2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

DYMO-Hair: Generalizable Volumetric DYnamics MOdeling for Robot Hair Manipulation

Chengyang Zhao[1], Uksang Yoo[1], Arkadeep Narayan Chaudhury[2], Giljoo Nam[3], Jonathan Francis[1,4], Jeffrey Ichnowski[1], Jean Oh[1]

***Abstract*— Hair care is an essential daily activity, yet it remains inaccessible to individuals with limited mobility and challenging for autonomous robot systems due to the fine-grained physical structure and complex dynamics of hair. In this work, we present DYMO-HAIR, a model-based robot hair care system. We introduce a novel dynamics learning paradigm that is suited for volumetric quantities such as hair, relying on an action-conditioned latent state editing mechanism, coupled with a compact 3D latent space of diverse hairstyles to improve generalizability. This latent space is pre-trained at scale using a novel hair physics simulator, enabling generalization across previously unseen hairstyles. Using the dynamics model with a Model Predictive Path Integral (MPPI) planner, DYMO-HAIR is able to perform visual goal-conditioned hair styling. Experiments in simulation demonstrate that DYMO-Hair's dynamics model outperforms baselines on capturing local deformation for diverse, unseen hairstyles. DYMO-Hair further outperforms baselines in closed-loop hair styling tasks on unseen hairstyles, with an average of 22% lower final geometric error and 42% higher success rate than the state-of-the-art system. Real-world experiments exhibit zero-shot transferability of our system to wigs, achieving consistent success on challenging unseen hairstyles where the state-of-the-art system fails. Together, these results introduce a foundation for model-based robot hair care, advancing toward more generalizable, flexible, and accessible robot hair styling in unconstrained physical environments. More details can be found at: https://dymohair.github.io.**

I. INTRODUCTION

Hair is central to personal identity and self-esteem [1], [2], yet routine care is difficult for individuals with limited mobility due to reduced coordination, strength, and flexibility [3]. To improve accessibility and autonomy, robot hair care systems have been explored [4]–[7], but existing approaches rely on either handcrafted trajectories or rule-based controllers, restricting generalization across diverse hairstyles and goals.

To address these limitations, we propose **DYMO-Hair**, a model-based robot hair care system. Our system is capable of generalizable and flexible visual goal-conditioned hair manipulation, across diverse hairstyles and objectives in unconstrained physical environments. At the core of our system is a dynamics model that captures diverse hair deformations across various hairstyles and combing motions.

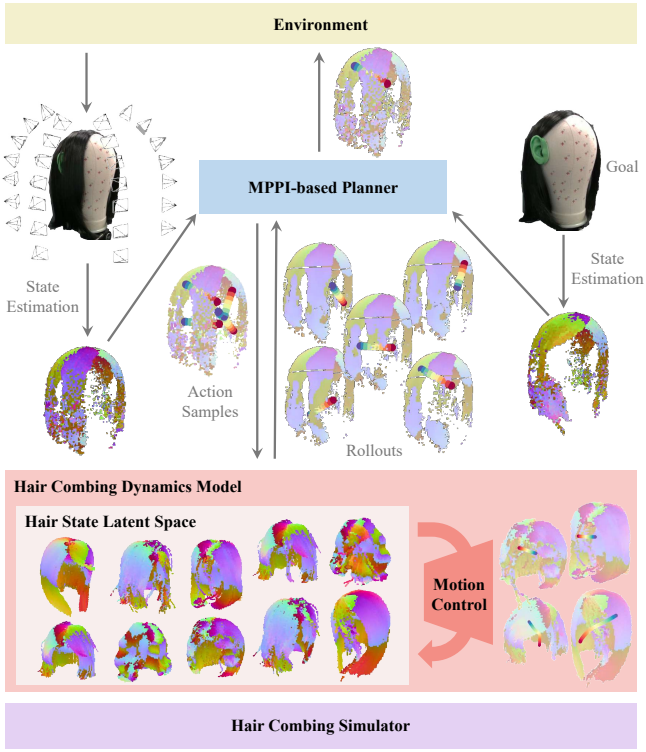


Fig. 1. **DYMO-HAIR Overview.** We introduce **DYMO-Hair**, a unified, model-based robot hair care system. We propose the first 3D volumetric hair-combing dynamics model, featuring a novel learning paradigm. It uses an action-conditioned latent state editing mechanism, coupled with a compact 3D latent space of diverse hairstyles, enabled by our novel hair-combing simulator, for generalizable dynamics modeling. Building on this model, we develop DYMO-Hair with a MPPI-based planner for closed-loop visual goal-conditioned hair styling.

For deformable objects like hair, complex structures and unobservable properties make accurate dynamics modeling difficult. While analytical physics-based models exist, they are computationally expensive and impractical for real-time control, motivating the use of learning-based neural dynamics as proxies [8]–[10]. However, hair poses unique challenges: 1) *Representation.* Low-resolution point clouds cannot capture strand-level geometry. 2) *Structure.* Graph-based methods scale poorly as point counts increase for higher resolution. 3) *Supervision.* Global metrics miss fine-scale deformations, while point-wise correspondence is impractical for strands. 4) *Data.* Hair entanglement makes real-world data collection slow and difficult to reset across styles.

To address these challenges, we introduce a novel

[1] Chengyang Zhao, Uksang Yoo, Jonathan Francis (by courtesy), Jeffrey Ichnowski, and Jean Oh are with Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. {chengyaz, uyoo, jmf1, jichnows, hyaejino}@andrew.cmu.edu
[2] Arkadeep Narayan Chaudhury is with Epic Games, Inc., Pittsburgh, Pennsylvania, USA. arkadeep.chaudhury@epicgames.com
[3] Giljoo Nam is with Meta Codec Avatars Lab, Pittsburgh, Pennsylvania, USA. giljoonam@meta.com
[4] Jonathan Francis is with Bosch Center for Artificial Intelligence, Pittsburgh, Pennsylvania, USA.

paradigm for generalizable volumetric hair dynamics modeling. We present the first 3D hair-combing dynamics model that leverages large-scale diverse synthetic data for hair dynamics learning and generalizes across various hairstyles. We represent hair as a high-resolution volumetric occupancy grid with a 3D orientation field to capture both hair position and local strand flow, which offers more geometric details and structural information of hair than sparse point clouds. It also allows dense supervision on both occupancy and orientation, providing sufficient local signals for the model to capture fine-grained deformations during learning. Our key innovation, inspired by ControlNet [11], is to pre-train a compact 3D latent space for diverse hair states and to introduce a control branch that models dynamics as action-conditioned state editing, enabling significantly improved generalizability through large-scale pre-training. To avoid time-consuming real-world data collection required by the pre-training, we further develop a hair-combing simulator based on Genesis [12]. It leverages a novel formulation of the position-based dynamics (PBD) method for strand-level, contact-rich hair simulation, enabling efficient large-scale generation of visually-realistic and physically-plausible synthetic dynamics data across diverse hairstyles. Experiments in simulation demonstrate that our model outperforms baselines on generalizable hair dynamics modeling for local hair deformation across diverse unseen hairstyles.

