HairStep: Transfer Synthetic to Real Using Strand and Depth Maps for Single-View 3D Hair Modeling

Y. Zheng, Z. Jin, M. Li, H. Huang, C. Ma, S. Cui, and X. Han

Computer Vision and Pattern Recognition, 2023

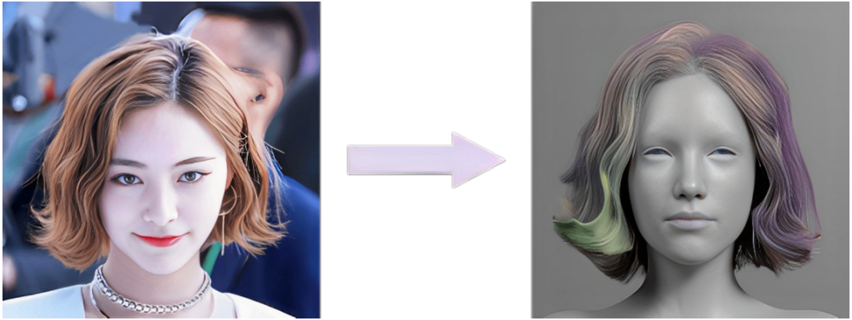
Alireza Heidari

December 2024

Outline

Objective

3D Hair Reconstruction from a Single Image

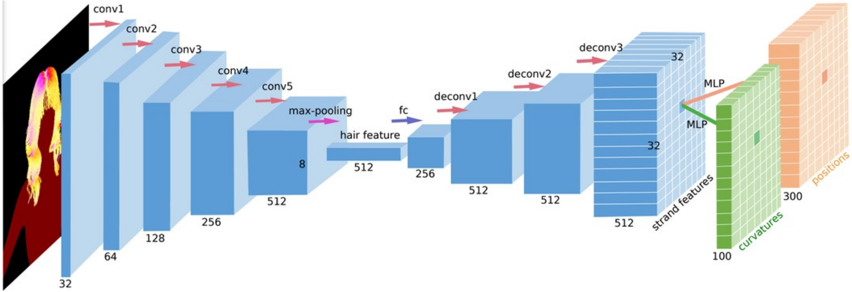


Single-view 3D Hair Modeling

Related Works

Related Works

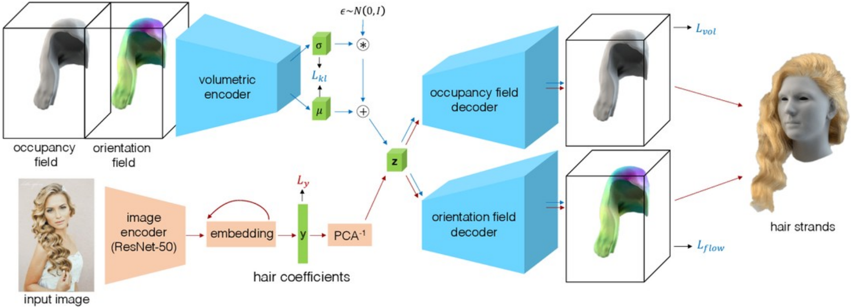
Single-View Hair Reconstruction using Convolutional Neural Networks



Zhou et al. 2018[1]

Related Works

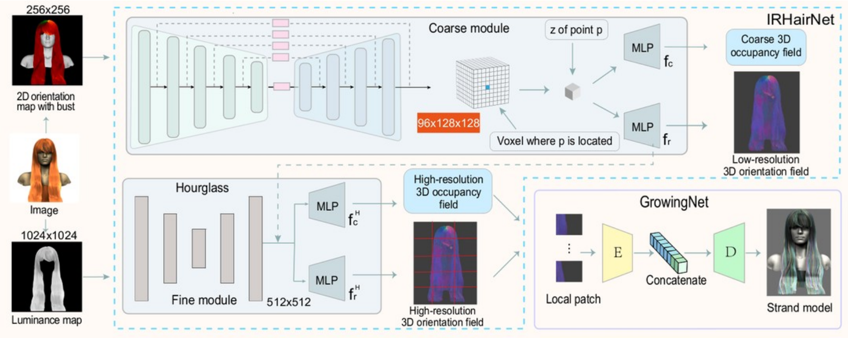
3D Hair Synthesis using Volumetric Variational Autoencoders



Saito et al. 2018[2]

Related Works

NeuralHDHair: Automatic High-fidelity Hair Modeling from a Single Image Using Implicit Neural Representations



Wu et al. 2022[3]

Motivation

Motivation

Main Challenge: Using synthetic data as a prior for real-world 3D hair modeling introduces a domain gap.

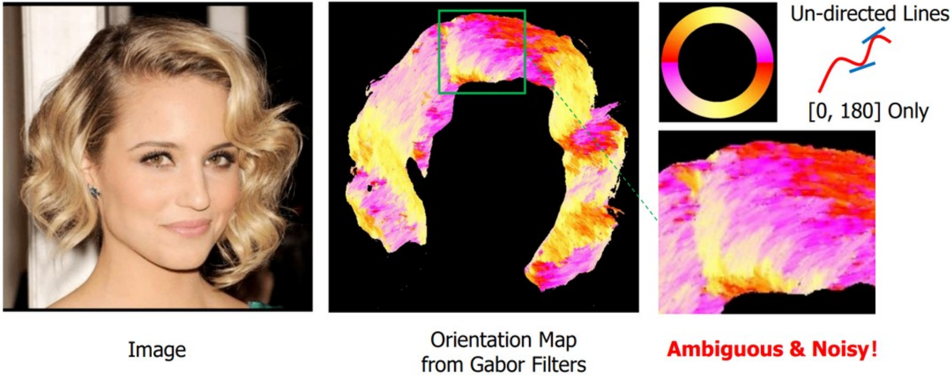
Existing Solutions:

- Utilize undirected 2D orientation maps as an intermediate representation between the input image and the 3D hair model.

