH (3rd stage)	11.238 ± 7.53	4.362	0.2318	0.352237	3.13	0.730	0.765	0.001600
w/ opt. length	17.169 ± 10.59	1.469	0.0048	0.000681	3.11	0.035	0.054	0.000637
w/ fix. length	6.000 ± 0.00	0.000	0.0014	0.000192	3.05	0.035	0.053	0.000257
w/ Unet	5.196 ± 1.38	0.285	0.0007	0.001282	1.35	0.157	0.200	0.000254
w/o $\mathcal{L}_{defur}$	5.197 ± 1.35	0.291	0.0019	0.000329	2.90	0.044	0.060	0.000802
w/o $\mathcal{L}_{chm}$	5.197 ± 1.38	0.283	0.0010	0.000149	2.65	0.033	0.033	0.000330
w/o $\mathcal{L}_{dir}^{gpt}$	5.197 ± 1.38	0.283	0.0012	0.000224	2.71	0.041	0.050	0.000261
w/o $\mathcal{L}_{shape}$	5.196 ± 1.38	0.283	0.0016	0.000284	2.91	0.033	0.053	0.000265
	"White Tiger"							
Ours	3.555 ± 1.46	0.355	0.0016	0.000210	3.05	0.030	0.035	0.000207
GH (2nd stage)	6.511 ± 17.86	8.190	0.0144	0.006519	3.12	0.526	0.469	0.000312
GH (3rd stage)	9.929 ± 17.24	8.087	0.4049	0.636722	3.14	0.885	0.970	0.000553
w/ opt. length	19.992 ± 15.20	1.756	0.0047	0.000637	3.09	0.036	0.050	0.000453
w/ fix. length	5.000 ± 0.00	0.000	0.0023	0.000352	3.06	0.037	0.050	0.000202
w/ Unet	3.556 ± 1.46	0.355	0.0008	0.001337	1.39	0.124	0.081	0.000379
w/o $\mathcal{L}_{defur}$	3.570 ± 1.48	0.357	0.0036	0.000728	2.96	0.051	0.082	0.000420
w/o $\mathcal{L}_{chm}$	3.555 ± 1.46	0.355	0.0021	0.000306	2.66	0.031	0.040	0.000258
w/o $\mathcal{L}_{dir}^{gpt}$	3.555 ± 1.46	0.355	0.0041	0.000862	3.04	0.049	0.037	0.000190
w/o $\mathcal{L}_{shape}$	3.555 ± 1.46	0.355	0.0034	0.000537	3.08	0.030	0.056	0.000202

Table 2. Unsupervised geometry consistency metrics for length, direction, and curvature, evaluated for both local and global cases, as well as for coverage of the outer surface.

to the method. Following [33, 42, 50], we measure precision, recall, and F-score between points with directions of reconstructed strands and the ground truth, see Table 1.

We further evaluate our method on two scenes in terms of consistency of length, directions, and curvature in local (using ten nearest neighbours) and global spaces, and distance to the surface, see Table 2. Specifically, we use unsupervised metrics that calculate: (1) fur length statistics based on global length mean μ_L and standard deviation of strands σ_L along with deviation in local region ($\sigma_{loc}(L)$), (2) fur curvature information based on local ($\text{Var}_{loc}(\kappa)$) and global variance ($\text{Var}_{glob}(\kappa)$) of curvature, and maximum bending angle κ_{max}, (3) fur orientation based on local variance of directions for the whole strand ($\text{Var}_{loc}(\text{dir})$) and considering only local variance of directions defined on first segment ($\text{Var}_{loc}^{first}(\text{dir})$), and (4) fur silhouette based on the two-way chamfer distance between the ground-truth geometry and the tips of the fur strands (CD).

Comparison with Gaussian Haircut. As shown in Tables 1 and 2, we compare results with Gaussian Haircut after the second and third stages (denoted as GH (2nd stage) and GH (3rd stage)), respectively. We observe a notable degradation in quality compared to our method.