Building on our dynamics model, we introduce DYMO-Hair, a unified, model-based robot hair care system for visual goal-conditioned hair styling. We adopt a Model Predictive Control (MPC) framework, using a Model Predictive Path Integral (MPPI)-based planner to optimize an action trajectory that minimizes the geometric distance between predicted hair states and the objective [13], [14]. Simulation experiments on diverse unseen hairstyles show that DYMO-Hair achieves superior effectiveness and generalizability for closed-loop hair styling compared to all system baselines, with an average of 22% lower final geometric error and 42% higher absolute success rate than the state-of-the-art system. Real-world demonstrations further exhibit zero-shot transferability of DYMO-Hair to physical wigs, achieving consistent success on challenging unseen hairstyles where the state-of-the-art system fails.

To summarize, our contributions are:

- A study of model-based approaches for robot hair manipulation.
- **DYMO-Hair**, a unified, model-based robot system for visual goal-conditioned hair styling, evaluated across diverse hairstyles in simulation and real-world settings.
- A 3D generalizable volumetric dynamics model for hair combing.
- A hair simulator with a novel PBD method for strand-level, contact-rich hair-combing simulation.

II. Related Works

A. Dynamics Modeling for Deformable Object Manipulation

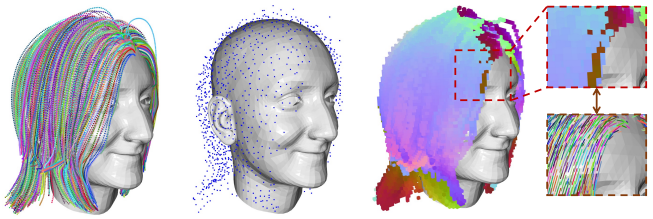
Physics-based modeling offers a first-principles dynamics formulation [15], [16], but is often impractical for deformable objects due to complex internal structures and unobservable material properties, limiting its real-time use in manipulations. Learning-based methods have gained attention as alternatives [10], with graph-based modeling proving effective in capturing spatial relationships in deformable dynamics [17], [18] for various object types like plasticine [8], [13], cloth [19], rope [20], and plush toys [14]. However, these approaches often assume easily observable deformations, use low-resolution state representations, and face scalability issues, which are unsuitable for hair modeling. Recent works [9], [10] explore particle-grid hybrid or diffusion-based modeling to improve scalability, but rely on point-wise correspondence for dense supervision, which is impractical for hair. We propose a new dynamics learning paradigm for hair that preserves geometric and structural details, avoids the scalability limits of graph-based methods and the point-wise correspondence requirement for dense supervision, and enables more generalizable dynamics modeling.

B. Robot System for Hair Manipulation

Robot hair manipulation integrates multiple research areas and remains underexplored due to its inherent complexity. One line of work [6] focuses on mechanical and human-robot interaction aspects, introducing a soft robot manipulator for a safer and more user-friendly system design. Another line of work approaches it algorithmically, often relying on 2D observations and rule-based strategies. Some studies [5], [21] tackle detangling with sensorized brushes using visual and force feedback, but rely on constrained settings and specific initial conditions. Another method [4] plans combing trajectories aligned with hair flow, but lacks goal-driven flexibility. Recent work [7] introduces rule-based, goal-conditioned planning based on 2D orientation differences, but is restricted to particular hairstyles and goals. We advance robot hair manipulation by capturing 3D hair states, explicitly modeling their 3D dynamics, and developing a model-based system for generalizable, flexible, goal-conditioned hair manipulation.

C. Hair Dynamics Simulation

Hair dynamics simulation is challenging due to hair's fine structure, complex inter-strand interactions, and large strand count. Physics-based methods have been explored for both clump-level and strand-level modeling. Clump-level methods [22], [23] represent hair as large bundles, achieving computational efficiency suitable for real-time applications but failing to capture strand-level hair behavior. Strand-based methods, on the other hand, use physically more accurate representations such as mass-spring models [24], [25] and Kirchhoff rod models [26], [27], enabling more accurate modeling of strand-level dynamics but remaining computationally expensive and inefficient. Recently, learning-based methods have been explored to accelerate simulation [28], [29], reduce the need for detailed physical modeling [30], and improve generalizability [31]. While efficient and visually-realistic, they still fall short of achieving strand-level, physically-accurate modeling. We propose a novel PBD method for strand-level hair simulation



(a) Strands (b) Point Cloud (c) Volumetric Grid

Fig. 2. **Comparison of Hair State Representations.** (a) Colors distinguish individual hair strands. (b) We show a resolution of 2K points, the maximum used for point cloud–based methods in our experiments (see Sec. VII-B for more details). (c) We show $64 \times 64 \times 128$ grids with a voxel size of about 5 mm. Colors denote local strand orientations. Red dashed box: a zoomed-in region. Brown dashed box: the corresponding local strand segments.

to enable efficient, both visually-realistic and physically-plausible simulation of contact-rich hair-combing dynamics, supporting synthetic data generation and closed-loop hair manipulation experiments.

III. Problem Formulation

We develop a model-based robot hair manipulation system for hair styling using a preset combing tool. Given a user-specified visual goal $\mathcal{G}$, at each timestep t the system receives the current observation o_t from the environment and estimates the underlying hair state s_t. The system then selects the next-step action a_t that best advances toward $\mathcal{G}$, guided by a dynamics model f_{dyn} capturing the hair deformation behavior under combing, *i.e.*, $\hat{s}_{t+1} = f_{\text{dyn}}(\hat{s}_t, \hat{a}_t)$. The action space consists of 3D combing motions, represented as sequences of tool positions and orientations. After executing a_t, the system receives the updated observation o_{t+1} and iterates this process until $\mathcal{G}$ is reached.