Limitations:

- Ambiguous directionality: Loses 3D cues from the image.
- Reliance on image filters: Adds noise and inaccuracies.

Undirected Orientation Maps Challenges



Example of a 2D orientation map used in existing solutions.

Contribution

Contribution

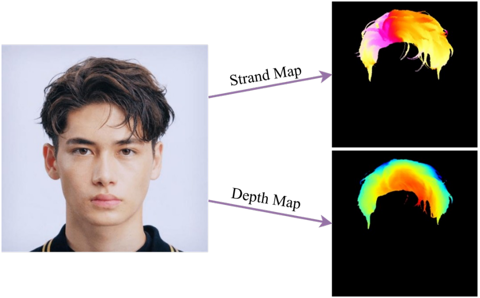
- Proposed *HairStep*, a novel intermediate representation combining strand maps and depth maps for 3D hair reconstruction.
- Developed a weakly-supervised domain adaptation method for depth estimation using synthetic priors and real-world sparse annotations.
- Created *HiSa* (strand annotations) and *HiDa* (relative depth annotations) datasets for 1,250 real portrait images.
- Introduced new metrics, *HairSale* (strand alignment error) and *HairRida* (relative depth accuracy), for quantitative evaluation of 3D hair modeling.
- Achieved state-of-the-art performance in single-view 3D hair modeling.

Proposed Method

HairStep Representation

HairStep is defined as $\mathbf{H} = \{\mathbf{O}, \mathbf{D}\}$ for each input image $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$, where:

- $\mathbf{O} \in \mathbb{R}^{W \times H \times 3}$ is the **Strand Map**.
- $\mathbf{D} \in \mathbb{R}^{W \times H \times 1}$ is the **Depth Map**.



Example of the *HairStep* representation.

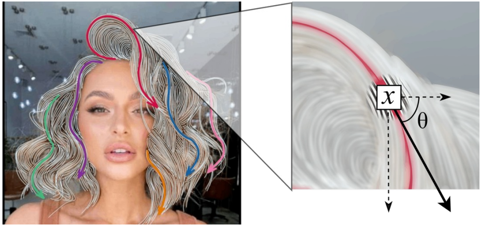
Strand Map Definition

The **Strand Map** $\mathbf{O} \in \mathbb{R}^{W \times H \times 3}$ is defined at each pixel x as:

$$\mathbf{O}(x) = \left(\mathbf{M}(x),\ \frac{\mathbf{O}_{\mathrm{2D}}(x)}{2} + 0.5\right), \tag{1}$$

where:

- $\mathbf{M}(x) \in \{0, 1\}$ is the hair mask indicating hair regions (1) and background (0).
- $\mathbf{O}_{\mathrm{2D}}(x) \in \mathbb{R}^2$ is the unit vector of 2D hair-growth orientation at pixel x.



Visualization of $\mathbf{O}_{\mathrm{2D}}(x)$, where $\mathbf{O}_{\mathrm{2D}}(x) = \begin{bmatrix} \cos(\theta) \\ -\sin(\theta) \end{bmatrix}$.

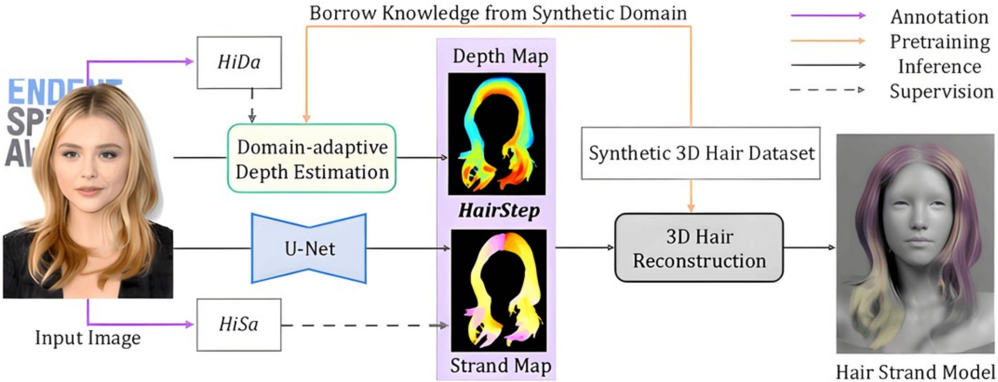
Depth Map Definition

The **Depth Map D** $\in \mathbb{R}^{W \times H \times 1}$ defines relative depth differences among hair strands.

- Each pixel $\mathbf{D}(x) \in [0, 1]$:
  - $\mathbf{D}(x) = 0$: Farthest from the camera (background or distant strands).
  - $\mathbf{D}(x) = 1$: Closest to the camera.

Example of the depth map $\mathbf{D}(x)$.

Method Overview



Pipeline of single-view 3D hair reconstruction using *HairStep*.

Method Overview

The pipeline consists of three main components:

1. **Strand Map Extraction and Prediction**
   - Extract strand maps from real images using the *HiSa* dataset.
   - Train a network to predict strand maps from input images.
2. **Domain-Adaptive Depth Estimation**
   - Estimate relative depth from real images using the *HiDa* dataset.
   - Employ domain adaptation techniques to refine depth estimation.
3. **3D Hair Reconstruction**
   - Reconstruct 3D hair strands from the predicted strand and depth maps.
   - Utilize implicit fields for volumetric hair representation.

Strand Map Extraction

Extracting strand maps is crucial for learning-based 3D hair modeling.

Approaches for Strand Map Extraction:

- **Synthetic Data:** Use rendering techniques (e.g., Soft Rasterizer [4]).
- **Real Data:** Use a U-Net architecture trained on the *HiSa* dataset.