Part-specific fur length. We ablate the importance of modeling different fur lengths for different parts, see Tables 1

Figure 5. Qualitative results of our reconstruction method compared with existing baselines. Surface reconstruction baselines produce very coarse geometry. Applying the existing state-of-the-art hair reconstruction method, Gaussian Haircut [74] leads to inconsistent strand lengths and noticeable artifacts. Our method produces accurate strand-based geometry. A digital zoom-in is recommended.

and 2. In Figure 6, we provide qualitative results of our method with fixed length (denoted as w/ fix. length), which exhibits artifacts especially when modeling long fur in the face region (see row 3). We also show results with implicitly optimized length, similar to Gaussian Haircut [74] (w/ opt. length). However, unsupervised length optimization leads to unrealistic fur length and noticeable inconsistencies.

Accurate body geometry. To demonstrate the importance of body geometry, we launch our reconstruction method with the body geometry replaced from M_{bald} to M_{NeuS}, see w/o defur in Figure 6. While qualitatively it shows similar results, there is an increased level of unrealistic strands, as we could see in the decreased spatial consistency of directions and curvatures in Table 2. Finally, we measured the bidirectional Chamfer distance (in mm) between the ground-truth bald geometry M_{bald}^{gt} and our reconstruction M_{bald},

as well as M_{NeuS}. The distances (mean / min / max) for M_{bald}: **6.83 / 0.45 / 69.72 mm** are significantly lower than for M_{NeuS}: **7.56 / 0.49 / 92.22 mm**.

Parametrization. We explore different latent space parametrization by replacing our NeuralFur MLP with the Unet architecture from Neural Haircut [50]. From Figure 6 and Tables 1 and 2, we can see significant degradation in consistency and realism of produced fur.

Prompting. To model accurate responses, we requested fur lengths specifically in centimeters, asked for specific body parts, and provided images from several views. We evaluate the performance of length prediction on a synthetic asset, including ablations for predictions without images, with a single image, and with several images. The results are reported in Table 3. For the structure of the prompt, please refer to the supplementary.
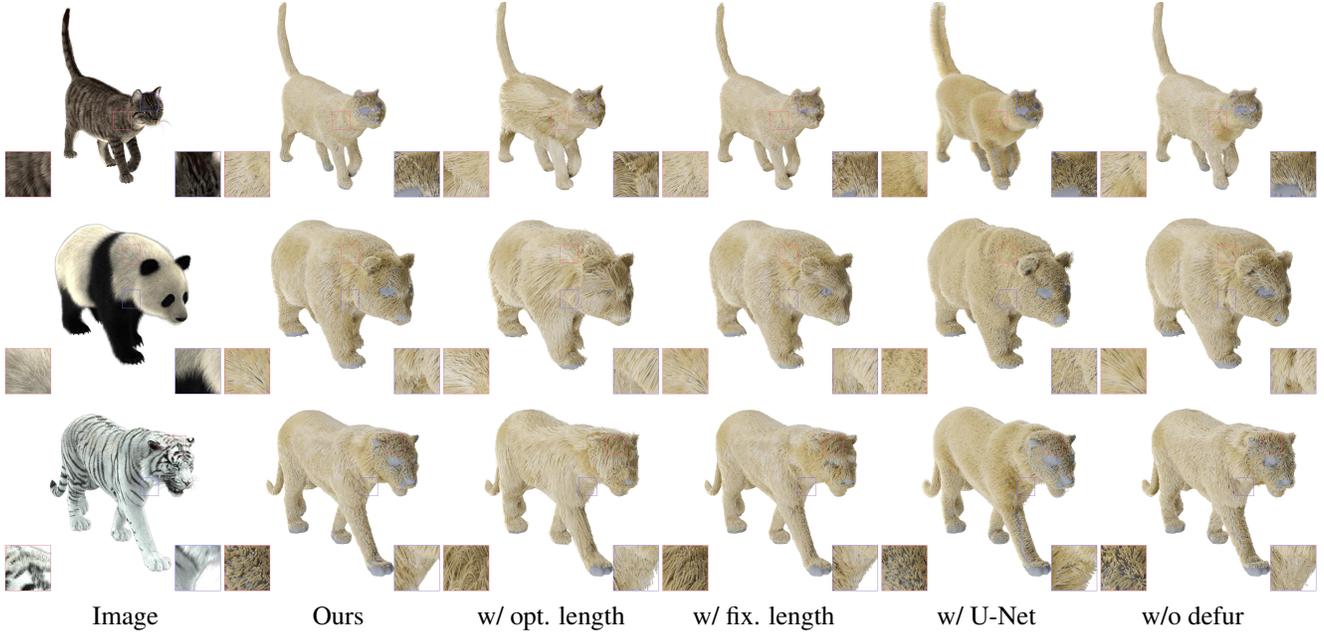
Figure 6. **Ablation study.** Qualitative evaluation of our design choices regarding length, fur parametrization, and the importance of the defurring approach for accurate geometry modeling.