IV. Hair Combing Dynamics Modeling

A. State Representation

Designing an effective state representation is fundamental to a dynamics model, as it should not only capture task-relevant information but also ensure robustness and efficiency for estimation. In the context of hair, a strand is the fundamental physical structure from which hair is formed, serving as a key geometric cue for describing hair state. While humans can perceive strand geometry almost instantly, even the most advanced methods for full 3D strand reconstruction require several minutes [32], [33], making them unsuitable for a real-time robot manipulation system.

To achieve both fidelity and efficiency, we draw inspiration from prior computer vision work [34] and represent the hair as a set of dense 3D points, each with a position and a unit direction vector describing the local strand orientation. This captures both the spatial distribution and the flow direction of hair, and can be estimated robustly and efficiently from multi-view RGB-D data. Compared with widely used standard point clouds for dynamics modeling, it encodes richer geometric and structural information while avoiding the computational cost of strand-level reconstruction. For more structured processing, following [35], we further discretize it into a high-resolution volumetric occupancy grid with a 3D orientation field as our final state representation:

$$s_t = (occ_t, ori_t),\ occ_t \in \{0,1\}^{\mathcal{V}_0},\ ori_t \in [0,1]^{\mathcal{V}_0 \times 3},$$

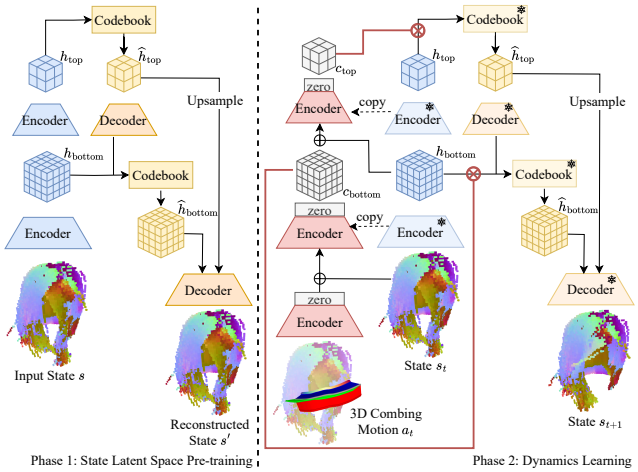


Fig. 3. **DYMO-Hair's Dynamics Model Overview. Left**: State latent space pre-training. A 3D volumetric hierarchical model with vector quantization enables compact compression while preserving detailed representation capability. **Right**: Dynamics learning. The pre-trained model is adapted to capture hair dynamics in a ControlNet-style framework, formulating dynamics as action-conditioned editing in the pre-trained state latent space. *zero*: zero-convolution; *copy*: weight copying for initialization; $\oplus$: element-wise addition; $\otimes$: 3D attention-based feature fusion. In this phase, only the motion encoding path is trainable, with all pre-trained components frozen.

where $\mathcal{V}_0$ is the spatial resolution of the grid. This voxelized form allows us to balance computational efficiency with geometric fidelity: it enables the use of highly optimized neural operators and dense voxel-level supervision in the learning process, while maintaining acceptable accuracy when resolution is sufficiently high. An illustration of our state representation is shown in Fig. 2.

B. State Latent Space Pre-training

Inspired by recent advances in conditioned image and video generation [11], [36], [37], we propose a novel ControlNet [11]-style two-phase paradigm for hair dynamics learning that is compatible with our high-resolution volumetric state representation and enables more generalizable modeling across diverse hairstyles.

In Phase 1, we use pre-training to encode diverse hairstyles with various deformations during combing into a unified, compact 3D latent space. As shown in Fig. 3, we adopt a hierarchical model structure, analogous to VQ-VAE-2 [38], in the 3D volumetric setting to achieve both compact state compression and detailed representation capability. The model is pre-trained for state reconstruction. Given a volumetric state representation s of any hairstyle with deformation, the model concatenates the occupancy and orientation components together and hierarchically encode it into lower resolutions with two 3D encoders, producing latent embeddings $h_{\text{bottom}} \in \mathbb{R}^{\mathcal{V}_1 \times D_1}$ and $h_{\text{top}} \in \mathbb{R}^{\mathcal{V}_2 \times D_2}$, where $\mathcal{V}_0 > \mathcal{V}_1 > \mathcal{V}_2$. The top-level codebook [38], [39] quantizes h_{top} into $\hat{h}_{\text{top}}$ by replacing each entry with its nearest codebook vector, which is further decoded into resolution $\mathcal{V}_1$ to serve as a prior for quantizing h_{bottom}. A decoder decodes the quantized $\hat{h}_{\text{bottom}}$ and the up-sampled $\hat{h}_{\text{top}}$ into the final reconstruction s'. This hierarchical design allows the two quantized latent spaces, *i.e.*, the bottom- and top-level codebooks, to capture

complementary local and global information respectively, enabling better state modeling and reconstruction than single-level quantization [39]. We use exponential moving averages (EMA) to update codebooks progressively during training.

C. Dynamics: Action-conditioned State Editing

In Phase 2, we formulate the dynamics as an action-conditioned state editing process, and adapt the pre-trained model with a ControlNet [11]-style framework to leverage the pre-trained state latent space for capturing hair dynamics.

The pre-trained state encoding path takes the initial state s_t as input and encodes it progressively, as in Phase 1. For dynamics modeling, we introduce an additional trainable motion encoding path that processes the combing motion a_t as a control signal in parallel with the state, while keeping all other pre-trained components, including the codebooks, frozen. This path consists of three cascaded encoders: the first preserves resolution, while the other two, sharing the state encoders' architecture, progressively compress the signal to lower resolutions, producing control embeddings $c_{\text{bottom}} \in \mathbb{R}^{\mathcal{V}_1 \times D_1}$ and $c_{\text{top}} \in \mathbb{R}^{\mathcal{V}_2 \times D_2}$. Following ControlNet [11], the path employs weight copying, zero-convolution, and cross-path feature fusion mechanisms to integrate state and motion features at fine-grained voxel level while stabilizing early training. The control embeddings are then fused with state embeddings h_{bottom} and h_{top} to perform edits in the pre-trained state latent space for the dynamics behavior. The edited embeddings are finally quantized and decoded into the end state s_{t+1}.

To enforce fine-grained spatial alignment between the state and the motion, we convert the motion into a volumetric grid matching the state resolution before feed it into the first motion encoder. Specifically, we sample the motion into up to K uniformly spaced key tool poses and, for each tool pose, compute the shortest distance from every voxel center in the grid to the tool's center line. Voxels outside a central cylindrical region, defined by the center line and a preset contact radius, are discarded to form a prior of the local contact region. This yields a time-indexed volumetric distance map in $\mathbb{R}^{\mathcal{V}_0 \times K \times 2}$, with distance and validity in the last dimension. The map is then fed into the motion encoding path for dynamics modeling.