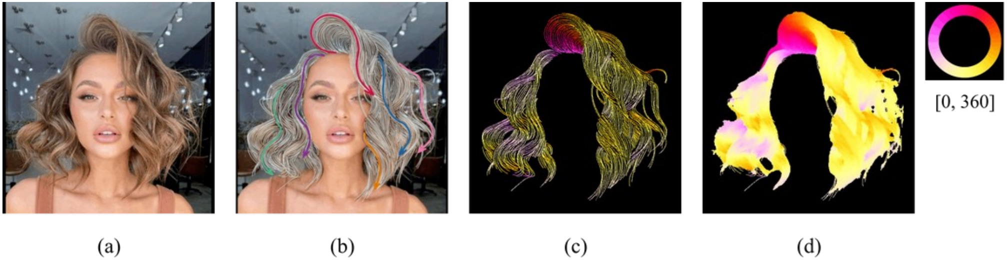
HiSa Dataset

The *HiSa* dataset provides strand maps for real images.

Dataset Details:

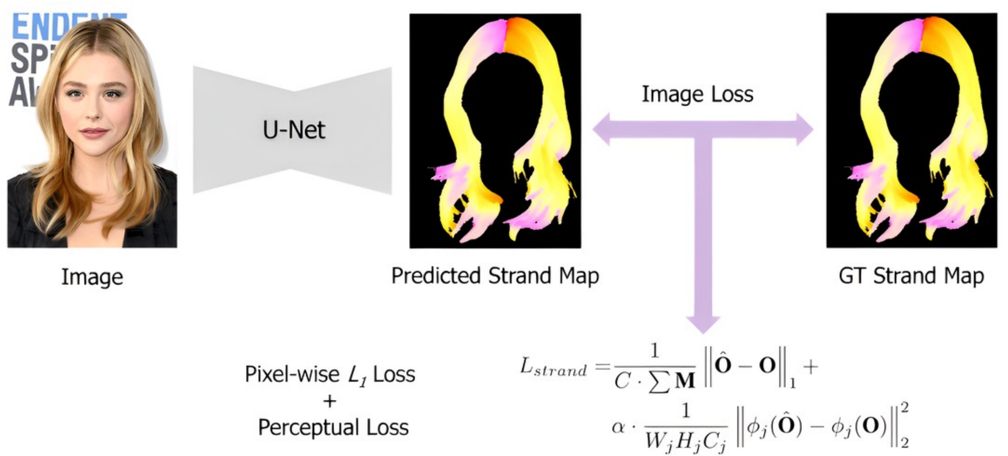
- **Collection:** 1,250 high-resolution portrait images.
- **Annotation Process:**
  - Professional artists draw directional vector curves from hair roots to ends.
  - Vector strokes are colored according to Eq. 1.
  - Colored strokes are interpolated to form dense strand maps.
- **Statistics:** On average, 300 strokes per portrait are annotated.

HiSa Dataset Visualization



Strand map extraction steps: (a) Portrait image, (b) Annotated vector strokes, (c) Colored strokes, (d) Final strand map.

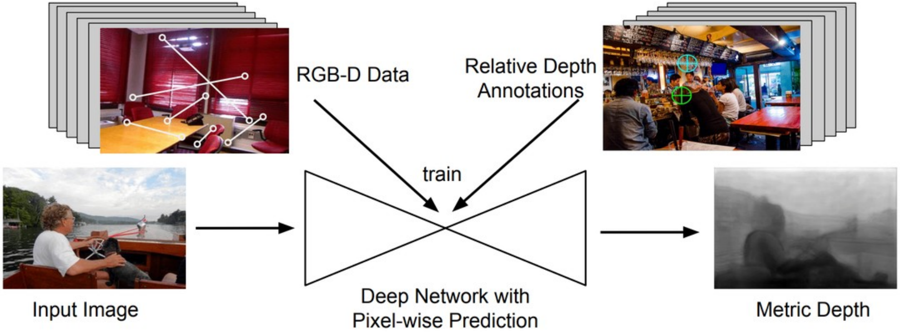
Strand Map Prediction Pipeline



Pipeline for strand map prediction using a U-Net architecture.

Relative Depth Estimation

Inspired by the depth-in-the-wild approach [5], relative depth estimation serves as a weak supervision signal.



Overview of the depth-in-the-wild pipeline.

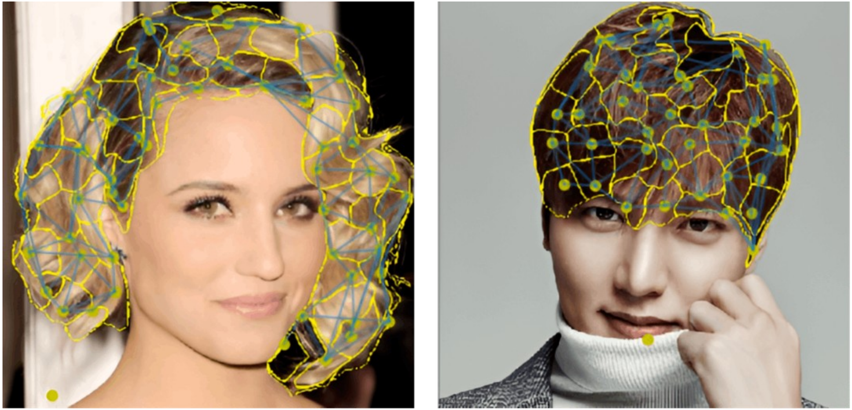
HiDa Dataset

The *HiDa* dataset provides relative depth annotations for hair regions in real images.