	Ground Truth (cm)	GPT responses (cm)		
	(mean / min / max)	No image	1 image	2 images
Face	0.61 / 0.00008 / 2.57	1 - 2	1 - 1.5	0.5 - 1
Body	2.30 / 0.0011 / 8.60	2 - 5	2 - 3	1 - 2
Legs	0.65 / 0.15 / 1.26	2 - 4	1.5 - 2.5	0.5-1.5
Mane	1.89 / 0.00 / 9.27	5 - 8	6 - 7	3 - 5

Table 3. Comparison of results on length estimation for different parts obtained from ChatGPT on a synthetic asset.

Figure 7. Simulation of a reconstructed panda with fur in Unreal Engine [11]. Note that the fur color is set by an artist, and the animation is applied via the SMAL model.

Losses. We ablate the importance of chamfer, curvature, and orientation consistency losses in Table 2, see w/o $\mathcal{L}_{chm}$, w/o $\mathcal{L}_{shape}$, and w/o $\mathcal{L}_{dir}^{gpt}$ correspondingly. Across several animal models, omitting $\mathcal{L}_{chm}$ results in degraded CD metric and reduced curvature consistency. Excluding $\mathcal{L}_{shape}$ diminishes both local and global directional consistency. Finally, discarding $\mathcal{L}_{dir}^{gpt}$ yields greater inconsistencies in directional alignment as well as curvature.

4.3. Discussion

Our method relies solely on multi-view images and effectively reconstructs detailed fur grooms by leveraging complementary information retrieved from a VLM. However, such multi-view data may not always be available for in-the-wild animals, and the results are dependent on the quality of the VLM outputs. Moreover, our framework is constrained by SMAL, limiting it to quadruped animals with a shared topology; extending the model to support a wider variety of animal types remains an exciting direction for future work. Another promising direction is the explicit modeling

of whiskers, which are essential features for many species, particularly, when close-up reconstructions of the face are required. Finally, it is important to note that our approach is restricted to recovering static geometry; appearance attributes and animations must be incorporated afterwards within a Computer Graphics framework, see Figure 7.

5. Conclusion

In this work, we introduced the first method capable of reconstructing strand-based fur geometry from multi-view images. Our method integrates furless geometry extraction with a subsequent fur optimization stage, leveraging prior knowledge from VLMs to estimate fur length across body regions and hair growth direction, resulting in notable improvements in reconstruction quality. We argue that leveraging external knowledge sources such as VLMs is particularly valuable for 3D reconstruction tasks where data is scarce and explicit priors cannot be easily learned.

Acknowledgements

Disclosure. While MJB is a co-founder and Chief Scientist at Meshcapade, his research in this project was performed solely at, and funded solely by, the Max Planck Society.