D. Supervision

We use the same composite loss for both phases. For occupancy, we combine the focal loss [40] and the soft Dice loss to address class imbalance and encourage volumetric overlap with ground truth, as occupied regions typically form a thin shell within the grid. For orientation, we use the L1 loss on normalized orientation vectors for each occupied voxel, with symmetry handling for directional equivalence, *i.e.*, v and $-v$ represent the same local strand orientation. For codebook matching, we adopt the EMA commitment loss from VQ-VAE [39], keeping latent entries close to their matched codebook vectors. The total loss is a weighted sum of these terms, jointly enforcing geometric accuracy, directional consistency, and latent representation quality.

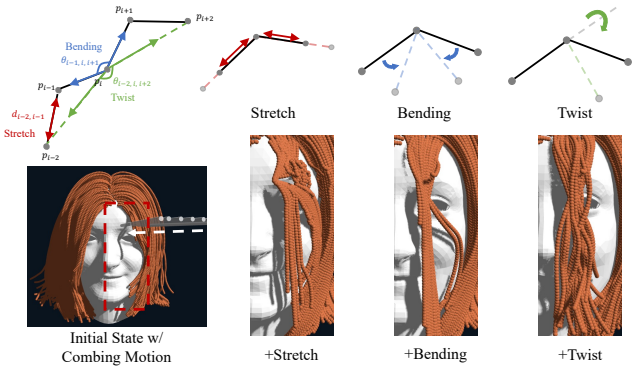


Fig. 4. **Constraints for PBD-based Strand-level, Contact-rich Hair Combing Simulation. Top:** Each constraint's formulation and intended effect. **Bottom:** Simulation results under progressive constraint addition. Starting from the initial state (far left), a combing motion is applied along the white dashed arrow. The red dashed box marks the contact-rich region, with its simulation results shown on the right. With all three constraints, the hair maintains a realistic shape; the twist constraint, in particular, preserves curvature and prevents gravity-induced oversmoothing.

V. Hair Dynamics Simulation

Existing neural dynamics models are often trained directly on real-world data for each object. However, for hair, fine-grained strand entanglement and deformation can create messy, hard-to-reset states, making real-world data collection highly time-consuming. Moreover, our approach requires large-scale data covering diverse hairstyles and deformations to build a strong, representative state latent space during pre-training, which further increases the impracticality of collecting such data in the real world. In this paper, we use fully synthetic data for dynamics learning. We develop a novel GPU-accelerated simulator based on Genesis [12] that enables efficient, both visually-realistic and physically-plausible strand-level, contact-rich hair-combing simulation.

At its core, our simulator uses a novel PBD method for strand-level hair simulation. We model each hair strand as a 3D particle sequence $[p_0, p_1, \cdots, p_n]$, where p_0 is the root connected to the scalp and fixed. Thousands of strands are simulated in parallel. We use three physics-informed constraints to model inner-strand physical properties. Let $d(p_i, p_j) = \|p_i - p_j\|$ denote the Euclidean distance between particles p_i and p_j, and $\theta(p_i, p_j, p_k) = \arccos\left(\frac{(p_i - p_j)\cdot(p_k - p_j)}{\|p_i - p_j\|\,\|p_k - p_j\|}\right)$ denote the angle at p_j formed between the vectors $p_i - p_j$ and $p_k - p_j$. The corresponding rest-state values are denoted $d^0(p_i, p_j)$ and $\theta^0(p_i, p_j, p_k)$.

- **Stretch:** Neighboring particles preserve their rest distances: $\forall\, i \in [0, n-1]$,

$$C_{\text{stretch}}(p_i, p_{i+1}) = d(p_i, p_{i+1}) - d^0(p_i, p_{i+1}).$$

- **Bending:** Local in-plane bending is regulated by constraining consecutive particle angles: $\forall\, i \in [1, n-1]$,

$$C_{\text{bending}}(p_{i-1}, p_i, p_{i+1}) = \theta(p_{i-1}, p_i, p_{i+1}) - \theta^0(p_{i-1}, p_i, p_{i+1}).$$

- **Twist:** The 3D twist constraint of the strand is approximated by multiple skip-connected 2D bending constraints with a fixed index gap $k > 1$: $\forall\, i \in [k, n-k]$,

$$C_{\text{twist}}(p_{i-k}, p_i, p_{i+k}) = \theta(p_{i-k}, p_i, p_{i+k}) - \theta^0(p_{i-k}, p_i, p_{i+k}).$$

They collectively form multiple 3D-intersecting 2D constraint planes to approximate the full 3D twist behavior.

The constraints with their effects are illustrated in Fig. 4. During simulation, the particle positions are iteratively updated toward satisfying all $C = 0$. Note that our twist model is a heuristic approximation, as true hair twist arises from the strand's internal physical microstructure. While more complex techniques like Kirchhoff rod models may offer more physically-accurate twist simulation, they typically remain computationally expensive and inefficient. Here we use the heuristics to balance accuracy and computation efficiency. We model inter-strand and hair-tool contacts at the particle level with standard PBD collision and friction handling.

Our simulator enables efficient simulation of visually-realistic and physically-plausible hair-combing dynamics across diverse hairstyles, supplying abundant data for model learning and serving as a testbed for closed-loop experiments with our hair styling system (see Sec. VII-C).

VI. Model-based Robot Hair Styling System

Building on our hair-combing dynamics model, we develop DYMO-Hair, a model-based robot hair care system for hair styling, capable of handling diverse hairstyles and goal configurations. As illustrated in Fig. 1, the system takes multi-view RGB-D observations from the environment and estimates the current volumetric hair state, following [34]. Given the current state and the visual goal configuration, the MPPI-based planner samples candidate actions, rolls out the dynamics model to predict possible future outcomes, and optimizes an action trajectory that minimizes the geometric distance between the predicted states and the goal, following prior neural dynamics-based planners [13], [14], [20]. A chunk of actions is executed, after which the system updates its observation and repeats the loop until the goal is reached.

VII. Experiments

A. Experiment Setup

Hairstyles Used. Fig. 5 shows the 10 hairstyles we use in simulation to train the hair dynamics model. For dynamics model and closed-loop system evaluation, we use 7 unseen hairstyles in simulation, and 2 physical wigs with different hairstyles for real-world test. The hairstyles and the mannequin head in simulation are from USC-HairSalon [41].