Dataset Details:

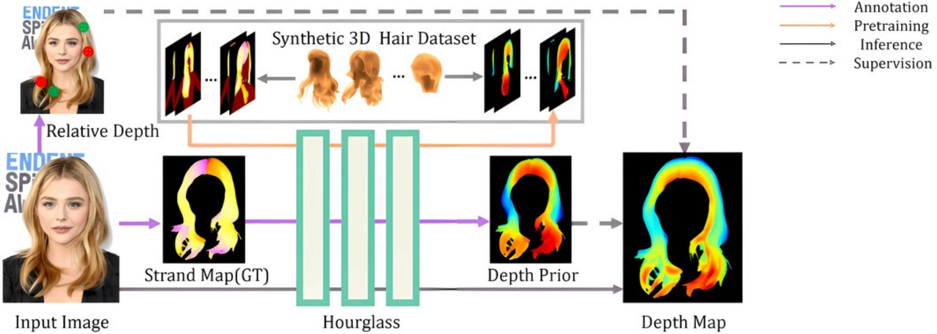
- **Collection:** 1,250 portrait images (same as *HiSa*).
- **Annotation Process:**
  - Generate super-pixels within the hair region.
  - Sample pixel pairs from adjacent super-pixels.
  - Present each pair to annotators to label which point is closer to the camera.
- **Statistics:**
  - On average, 140 pairs per portrait.
  - Total of 129,079 annotated pixel pairs.

HiDa Dataset Visualization



Example of super-pixels generated for the *HiDa* dataset.

Domain-Adaptive Depth Estimation Pipeline



$$L_{rank} = \frac{1}{N}\sum_{i=1}^{N} \max(0, -(\mathbf{D}_r(p_1^i) - \mathbf{D}_r(p_2^i)) \cdot r^i + \varepsilon) \qquad L_{depth} = \beta \cdot \left\| \mathbf{D}_r - \bar{\mathbf{D}} \right\|_1 + L_{rank}$$

Overview of the domain-adaptive depth estimation approach.

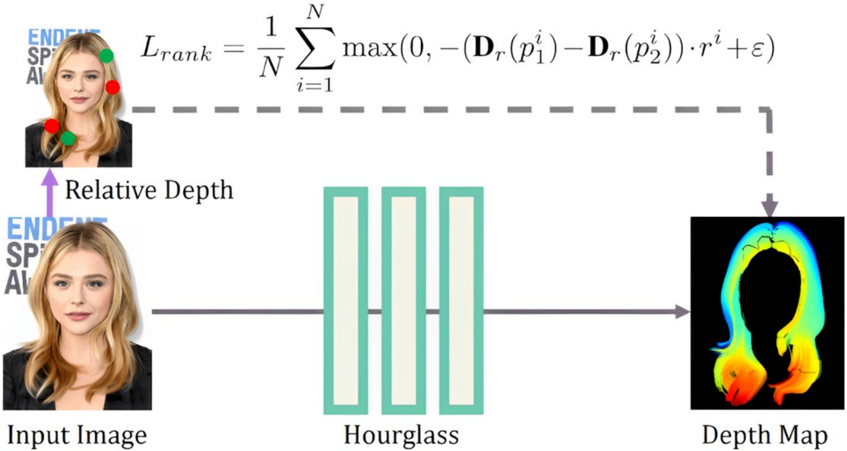
Depth Estimation Methodology

An **Hourglass Network** is used to predict depth maps from input images.

- **Relative Depth Supervision:** Use margin-based ranking loss with *HiDa* depth pairs.
- **Domain Adaptation:** Enhance depth prediction using synthetic data.
- **Loss Function:** Combine L_1 loss and ranking loss:

$$L_{\text{depth}} = \beta \|\mathbf{D}_r - \bar{\mathbf{D}}\|_1 + L_{\text{rank}}. \tag{2}$$

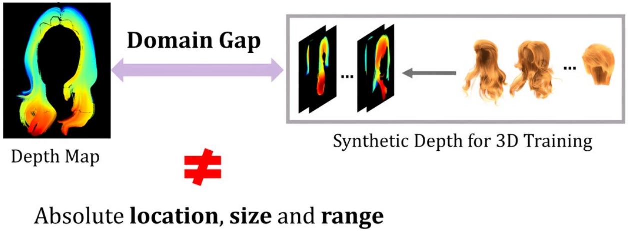
Depth Estimation Methodology



Hourglass network architecture for depth estimation.

Challenges in Depth Estimation

Training with only ordinal labels can introduce ambiguity and artifacts in depth prediction, resulting in noisy or coarse 3D hair models.



Domain gaps and artifacts in predicted depth from ordinal labels.

Domain-Adaptive Depth Estimation

To mitigate artifacts, domain-adaptive depth estimation pipeline is proposed.

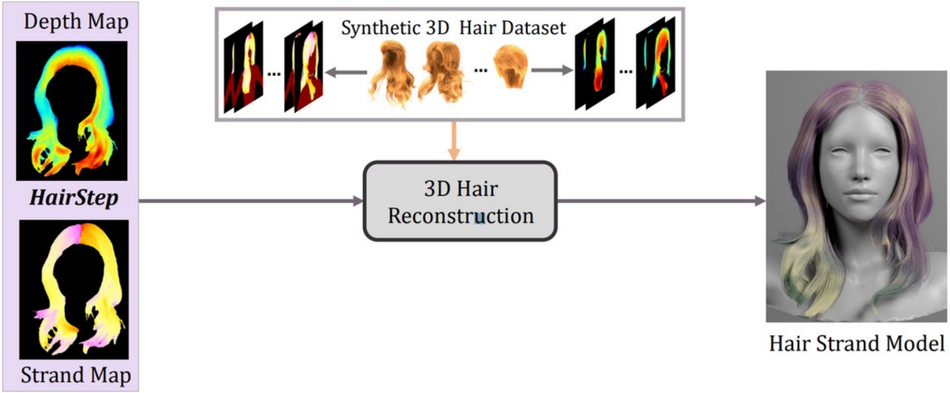
Step 1: Pre-training on Synthetic Data

- **Network:** Hourglass network (`Depth_syn`).
- **Input:** Ground-truth strand maps from synthetic data.
- **Output:** Depth map $\bar{\mathbf{D}}$.
- **Loss:** L_1 loss on $\bar{\mathbf{D}}$.