References

[1] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katie Millicah, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. Flamingo: a visual language model for few-shot learning. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2022. Curran Associates Inc. 3

[2] Tobias Grønbeck Andersen, Viggo Falster, Jeppe Revall Frisvad, and Niels Jørgen Christensen. Hybrid fur rendering: combining volumetric fur with explicit hair strands. *The Visual Computer*, 32(6):739–749, 2016. 3

[3] Benjamin Biggs, Thomas Roddick, Andrew Fitzgibbon, and Roberto Cipolla. Creatures great and smal: Recovering the shape and motion of animals from video. In *Asian Conference on Computer Vision*, pages 3–19. Springer, 2018. 2

[4] Benjamin Biggs, Oliver Boyne, James Charles, Andrew Fitzgibbon, and Roberto Cipolla. Who left the dogs out?: 3D animal reconstruction with expectation maximization in the loop. In *ECCV*, 2020. 2, 3, 4

[5] Thomas J Cashman and Andrew W Fitzgibbon. What shape are dolphins? building 3d morphable models from 2d images. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):232–244, 2012. 3

[6] Menglei Chai, Linjie Luo, Kalyan Sunkavalli, Nathan Carr, Sunil Hadap, and Kun Zhou. High-quality hair modeling from a single portrait photo. *ACM Transactions on Graphics*, 34(6):1–10, 2015. 3

[7] Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. Autohair: fully automatic hair modeling from a single image. *ACM Trans. Graph.*, 35:116:1–116:12, 2016. 3

[8] Matt Jen-Yuan Chiang, Benedikt Bitterli, Chuck Tappan, and Brent Burley. A practical and controllable hair and fur model for production path tracing. In *Computer Graphics Forum*, pages 275–283. Wiley Online Library, 2016. 3

[9] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2023. 6

[10] Gilles Daviet. Interactive hair simulation on the gpu using admm. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 3

[11] Epic Games. Unreal engine. 8

[12] Luca Fascione, Johannes Hanika, Rob Pieké, Ryusuke Villemin, Christophe Hery, Manuel Gamito, Luke Emrose, and André Mazzone. Path tracing in production. In *ACM SIGGRAPH 2018 Courses*, pages 1–79. 2018. 3

[13] Yun Fei, Henrique Teles Maia, Christopher Batty, Changxi Zheng, and Eitan Grinspun. A multi-scale model for simulating liquid-hair interactions. *ACM Transactions on Graphics (TOG)*, 36(4):1–17, 2017. 3

[14] Philip-William Grassal, Luca Hormann, Nadia Hamlaoui, Titus Leistner, and Lynton Ardizzone. A mobile scanning solution to reconstruct strand-based hairstyles. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks*, New York, NY, USA, 2025. Association for Computing Machinery. 1, 3

[15] Jerry Hsu, Tongtong Wang, Zherong Pan, Xifeng Gao, Cem Yuksel, and Kui Wu. Sag-free initialization for strand-based hybrid hair simulation. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2023)*, 42(4), 2023. 3

[16] Angjoo Kanazawa, Shahar Kovalsky, Ronen Basri, and David Jacobs. Learning 3d deformation of animals from 2d images. In *Computer Graphics Forum*, pages 365–374. Wiley Online Library, 2016. 3

[17] Angjoo Kanazawa, Shubham Tulsiani, Alexei A Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *Proceedings of the European conference on computer vision (ECCV)*, pages 371–386, 2018. 3

[18] Sinead Kearney, Wenbin Li, Martin Parsons, Kwang In Kim, and Darren Cosker. Rgbd-dog: Predicting canine pose from rgbd sensors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8336–8345, 2020. 3

[19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 5

[20] Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. Deepmvshair: Deep hair modeling from sparse views. *SIGGRAPH Asia 2022 Conference Papers*, 2022. 1

[21] Peter Kulits, Michael J Black, and Silvia Zuffi. Reconstructing animals and the wild. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16565–16577, 2025. 3

[22] Rachmadio Noval Lazuardi, Artem Sevastopolsky, Egor Zakharov, Matthias Niessner, and Vanessa Sklyarova. Geomhair: Reconstruction of hair strands from colorless 3d scans. *arXiv preprint arXiv:2505.05376*, 2025. 1