Simulation Setup. The simulation workspace consists of a mannequin head and a hair model with a realistic hairstyle, as shown in Fig. 5. The tool for manipulation is a thin cylinder. For simulation stability, we preprocess each strand for an appropriate linear density while preserving its curvature.

Real-world Setup. Fig. 5 also shows our real-world workspace, consisting of a 25 cm×19 cm×30 cm mannequin head with a physical wig. The head is painted with calibration marks. We use a UFactory 850 6-axis robot arm with a 3D-printed tool of length 16 cm and tip radius 4 mm, designed to resemble the cylindrical tool used in simulation to reduce the sim-to-real gap. We add a TPU-printed deformable finger tip on it to ensure better contact between the tool and the head. For perception, we use a wrist-mounted RealSense D405 RGB-D camera to capture multi-view observations of the hair by varying the robot pose. Both the head and the camera are calibrated to the robot frame.

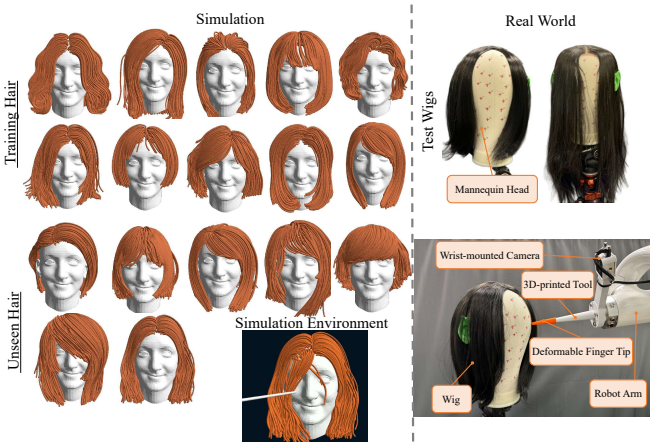


Fig. 5. **Experiment Setup. Left**: Synthetic hair used for training and evaluation; simulation environment. **Right**: Real-world setup for evaluation.

B. Hair Combing Dynamics Learning

Data. We construct a large-scale synthetic hair-combing dynamics dataset using our simulator for generalizable dynamics learning. For each of 10 training hairstyles, we randomly sample over 1K diverse tool motions, rolling them out in simulation to either mess or clean the hair, and recording intermediate states and sub-motions. After filtering low-quality samples, we obtain a 10K training dataset. For latent space pre-training, we combine these synthetic hair states with deformed variants from our dataset and an existing large-scale neat hairstyle database [42], [43], yielding 47K hair states to build a strong and representative latent space. Dynamics learning is then trained on our 10K combing dataset. For evaluation, we generate a separate dynamics dataset from 7 unseen hairstyles, producing 800 filtered transitions for testing the model's generalizability.

Baselines. We compare our method against three baselines:

- **PC-GNN**: represents hair as point clouds with per-point orientation, applies a GNN-based structure without pre-training, following the most common paradigm for neural deformable object dynamics [13].
- **V-UNet**: represents hair as volumetric grids as ours does, but uses a UNet-based architecture with pyramid fine-grained state-action fusion and no pre-training; serves as an ablation to test the benefit of our pre-trained latent space.
- **V-FiLM**: represents hair as volumetric grids as ours does, uses a FiLM [44]-style state-action fusion mechanism, and is trained on the same pre-trained latent space as ours; serves as an ablation to test the effectiveness of our ControlNet-style design for fine-grained state-action fusion.

Evaluation Metrics. We convert the outputs of all volumetric methods into point clouds with per-point orientation and down-sample them to the same resolution as **PC-GNN** for fair comparison. Three metrics are used: 1) $\mathrm{CD}_{\mathrm{point}}$: Chamfer Distance (CD) between predicted and ground-truth point clouds, ignoring orientation. 2) $\mathrm{Err}_{\mathrm{ori}}$: point-level orientation error, computed as the average angular difference between

TABLE I

HAIR COMBING DYNAMICS MODEL EVALUATION ON UNSEEN HAIR

Method	CD_{point} ↓		Err_{ori} ↓		CD_{strand} ↓	
	mean	90th	mean	90th	mean	90th
PC-GNN [13]	0.0814	0.1359	13.34	30.16	0.1052	0.1983
V-UNet	0.0792	0.1345	14.25	31.00	0.1047	0.1966
V-FiLM	0.0807	0.1334	13.60	29.59	0.1065	0.2011
Ours	**0.0775**	**0.1240**	**12.03**	**26.12**	**0.1005**	**0.1878**

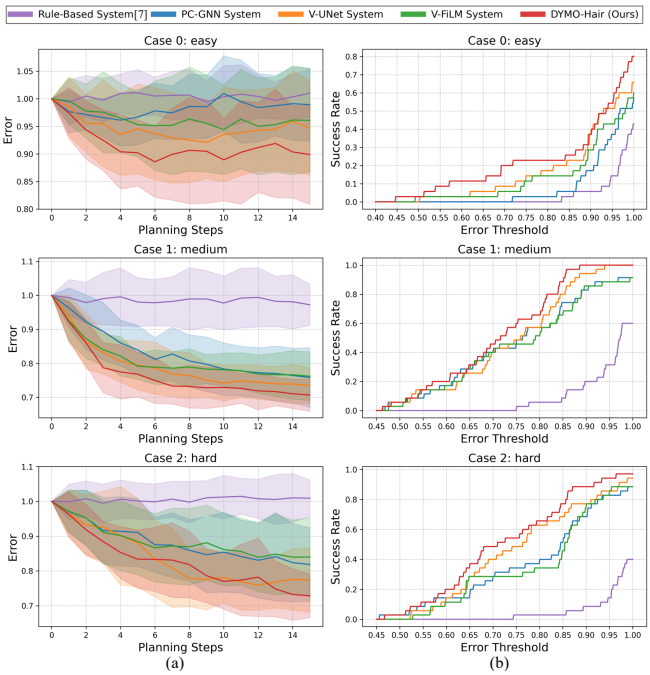


Fig. 6. **Quantitative Results for Closed-loop Hair Styling in Simulation.** Experiments are conducted on 7 unseen hairstyles, each with 3 cases repeated 5 times. (a) Error curves over planning steps, where solid lines and shaded regions denote mean and standard deviation across hairstyles. Faster error reduction and lower final error indicate higher effectiveness. (b) Success rate curves w.r.t. error thresholds used to determine success, reflecting the distribution of final errors. Curves closer to the top-left correspond to more low-error outcomes and better performance. In both (a) and (b), error is defined as relative error: the ratio of the current strand-level distance to the initial strand-level distance, providing a unified metric across hairstyles with varying absolute error magnitudes.

predicted orientations and those of the closest ground-truth points, with symmetry tolerated. Predicted points farther than 2 cm from the ground truth are discarded to avoid invalid matches. 3) CD_{strand}: strand-level CD, where predicted point clouds with orientations are reconstructed into strand segments to evaluate the strand structure they convey, measuring on average how well each predicted segment aligns with its closest ground-truth segment and vice versa.