Step 2: Training on Real Data

- **Network:** Hourglass network (`Depth_r`).
- **Input:** Real images with predicted strand maps.
- **Output:** Depth map $\mathbf{D}_r$.
- **Supervision:** Depth prior $\bar{\mathbf{D}}$ from `Depth_syn`.
- **Loss:** Combined L_{depth} as in the loss function.

Single-View 3D Hair Modeling Pipeline



Pipeline for single-view 3D hair modeling using *HairStep*.

Modeling Details

Objective: Reconstruct strand-level 3D hair from a single-view portrait image using the *HairStep* representation $\{\mathbf{O}, \mathbf{D}\}$.

Key Components:

1. **Implicit Representation**
   - Predict occupancy and orientation fields in a canonical head space.
   - Utilize a neural network to model volumetric hair structure.
2. **Strand Generation**
   - Convert implicit fields into explicit 3D hair strands.
   - Follow orientation field vectors to grow strands from scalp.

Implicit 3D Hair Representation

Why Use Implicit Fields?

- Efficiently represent complex volumetric structures.
- Capture continuous geometry without discretizing every strand.

Definitions (Following NeuralHDHair [3]):

- **Occupancy Field** $f_{\text{occ}}(\mathbf{x}) \in [0, 1]$
  - Indicates whether point **x** is inside the hair volume.
- **Orientation Field** $f_{\text{orient}}(\mathbf{x}) \in \mathbb{R}^3$
  - Provides the local hair-growth direction at point **x**.

Neural Network Prediction (NeuralHDHair*)

Adapted NeuralHDHair Framework:

Input:

- Strand map $\mathbf{O}$ (from *HairStep*).
- Depth map $\mathbf{D}$ (from domain-adaptive depth estimation).

Output:

- Implicit occupancy field $f_{\text{occ}}(\mathbf{x})$.
- Implicit orientation field $f_{\text{orient}}(\mathbf{x})$.

Modifications:

- *No Luminance Map*: Exclude luminance to reduce domain gap from lighting variations.
- *Omit GrowingNet*: Focus on direct strand generation from implicit fields.

Hair Strand Generation Process

After predicting the implicit fields, hair strands are generated from the scalp.

1. **Initialization**
   - Place hair roots uniformly on a standard scalp model.
2. **Strand Growing**
   - From each root, iteratively follow orientation vectors from $f_{\text{orient}}(\mathbf{x})$.
   - Continue until $f_{\text{occ}}(\mathbf{x}) = 0$ or reaching maximum strand length.
3. **Result**
   - Obtain a dense set of 3D hair strands (approximately 10,000 strands) that replicate the input hairstyle.

Experiments

Experimental Objectives

Goal: Validate that *HairStep* reduces the domain gap between synthetic and real data, thereby improving single-view 3D hair modeling quality.

Key Objectives to Validate:

- Demonstrate that *HairStep* outperforms traditional orientation-based methods.
- Show improved 3D hair reconstruction on synthetic and real images.
- Introduce and validate new metrics (*HairSale* and *HairRida*) for objective evaluation.

Overview of Core Experiments

1. **HairStep Extraction**
   - Compare *HairStep*'s strand maps with orientation maps generated using Gabor filters.
   - Evaluate depth maps extracted using different strategies within the *HairStep* framework.
2. **Single-View 3D Hair Modeling**
   - Compare the quality of 3D hair reconstruction with existing orientation-based methods.
3. **Intermediate Representation Evaluation**
   - Assess the impact of different intermediate representations on 3D reconstruction quality.
   - Evaluate performance on both synthetic and real data using various methods.
4. **Ablation Study**
   - Experiment with different depth estimation strategies within the *HairStep* framework.
   - Evaluate their impact on the final 3D hair reconstruction quality.

Datasets

Synthetic Dataset: *USC-HairSalon* [6]

- Contains 343 3D hair models, each with multiple camera views.

Real Datasets: *HiSa* (Strand Map Annotation) and *HiDa* (Depth Annotation)

- *HiSa*: 1,250 real portrait images with dense, pixel-level strand direction annotations.
- *HiDa*: 1,250 real portrait images with carefully annotated relative depth pairs.

Evaluation Metrics

Quantitative Evaluation:

- **For Real Data:**
  - *HairSale*: Mean angular error of predicted strand directions.
  - *HairRida*: Accuracy of predicted relative depth orderings.
- **For Synthetic Data:**
  - *Orientation Error*: L_2 error between predicted and ground-truth 3D orientation fields.
  - *Occupancy Accuracy*: Precision of predicted 3D occupancy fields relative to ground truth.

Qualitative Evaluation:

- *Visual Quality*: Subjective assessment of reconstructed geometry.
- *User Study*: Preferences gathered from human subjects comparing reconstruction results.

Evaluation Metric: *HairSale*

HairSale: Mean Angular Error of Strand Directions.

$$\text{HairSale} = \frac{1}{K} \sum_{i=1}^{K} \arccos\big(V(O_r(x_i)) \cdot V(O_{gt}(x_i))\big)$$

- **Notation:**
  - K: Number of pixels within the overlap of predicted and ground-truth masks.
  - $V(O_r(x_i))$: Unit vector of predicted orientation at pixel x_i.
  - $V(O_{gt}(x_i))$: Unit vector of ground-truth orientation at pixel x_i.