[23] Ci Li, Nima Ghorbani, Sofia Broomé, Maheen Rashid, Michael J Black, Elin Hernlund, Hedvig Kjellström, and Silvia Zuffi. hsmal: Detailed horse shape and pose reconstruction for motion pattern recognition. *arXiv preprint arXiv:2106.10102*, 2021. 3

[24] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *ICML*, 2022. 3

[25] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2023. Curran Associates Inc. 3

[26] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26296–26306, 2024. 3

[27] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, 1987. 4

[28] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, page 163–169, New York, NY, USA, 1987. Association for Computing Machinery. 4

[29] H. Luo, A. Chen, Q. Zhang, B. Pang, M. Wu, L. Xu, and J. Yu. Convolutional neural opacity radiance fields. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, Los Alamitos, CA, USA, 2021. IEEE Computer Society. 3, 6

[30] Haimin Luo, Ouyang Min, Zijun Zhao, Suyi Jiang, Longwen Zhang, Qixuan Zhang, Wei Yang, Lan Xu, and Jingyi Yu. Gaussianhair: Hair modeling and rendering with light-aware gaussians. *ArXiv*, abs/2402.10483, 2024. 3

[31] Linjie Luo, Hao Li, Sylvain Paris, Thibaut Weise, Mark Pauly, and Szymon Rusinkiewicz. Multi-view hair capture using orientation fields. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1490–1497. IEEE, 2012. 3

[32] Jin Lyu, Tianyi Zhu, Yi Gu, Li Lin, Pujin Cheng, Yebin Liu, Xiaoying Tang, and Liang An. Animer: Animal pose and shape estimation using family aware transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17486–17496, 2025. 2, 3

[33] Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheikh. Strand-accurate multi-view hair capture. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 155–164, 2019. 3, 6

[34] Tomasz Niewiadomski, Anastasios Yiannakidis, Hanz Cuevas-Velasquez, Soubhik Sanyal, Michael J. Black, Silvia Zuffi, and Peter Kulits. Generative zoo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. 2, 3, 6, 7

[35] OpenAI. Gpt-5, 2025. ChatGPT, accessed 2025-08-18. 3, 4

[36] Sylvain Paris, Héctor M. Briceño, and François X. Sillion. Capture of hair geometry from multiple images. *ACM SIGGRAPH 2004 Papers*, 2004. 3, 5

[37] Sylvain Paris, Will Chang, Oleg I Kozhushnyan, Wojciech Jarosz, Wojciech Matusik, Matthias Zwicker, and Frédo Durand. Hair photobooth: geometric and photometric acquisition of real hairstyles. *ACM Transactions on Graphics*, 27 (3):30, 2008. 3

[38] Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S Deutsch, Z Yan Wang, et al. Sleap: A deep learning system for multi-animal pose tracking. *Nature methods*, 19(4):486–495, 2022. 2

[39] Emmanuel Piuze, Paul G Kry, and Kaleem Siddiqi. Generalized helicoids for modeling hair geometry. In *Computer Graphics Forum*, pages 247–256. Wiley Online Library, 2011. 3

[40] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021. 3

[41] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Dräxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron C. Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, 2018. 5

[42] Radu Alexandru Rosu, Shunsuke Saito, Ziyan Wang, Chenglei Wu, Sven Behnke, and Giljoo Nam. Neural strands: Learning hair geometry and appearance from multi-view images. In *European Conference on Computer Vision*, 2022. 1, 3, 5, 6

[43] Radu Alexandru Rosu, Keyu Wu, Yao Feng, Youyi Zheng, and Michael J. Black. DiffLocks: Generating 3d hair from a single image using diffusion models. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2025. 1

[44] Nadine Rüegg, Shashank Tripathi, Konrad Schindler, Michael J. Black, and Silvia Zuffi. 3d animal reconstruction with anatomical awareness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8867–8876, 2023. 2