Implementation Details. For **PC-GNN**, we use 2K-point resolution, the highest feasible while maintaining enough message-passing capability under the same computation budget as other methods. For all volumetric methods, we use $64 \times 64 \times 128$ grids with a $\sim$5 mm voxel size. All models are trained to convergence on 2 NVIDIA RTX 4090 GPUs.

Results. We evaluate all methods on 7 unseen hairstyles to test generalizability, with results shown in Tab. I. Since in hair-combing scenarios deformations are highly localized, occurring mainly near the comb while most of the hair remains remains stable and unchanged, evaluating the entire hairstyle can obscure local effects. We therefore focus on the near-motion region and report both the mean, reflecting overall performance, and the 90th percentile, which captures sparse but significant deformations, e.g., deformations that occur in only a few strands, and reduces averaging bias. The results show that our method outperforms all baselines in capturing local hair deformation behavior in the near-motion region for unseen hairstyles. Our model surpasses the widely used **PC-GNN** paradigm, underscoring its suitability for generalizable hair dynamics. Compared with other voxel-based methods, our model leverages latent space pre-training for stronger generalization than **V-UNet**, and finer state-action fusion for more accurate future state prediction than **V-FiLM** based on the same pre-trained state latent space. Overall, the results highlight the effectiveness and generalizability of our model on hair-combing dynamics, validating our design choices of pre-training and fine-grained state-action fusion.

C. Closed-loop Goal-conditioned Hair Styling

Baselines. We compare DYMO-Hair against four baselines. **Rule-based System**: the only existing method for visual goal-conditioned robot hair styling [7], which uses a single front-view 2D observation and handcrafted rules to derive actions from 2D orientation differences between current and goal states. **PC-GNN System**, **V-UNet System**, and **V-FiLM System**: three model-based systems that replace DYMO-Hair's dynamics model with the baseline models described in Sec. VII-B, while keeping all other parts unchanged.

Implementation Details. All systems for visual goal-conditioned hair styling assume geometrically grounded goal configurations that are well-aligned with the states, consistent with existing neural dynamics-based deformable object manipulation methods [13], [14], [20]. For all model-based systems, the MPPI planning cost is defined as the strand-level distance between strand segments reconstructed from the predicted future states and the goal state, incorporating both point and orientation predictions in a unified manner.

Simulation Experiments. We first evaluate DYMO-Hair's effectiveness and generalizability for visual goal-conditioned closed-loop hair styling thoroughly in simulation, using diverse unseen hairstyles and varying initial states. Specifically, we test on 7 unseen hairstyles with 3 cases per hairstyle at different levels of messiness, and repeat each case 5 times with different random seeds to reduce randomness. All experiments are conducted with a maximum budget of 15 steps. Quantitative results in Fig. 6 show that all model-based methods outperform **Rule-based System**. Among them, **DYMO-Hair** achieves the best performance, with an average of 22% lower final geometric error and 42% higher absolute success rate across three cases (using thresholds of 0.90, 0.70, and 0.70, chosen as reasonable success criteria based on experiments and visualization), demonstrating both the advantage of incorporating a dynamics model for improving system capability and generalizability, and the superior effectiveness of our advanced dynamics model in boosting closed-loop manipulation at the system level. Qualitative results are shown in Fig. 7.

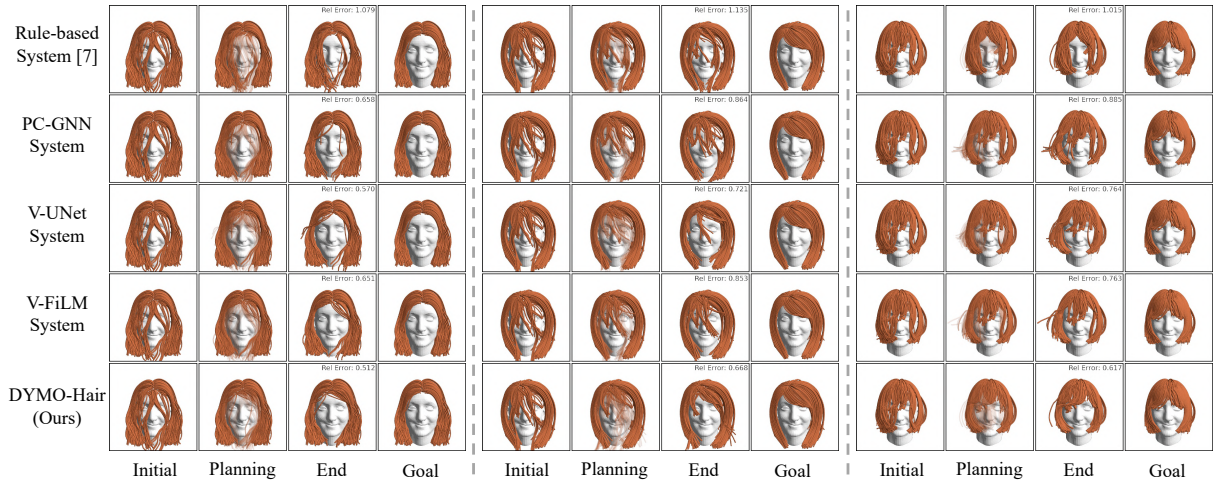


Fig. 7. **Qualitative Results for Closed-loop Hair Styling in Simulation.** Three *hard* cases of different unseen hairstyles are shown, with columns (left to right) illustrating the initial state, the intermediate planning steps, the end state with relative error, and the goal.