Evaluation Metric: *HairRida*

HairRida: Relative Depth Ordering Accuracy.

$$\text{HairRida} = \frac{1}{Q}\sum_{i=1}^{Q} \max\big(0, r_i \cdot \text{sign}(D_r(p_{i1}) - D_r(p_{i2}))\big)$$

- **Notation:**
  - Q: Number of annotated pixel pairs.
  - r_i: Ground-truth relative depth order for pair i ($+1$ or -1).
  - $D_r(p_{i1})$, $D_r(p_{i2})$: Predicted depth values at the corresponding pixels.

Evaluation Metrics Illustration

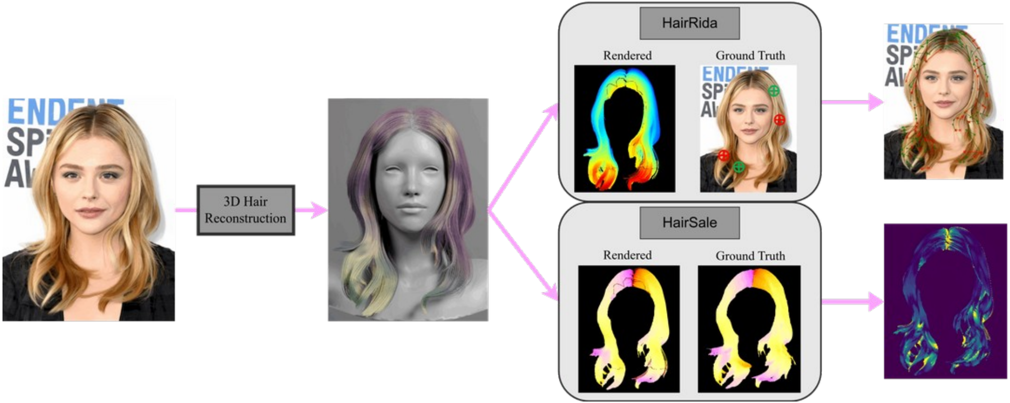


Illustration of the *HairSale* and *HairRida* metrics.

Qualitative Evaluation Examples

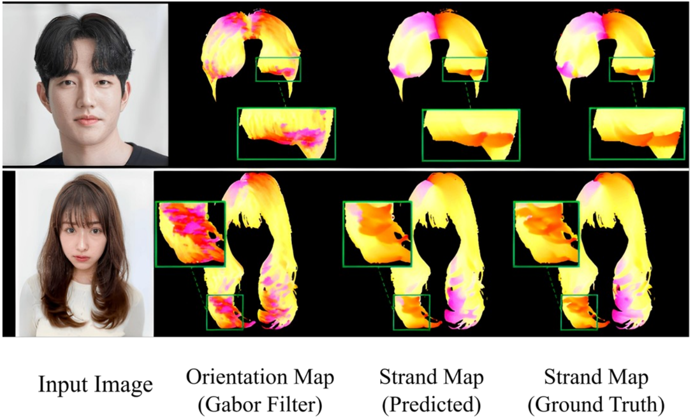
Comparing input images with reconstructed 3D hair models.

HairStep Extraction Evaluation

Objective: Evaluate the quality of strand and depth maps extracted using *HairStep*.

- **Strand Maps:**
  - Compare *HairStep*'s strand maps with orientation maps generated using Gabor filters.
  - Quantitative evaluation using the *HairSale* metric (undirected).
- **Depth Maps:**
  - Compare the impact of different depth estimation strategies within the *HairStep* framework.
  - Quantitative evaluation using the *HairRida* and L_1 loss metrics.

Strand Map Extraction: Qualitative Comparison



Qualitative comparison of predicted strand maps versus ground truth.

Strand Map Extraction: Quantitative Results

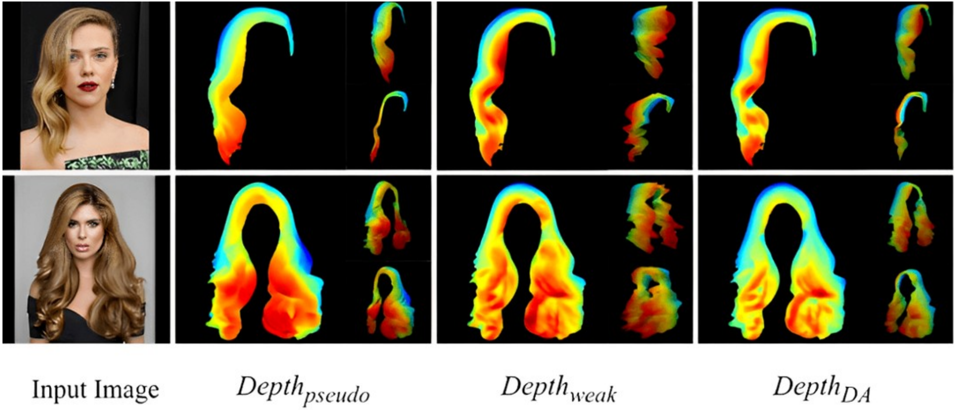
Using the *HairSale* metric for evaluation, *HairStep*'s strand map is converted to an undirected form to match Gabor filters' orientation ambiguity.

Method	*HairSale* $\downarrow$ (Undirected)
Gabor Filters	18.4
HairStep	**14.2**

Undirected *HairSale* comparison: *HairStep* improves over Gabor filters by 22.8%.

Note: Undirected evaluation introduces bi-directional ambiguity, increasing error. Despite this, *HairStep* still outperforms Gabor filters.

Depth Map Estimation: Qualitative Comparison



Qualitative comparison of predicted depth maps.

Depth Map Estimation: Quantitative Results

Methods:

- $Depth_{\text{pseudo}}$: Synthetic pseudo-label-based depth estimation.
- $Depth_{\text{weak}}$: Weakly supervised depth estimation using ordinal cues.
- $Depth_{\text{DA}}$: Domain-adaptive depth estimation.