[45] Nadine Rüegg, Shashank Tripathi, Konrad Schindler, Michael J. Black, and Silvia Zuffi. BITE: Beyond priors for improved three-D dog pose estimation. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 8867–8876, 2023. 3

[46] Remy Sabathier, Niloy J Mitra, and David Novotny. Animal avatars: Reconstructing animatable 3d animals from casual videos. In *European Conference on Computer Vision*, pages 270–287. Springer, 2024. 2

[47] Mike Seymour. Behind the scenes on weta digital's real-time hair and fur short meerkat, 2020. Accessed: 2025-08-10. 1

[48] Yuefan Shen, Changgeng Zhang, Hongbo Fu, Kun Zhou, and Youyi Zheng. Deepsketchhair: Deep sketch-based 3d hair modeling. *IEEE transactions on visualization and computer graphics*, 27(7):3250–3263, 2020. 3

[49] Yuefan Shen, Shunsuke Saito, Ziyan Wang, Olivier Maury, Chenglei Wu, Jessica Hodgins, Youyi Zheng, and Giljoo

Nam. Ct2hair: High-fidelity 3d hair modeling using computed tomography. *ACM Transactions on Graphics*, 42(4): 1–13, 2023. 1, 3

[50] Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. Neural haircut: Prior-guided strand-based hair reconstruction. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2023. 1, 3, 4, 5, 6, 7

[51] Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J Black, and Justus Thies. Text-conditioned generative model of 3d strand-based human hairstyles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4703–4712, 2024. 1, 3

[52] Vanessa Sklyarova, Egor Zakharov, Malte Prinzler, Giorgio Becherini, Michael J. Black, and Justus Thies. Im2haircut: Single-view strand-based hair reconstruction for human avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10656–10665, 2025. 1, 3

[53] M. Spivak. *A Comprehensive Introduction to Differential Geometry Vol. II*. Publish or Perish, 1979. 5

[54] Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, and Bo Zheng. Dr.hair: Reconstructing scalp-connected hair strands without pre-training via differentiable rendering of line segments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20601–20611, 2024. 1, 3

[55] Amir Vaxman et al. Directional: A library for directional field synthesis, design, and processing. 5

[56] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2, 3, 4, 6, 7

[57] Yufu Wang, Nikos Kolotouros, Kostas Daniilidis, and Marc Badger. Birds of a feather: Capturing avian shape models from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14739–14749, 2021. 3

[58] Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. MonoHair: High-fidelity hair modeling from a monocular video. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 24164–24173, 2024. 1, 3

[59] Shangzhe Wu, Ruining Li, Tomas Jakab, Christian Rupprecht, and Andrea Vedaldi. Magicpony: Learning articulated 3d animals in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8792–8802, 2023. 2

[60] Yuefan Wu, Zeyuan Chen, Shaowei Liu, Zhongzheng Ren, and Shenlong Wang. Casa: Category-agnostic skeletal animal reconstruction. In *Advances in Neural Information Processing Systems*, pages 28559–28574. Curran Associates, Inc., 2022. 2

[61] Jun Xing, Koki Nagano, Weikai Chen, Haotian Xu, Li-yi Wei, Yajie Zhao, Jingwan Lu, Byungmoon Kim, and Hao Li. Hairbrush for immersive data-driven hair modeling. In *Proceedings of the 32Nd Annual ACM Symposium on User Interface Software and Technology*, pages 263–279, 2019. 3

[62] Jiacong Xu, Yi Zhang, Jiawei Peng, Wufei Ma, Artur Jesslen, Pengliang Ji, Qixin Hu, Jiehua Zhang, Qihao Liu, Jiahao Wang, et al. Animal3d: A comprehensive dataset of 3d animal pose and shape. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9099–9109, 2023. 2, 3

[63] Ling-Qi Yan, Chi-Wei Tseng, Henrik Wann Jensen, and Ravi Ramamoorthi. Physically-accurate fur reflectance: modeling, measurement and rendering. *ACM Trans. Graph.*, 34(6), 2015. 3