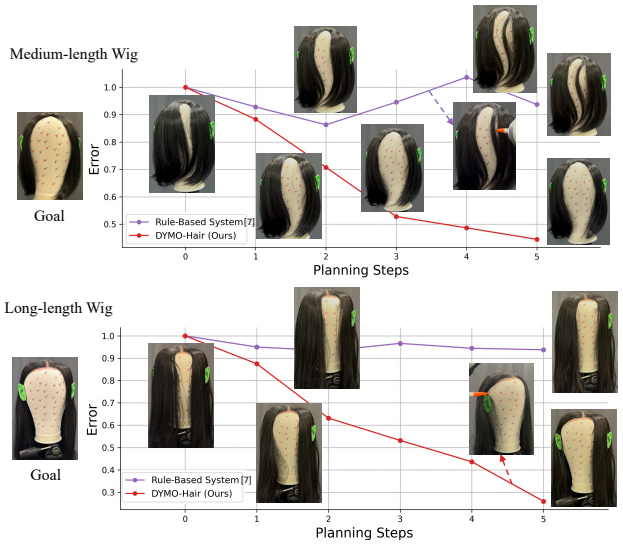


Fig. 8. **Results for Closed-loop Hair Styling in the Real World.** For each case, the visual goal is shown on the left, with key observations and actions for each method displayed alongside the curve. Error is defined as relative error: the ratio of the current strand-level distance to the initial distance.

Real-world Experiments. We further evaluate the zero-shot transferability of DYMO-Hair from simulation to unconstrained real-world settings. Experiments on 2 physical wigs with different hairstyles, using the setup in Fig. 5, compare our system against **Rule-based System**, the state-of-the-art robot hair styling system. As shown in Fig. 8, DYMO-Hair consistently outperforms the baseline. While the **Rule-based System** fails in both cases, **DYMO-Hair** achieves rapid, effective progress toward the target style even in the challenging long-length wig case, where success requires pushing hair behind the ear to achieve a long-horizon spatial transformation. **Rule-based System** fails mainly due to two weaknesses: 1) its handcrafted rules rely on strand-level correspondence for geometric difference computation, requiring accurate 2D orientation maps and strand tracking. These dependencies are brittle as errors accumulate with longer hair and performance degrades in unconstrained settings with variations in lighting, resolution, *etc.*; and 2) it relies solely on a single front-view observation, which is inadequate for styles demanding long-horizon changes across both front and side views, as in the long-length wig case in Fig. 8. In contrast, DYMO-Hair leverages multi-view 3D hair state estimations and evaluates geometric differences without the need for strand-level correspondence. This design is more robust to environmental variations, generalizes across hair lengths, and effectively captures long-horizon goals. The dynamics model further provides predictive capability, improving both action efficiency and manipulation effectiveness.

VIII. Conclusion

This paper presents the first study on model-based robot hair manipulation. We introduce the first 3D hair-combing dynamics model with a novel volumetric learning paradigm for generalizable dynamics modeling. We also develop a simulator with a novel PBD method for strand-level, contact-rich hair-combing simulation to support data requirements for large-scale pre-training. Together, these contributions yield the first unified model-based robot hair care system, DYMO-Hair, for visual goal-conditioned hair styling, generalizable to novel hairstyles and evaluated in both simulation and the real world. Several limitations remain for future work. First, we focus on model-based planning with limited consideration of human-robot interaction aspects, such as enforcing action-space constraints to avoid safety-critical regions like the eyes to enhance real-world usability. Second, the system assumes privileged mannequin head information and perfect calibration; incorporating online head estimation would improve practicality. Finally, we use a simple 3D-printed combing tool; replacing it with advanced designs like soft robot fingers [6], [45] could further enhance usability.

Acknowledgments

We would like to thank Feiyu Zhu and John Z. Zhang for their valuable feedback and discussions. This work was supported by NSF IIS-2112633 and NSF Graduate Research Fellowship under Grant No. DGE2140739.

REFERENCES

[1] C. M. McFarquhar and M. J. Lowis, "The effect of hairdressing on the self-esteem of men and women," *Mankind quarterly*, vol. 41, no. 2, p. 181, 2000. 1

[2] U. Yoo, N. Dennler, S. Patil, J. Oh, and J. Ichnowski, "Inclusion in assistive haircare robotics: Practical and ethical considerations in hair manipulation," *arXiv preprint arXiv:2411.05137*, 2024. 1

[3] A. Waugh, "Personal care, sensory impairment and unconsciousness." *Foundations of Nursing Practice: Fundamentals of Holistic Care*, p. 363, 2013. 1

[4] N. Dennler, E. Shin, M. Matarić, and S. Nikolaidis, "Design and evaluation of a hair combing system using a general-purpose robotic arm," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 3739–3746. 1, 2

[5] J. Hughes, T. Plumb-Reyes, N. Charles, L. Mahadevan, and D. Rus, "Detangling hair using feedback-driven robotic brushing," in *2021 IEEE 4th International Conference on Soft Robotics (RoboSoft)*, 2021, pp. 487–494. 1, 2

[6] U. Yoo, N. Dennler, E. Xing, M. Matarić, S. Nikolaidis, J. Ichnowski, and J. Oh, "Soft and compliant contact-rich hair manipulation and care," in *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2025, pp. 610–619. 1, 2, 7

[7] S. Kim, N. Kanazawa, S. Hasegawa, K. Kawaharazuka, and K. Okada, "Front hair styling robot system using path planning for root-centric strand adjustment," in *2025 IEEE/SICE International Symposium on System Integration (SII)*, 2025, pp. 544–549. 1, 2, 6

[8] H. Shi, H. Xu, S. Clarke, Y. Li, and J. Wu, "Robocook: Long-horizon elasto-plastic object manipulation with diverse tools," *arXiv preprint arXiv:2306.14447*, 2023. 1, 2

[9] K. Zhang, B. Li, K. Hauser, and Y. Li, "Particle-grid neural dynamics for learning deformable object models from rgb-d videos," *arXiv preprint arXiv:2506.15680*, 2025. 1, 2

[10] T. Tian, H. Li, B. Ai, X. Yuan, Z. Huang, and H. Su, "Diffusion dynamics models with generative state estimation for cloth manipulation," *arXiv preprint arXiv:2503.11999*, 2025. 1, 2

[11] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *IEEE International Conference on Computer Vision (ICCV)*, 2023. 2, 3, 4

[12] G. Authors, "Genesis: A generative and universal physics engine for robotics and beyond," December 2024. [Online]. Available: https://github.com/Genesis-Embodied-AI/Genesis 2, 4

[13] H. Shi, H. Xu, Z. Huang, Y. Li, and J. Wu, "Robocraft: Learning to see, simulate, and shape elasto-plastic objects with graph networks," *arXiv preprint arXiv:2205.02909*, 2022. 2, 5, 6