Method	*HairRida* ↑	L_1 ↓
$Depth_{\text{pseudo}}$	80.47%	-
$Depth_{\text{weak}}$	85.17%	0.2470/3.125
$Depth_{\text{DA}}$	**85.20%**	**0.1768/0.1188**

Relative depth accuracy (*HairRida*) and L_1 loss comparisons for different depth estimation strategies.

Single-View 3D Hair Modeling Comparison

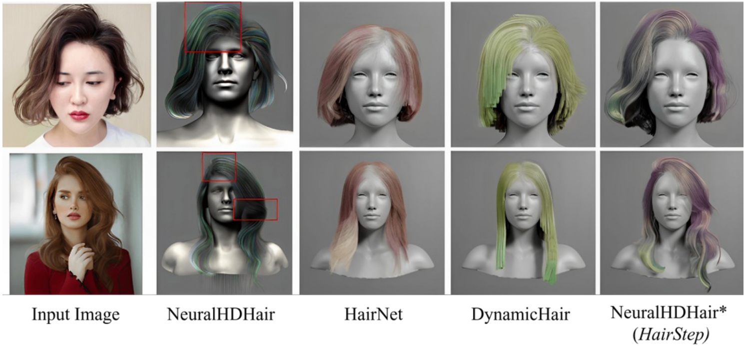
Methods Compared:

- HairNet [1]
- DynamicHair [7]
- NeuralHDHair [3]
- **NeuralHDHair* + HairStep**: Proposed method.

Modifications for NeuralHDHair*:

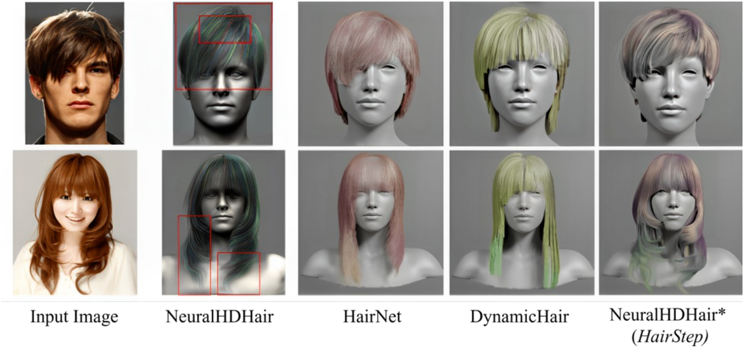
- *No luminance map*: Reduces domain gap from varying lighting conditions.
- *Omit GrowingNet*: Focus on reconstruction quality, not scalability.

Single-View 3D Hair Modeling: Qualitative Comparison



Qualitative comparison with existing single-view 3D hair reconstruction methods.

Single-View 3D Hair Modeling: Additional Comparisons



Additional qualitative comparisons with baseline methods.

Single-View 3D Hair Modeling: Observations

Key Observations:

- **HairNet** and **DynamicHair**:
  - Produce coarse, less detailed shapes.
  - Struggle with complex hairstyles.
- **NeuralHDHair**:
  - Faces challenges with sharp depth variations.
  - Struggles with intricate hair growth patterns.
- **Orientation-based methods**:
  - Lack fine-grained detail necessary for accurate 3D reconstruction.

Intermediate Representation Evaluation

Goal: Examine how different intermediate representations affect the final 3D hair reconstruction quality.

Intermediate Representations:

- Orientation Map
- Strand Map
- *HairStep* (Strand + Depth Maps)

Methods:

- HairNet [1]
- DynamicHair [7]
- NeuralHDHair* + *HairStep*

Quantitative Results on Synthetic Data

Method	Orientation Error ↓	Occupancy Accuracy ↑
HairNet (Orientation Map)	0.02349	–
HairNet (Strand Map)	0.02206 (-6.1%)	–
HairNet (HairStep)	**0.02184** (-7.0%)	–
DynamicHair (Orientation Map)	0.1352	78.19%
DynamicHair (Strand Map)	0.1185 (-12.4%)	79.62%
DynamicHair (HairStep)	**0.1174** (-13.2%)	**79.78%**
NeuralHDHair* (Orientation Map)	0.1324	82.59%
NeuralHDHair* (Strand Map)	0.0722 (-41.7%)	84.18%
NeuralHDHair* (HairStep)	**0.0658** (-50.3%)	**86.77%**

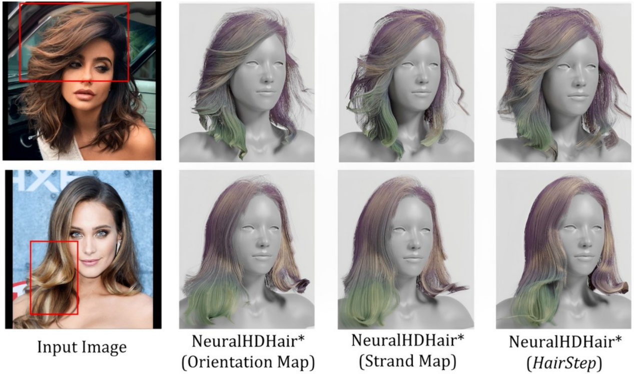
Quantitative results on synthetic data (USC-HairSalon). *HairStep* consistently improves performance.