[64] Ling-Qi Yan, Henrik Wann Jensen, and Ravi Ramamoorthi. An efficient and practical near and far field fur reflectance model. *ACM Trans. Graph.*, 36(4), 2017. 3

[65] Gengshan Yang, Deqing Sun, Varun Jampani, Daniel Vlasic, Forrester Cole, Huiwen Chang, Deva Ramanan, William T. Freeman, and Ce Liu. Lasr: Learning articulated shape reconstruction from a monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15980–15989, 2021. 2

[66] Gengshan Yang, Deqing Sun, Varun Jampani, Daniel Vlasic, Forrester Cole, Huiwen Chang, Deva Ramanan, William T Freeman, and Ce Liu. Lasr: Learning articulated shape reconstruction from a monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15980–15989, 2021. 2

[67] Gengshan Yang, Deqing Sun, Varun Jampani, Daniel Vlasic, Forrester Cole, Ce Liu, and Deva Ramanan. Viser: Video-specific surface embeddings for articulated 3d shape reconstruction. In *Advances in Neural Information Processing Systems*, pages 19326–19338. Curran Associates, Inc., 2021. 2

[68] Gengshan Yang, Deqing Sun, Varun Jampani, Daniel Vlasic, Forrester Cole, Ce Liu, and Deva Ramanan. Viser: Video-specific surface embeddings for articulated 3d shape reconstruction. *Advances in Neural Information Processing Systems*, 34:19326–19338, 2021.

[69] Gengshan Yang, Minh Vo, Natalia Neverova, Deva Ramanan, Andrea Vedaldi, and Hanbyul Joo. Banmo: Building animatable 3d neural models from many casual videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2863–2873, 2022. 2

[70] Gengshan Yang, Shuo Yang, John Z Zhang, Zachary Manchester, and Deva Ramanan. Ppr: Physically plausible reconstruction from monocular videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3914–3924, 2023. 3

[71] Chun-Han Yao, Wei-Chih Hung, Yuanzhen Li, Michael Rubinstein, Ming-Hsuan Yang, and Varun Jampani. Lassie: Learning articulated shapes from sparse image ensemble via 3d part discovery. *Advances in Neural Information Processing Systems*, 35:15296–15308, 2022. 3

[72] Hang Yu, Yufei Xu, Jing Zhang, Wei Zhao, Ziyu Guan, and Dacheng Tao. Ap-10k: A benchmark for animal pose estimation in the wild. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021. 2, 3

[73] Cem Yuksel, Scott Schaefer, and John Keyser. Hair meshes. *ACM Transactions on Graphics*, 28(5):1–7, 2009. 3

[74] Egor Zakharov, Vanessa Sklyarova, Michael Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. Human hair reconstruction with strand-aligned 3d gaussians. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XVI*, page 409–425, Berlin, Heidelberg, 2024. Springer-Verlag. 1, 2, 3, 5, 6, 7

[75] Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus Gross, and Thabo Beeler. Groomgen: A high-quality generative hair model using hierarchical latent representations. *ACM Transactions on Graphics (TOG)*, 42(6):1–16, 2023. 3

[76] Yuxiao Zhou, Menglei Chai, Daoye Wang, Sebastian Winberg, Erroll Wood, Kripasindhu Sarkar, Markus Gross, and Thabo Beeler. Groomcap: High-fidelity prior-free hair capture. *ACM Trans. Graph.*, 43(6), 2024. 1, 3

[77] Silvia Zuffi and Michael J Black. Awol: Analysis without synthesis using language. In *European Conference on Computer Vision*, pages 1–19. Springer, 2024. 3

[78] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 3, 4, 6, 7

[79] Silvia Zuffi, Angjoo Kanazawa, Tanya Berger-Wolf, and Michael J Black. Three-d safari: Learning to estimate zebra pose, shape, and texture from images" in the wild". In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5359–5368, 2019. 3