[14] M. Zhang, K. Zhang, and Y. Li, "Dynamic 3d gaussian tracking for graph-based neural dynamics modeling," *arXiv preprint arXiv:2410.18912*, 2024. 2, 5, 6

[15] T. Tang, C. Wang, and M. Tomizuka, "A framework for manipulating deformable linear objects by coherent point drift," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3426–3433, 2018. 2

[16] S. Tiburzio, T. Coleman, and C. Della Santina, "Model-based manipulation of deformable objects with non-negligible dynamics as shape regulation," *arXiv preprint arXiv:2402.16114*, 2024. 2

[17] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, "Learning mesh-based simulation with graph networks," *arXiv preprint arXiv:2010.03409*, 2021. 2

[18] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, "Learning to simulate complex physics with graph networks," in *International conference on machine learning*. PMLR, 2020, pp. 8459–8468. 2

[19] X. Lin, Y. Wang, Z. Huang, and D. Held, "Learning visible connectivity dynamics for cloth smoothing," in *Conference on Robot Learning*. PMLR, 2022, pp. 256–266. 2

[20] K. Zhang, B. Li, K. Hauser, and Y. Li, "Adaptigraph: Material-adaptive graph-based neural dynamics for robotic manipulation," *arXiv preprint arXiv:2407.07889*, 2024. 2, 5, 6

[21] T. B. Plumb-Reyes, N. Charles, and L. Mahadevan, "Combing a double helix," *arXiv preprint arXiv:2103.05211*, 2021. 2

[22] C. K. Koh and Z. Huang, "A simple physics model to animate human hair modeled in 2d strips in real time," in *Proceedings of the Eurographic Workshop on Computer Animation and Simulation*. Berlin, Heidelberg: Springer-Verlag, 2001, p. 127–138. 2

[23] K. Wu and C. Yuksel, "Real-time hair mesh simulation," in *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ser. I3D '16, 2016, p. 59–64. 2

[24] R. E. Rosenblum, W. E. Carlson, and E. Tripp III, "Simulating the structure and dynamics of human hair: Modelling, rendering and animation," *The Journal of Visualization and Computer Animation*, vol. 2, no. 4, pp. 141–148, 1991. 2

[25] A. Selle, M. Lentine, and R. Fedkiw, "A mass spring model for hair simulation," *ACM Trans. Graph.*, vol. 27, no. 3, p. 1–11, Aug. 2008. 2

[26] M. Bergou, M. Wardetzky, S. Robinson, B. Audoly, and E. Grinspun, "Discrete elastic rods," in *ACM SIGGRAPH 2008 Papers*, ser. SIGGRAPH '08, 2008. 2

[27] T. Kugelstadt and E. Schömer, "Position and orientation based cosserat rods," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '16. Eurographics Association, 2016, p. 169–178. 2

[28] Q. Lyu, M. Chai, X. Chen, and K. Zhou, "Real-time hair simulation with neural interpolation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 4, pp. 1894–1905, 2020. 2

[29] T. Stuyck, G. W.-C. Lin, E. Larionov, H. yu Chen, A. Bozic, N. Sarafianos, and D. Roble, "Quaffure: Real-time quasi-static neural hair simulation," *arXiv preprint arXiv:2412.10061*, 2025. 2

[30] Z. Wang, G. Nam, T. Stuyck, S. Lombardi, C. Cao, J. Saragih, M. Zollhöfer, J. Hodgins, and C. Lassner, "Neuwigs: A neural dynamic model for volumetric hair capture and animation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 8641–8651. 2

[31] J. X. Zhang, J. Zhu, H. Chen, and S. Marschner, "Hairformer: Transformer-based dynamic neural hair simulation," *arXiv preprint arXiv:2507.12600*, 2025. 2

[32] T. Sun, G. Nam, C. Aliaga, C. Hery, and R. Ramamoorthi, "Human hair inverse rendering using multi-view photometric data," in *Eurographics Symposium on Rendering*, 2021. 3

[33] R. A. Rosu, S. Saito, Z. Wang, C. Wu, S. Behnke, and G. Nam, "Neural strands: Learning hair geometry and appearance from multi-view images," *ECCV*, 2022. 3

[34] G. Nam, C. Wu, M. H. Kim, and Y. Sheikh, "Strand-accurate multi-view hair capture," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3, 5

[35] S. Saito, L. Hu, C. Ma, H. Ibayashi, L. Luo, and H. Li, "3d hair synthesis using volumetric variational autoencoders," *ACM Trans. Graph.*, vol. 37, no. 6, Dec. 2018. 3

[36] D. Geng, C. Herrmann, J. Hur, F. Cole, S. Zhang, T. Pfaff, T. Lopez-Guevara, C. Doersch, Y. Aytar, M. Rubinstein, C. Sun, O. Wang, A. Owens, and D. Sun, "Motion prompting: Controlling video generation with motion trajectories," *arXiv preprint arXiv:2412.02700*, 2024. 3

[37] Z. Liu, H. Zhu, R. Chen, J. Francis, S. Hwang, J. Zhang, and J. Oh, "Mosaic: Generating consistent, privacy-preserving scenes from multiple depth views in multi-room environments," *arXiv preprint arXiv:2503.13816*, 2025. 3

[38] A. Razavi, A. Van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with vq-vae-2," *Advances in neural information processing systems*, vol. 32, 2019. 3

[39] A. Van Den Oord, O. Vinyals, *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017. 3, 4

[40] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2018. 4

[41] L. Hu, C. Ma, L. Luo, and H. Li, "Single-view hair modeling using a hairstyle database," *ACM Transactions on Graphics (Proceedings SIGGRAPH 2015)*, vol. 34, no. 4, July 2015. 5

[42] C. He, X. Sun, Z. Shu, F. Luan, S. Pirk, J. A. A. Herrera, D. L. Michels, T. Y. Wang, M. Zhang, H. Rushmeier, and Y. Zhou, "Perm: A parametric representation for multi-style 3d hair modeling," in *International Conference on Learning Representations*, 2025. 5

[43] C. H. Yi Zhou, Xin Sun, "Hair20k: A large 3d hairstyle database for hair modeling," 2024. [Online]. Available: https://zhouyisjtu.github.io/project_hair/hair20k.html 5

[44] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018. 5

[45] U. Yoo, J. Francis, J. Oh, and J. Ichnowski, "Kinesoft: Learning proprioceptive manipulation policies with soft robot hands," in *Proceedings of the 9th Conference on Robot Learning (CoRL)*, 2025. 7