Quantitative Results on Real Data

Method	IoU ↑	*HairSale* ↓	*HairRida* ↑
HairNet (Orientation Map)	57.15%	31.97	75.65%
HairNet (Strand Map)	57.48%	28.60 (-10.5%)	74.81%
HairNet (HairStep)	57.01%	**27.68** (-13.4%)	**74.97%**
DynamicHair (Orientation Map)	56.39%	32.66	74.08%
DynamicHair (Strand Map)	59.51%	**26.53** (-18.8%)	73.42%
DynamicHair (HairStep)	59.14%	27.51 (-15.8%)	**73.58%**
NeuralHDHair* (Orientation Map)	77.56%	19.60	70.67%
NeuralHDHair* (Strand Map)	77.60%	**16.00** (-18.4%)	72.37%
NeuralHDHair* (HairStep)	77.22%	16.36 (-16.5%)	**76.79%**

Quantitative comparisons on real data. *HairStep* yields consistently improved performance.

Qualitative Comparison on Real Data



Qualitative results of NeuralHDHair* using (left to right) orientation map, strand map, and *HairStep*.

Intermediate Representation Impact: Observations

HairStep consistently improves performance on both synthetic and real data.

User Study (10 samples, 39 participants):

- 64.87% preferred reconstructions using *HairStep*.
- 21.28% preferred strand map-based reconstructions.
- 13.85% preferred orientation map-based reconstructions.

Ablation Study: Depth Estimation

Objective: Evaluate the impact of different depth estimation approaches on final reconstruction.

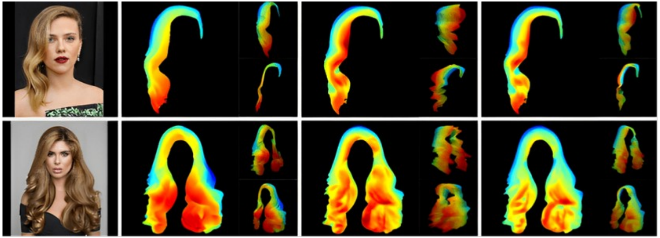
Configurations:

- C_0: Strand Map + $Depth_{\text{pseudo}}$
- C_1: Strand Map + $Depth_{\text{weak}}$
- **Full**: Strand Map + $Depth_{\text{DA}}$

Method	IoU ↑	*HairSale* ↓	*HairRida* ↑
$\mathbf{C_0}$	77.75%	16.03 (-18.2%)	73.57%
$\mathbf{C_1}$	77.11%	16.54 (-15.6%)	75.80%
Full	**77.22%**	**16.36** (-16.5%)	**76.79%**

Ablation study: Evaluating depth estimation methods within the *HairStep* framework.

Ablation Study: Configurations Comparison



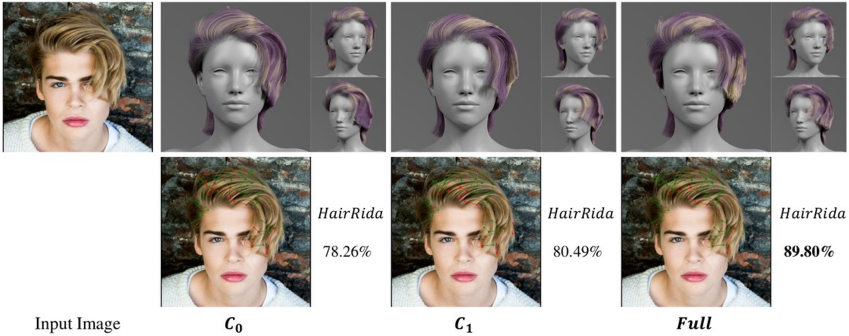
Input Image $Depth_{pseudo}$ $Depth_{weak}$ $Depth_{DA}$

C_1 gives **worse alignment of hair growth**.

C_0 suffers from the **flat geometry of depth**.

Qualitative comparison of different depth estimation configurations.

Ablation Study: Visual Comparison



From left to right: Input image, C_0, C_1, and Full method. Green/red lines indicate correct/incorrect relative depth predictions (*HairRida*).

Conclusion

Conclusion

Contributions:

- Proposed *HairStep*, a novel intermediate representation combining strand and depth maps.
- Collected new datasets *HiSa* and *HiDa* with annotated real images.
- Proposed two quantitative metrics: *HairSale* and *HairRida*.
- Achieved state-of-the-art performance in single-view 3D hair modeling.

Limitations:

- Manual annotation is time-consuming, limiting scalability.
- Generalization to unseen hairstyles and diverse real-world conditions needs further investigation.

References I

[1] Y. Zhou, L. Hu, J. Xing, *et al.*, "Single-view hair reconstruction using convolutional neural networks," in *European Conference on Computer Vision*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:49666680.

[2] S. Saito, L. Hu, C. Ma, H. Ibayashi, L. Luo, and H. Li, "3d hair synthesis using volumetric variational autoencoders," *ACM Transactions on Graphics (TOG)*, vol. 37, pp. 1–12, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:54101192.

[3] K. Wu, Y. Ye, L. Yang, H. Fu, K. Zhou, and Y. Zheng, "Neuralhdhair: Automatic high-fidelity hair modeling from a single image using implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1526–1535.

[4] S. Liu, T. Li, W. Chen, and H. Li, "Soft rasterizer: A differentiable renderer for image-based 3d reasoning," *The IEEE International Conference on Computer Vision (ICCV)*, Oct. 2019.

References II

[5] W. Chen, Z. Fu, D. Yang, and J. Deng, "Single-image depth perception in the wild," *Advances in neural information processing systems*, vol. 29, 2016.

[6] L. Hu, C. Ma, L. Luo, and H. Li, "Single-view hair modeling using a hairstyle database," *ACM Transactions on Graphics (TOG)*, vol. 34, pp. 1–9, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:18205814.

[7] L. Yang, Z. Shi, Y. Zheng, and K. Zhou, "Dynamic hair modeling from monocular videos using deep neural networks," *ACM Transactions on Graphics (TOG)*, vol. 38, pp. 1–12, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:207997743.

[8] Y. Zheng, Z. Jin, M. Li, *et al.*, "Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 12 726–12 735.