

HairStep: 使用 Strand maps 和 Depth maps 结构, 使用合成头发数据和真实头发数据的组合, 实现单视角 3D 的头发模型的创建

郑玉健¹⁾ 金子荣²⁾ 李莫兰³⁾ 黄海滨³⁾ 马崇阳³⁾ 崔曙光²⁾¹⁾ 韩晓光²⁾¹⁾

¹⁾未来智联网研究院 (FNii), 香港中文大学 (CUHKSZ)

²⁾上海证券研究所 (SEE) 香港中文大学 (CUHKSZ) ³⁾快手 (Kuaishou) 香港中文大学 (CUHKSZ)

摘要 在这项工作中, 我们解决了基于学习的单视图三维头发建模的挑战性问题。由于收集成对的真实现像和三维头发数据的困难, 利用合成数据为真实领域提供先验知识成为领先的解决方案。不幸的是, 这就引入了领域差距的挑战, 由于真实头发渲染的固有困难, 现有的方法通常使用方向图而不是头发图像作为输入来弥补差距。我们坚定地认为一个中间表示是必要的, 但我们认为, 使用主要的基于滤波的方法的方向映射对不确定的误差很敏感, 而且远不是一个合格的表示。因此, 我们首先提出了这个问题, 并提出了一种新的中间表示方法, 称为 HairStep, 它由链图和深度图组成。结果表明, 发步不仅为精确的三维头发建模提供了足够的信息, 而且可以从真实现像中进行推断出来。具体来说, 我们收集了一个包含 1250 幅肖像图像的数据集。进一步设计了一个学习框架, 将真实现像转移到链图和深度图上。实验表明, 发步缩小了合成和真实之间的域差距, 并在单视角三维头发重建上取得了最新的性能。

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

Yujian Zheng¹⁾ Zirong Jin²⁾ Moran Li³⁾ Haibin Huang³⁾ Chongyang Ma³⁾

Shuguang Cui²⁾¹⁾ Xiaoguang Hang²⁾¹⁾

¹⁾(FNii, CUHKSZ) ²⁾(SSE, CUHKSZ) ³⁾(Kuaishou, CUHKSZ)

Abstract In this work, we tackle the challenging problem of learning-based single-view 3D hair modeling. Due to the great difficulty of collecting paired real image and 3D hair data, using synthetic data to provide prior knowledge for real domain becomes a leading solution. This unfortunately introduces the challenge of domain gap. Due to the inherent difficulty of realistic hair rendering, existing methods typically use orientation maps instead of hair images as input to bridge the gap. We firmly think an intermediate representation is essential, but we argue that orientation map using the dominant filtering-based methods is sensitive to uncertain noise and far from a competent representation. Thus, we first raise this issue up and propose a novel intermediate representation, termed as HairStep, which consists of a strand map and a depth map. It is found that HairStep not only provides sufficient information for accurate 3D hair modeling, but also is feasible to be inferred from real images. Specifically, we collect a dataset of 1,250 portrait images with two types of annotations. A learning framework is further designed to transfer real images to the strand map and depth map. It is noted that, an extra bonus of our new dataset is the first quantitative metric for 3D hair modeling. Our experiments show that HairStep narrows the domain gap between synthetic and real and achieves state-of-the-art performance on single-view 3D hair reconstruction.

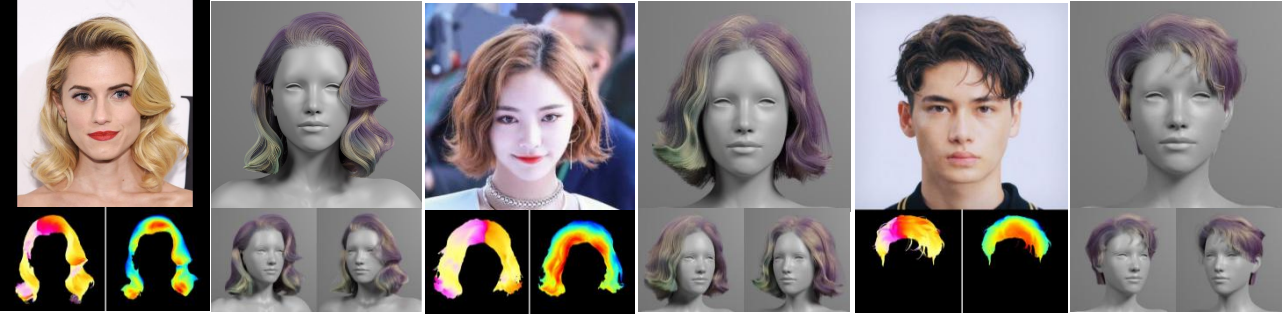


Figure 1. 给定一幅肖像图像，我们首先将其转换为一个中间表示的链图和一个深度图（如图左下角和右下角所示），然后在链级恢复相应的 3D 头发模型。我们的 HairStep 能够弥补合成数据和真实数据之间的领域差距，并实现高保真的头发建模结果。

1 介绍

高保真的 3D 头发建模是数字人类创造的关键组成部分。一个人的发型通常由大约 10 万股[1]组成。由于复杂性，获得高质量的 3D 头发模型成本昂贵。虽然高端捕获系统[9, 19]相对成熟，但仍难以重建具有复杂几何形状的令人满意的 3D 头发。

柴梦蕾等人[3, 4]首先提出了来自单视图图像的简单头发建模方法，使获取 3D 头发的用户更加友好。但是这些早期的系统需要额外的输入，比如用户的敲击。此外，它们只对头发的可见部分有效，而不能忠实地恢复不可见的几何形状。近年来，基于检索的[2, 10]方法减少了用户输入的依赖性，提高了重建的三维头发模型的质量。然而，这些方法的准确性和效率直接受到三维头发数据库的大小和多样性的影响。受学习型形状侦察技术进步的启发。三维发型模型由神经网络生成，如显式点序列[46]、体积定向场[26, 30, 41]和隐式取向场[37]从单视图输入。随着上述三维头发表示的发展，恢复形状的质量得到了显著的提高。由于对 3D 头发和真实图像进行填充具有挑战性，现有的基于学习的方法[26, 30, 37, 40, 46]在应用于真实肖像之前，只是在合成数据上进行训练。然而，渲染图像（来自合成头发模型）和真实图像之间的域差距对重建结果的质量有很大的负面影响。通过这些方法恢复的三维发型通常在一些重要的细节（如方向、卷曲和遮挡）上与给定的图像不匹配。

为了缩小合成数据与真实图像之间的域差距，现有的方法[37, 38, 41, 46]大多采用二维方向图[23]作为输入图像与三维头发模型之间的中间表示。然而，这种无向的二维方向图在增长方向上是

模糊的，并且失去了图像中给出的三维提示。更重要的是，它依赖于图像滤波器，这导致了有误差的方向图。在这项工作中，我们重新考虑了当前单视角三维头发建模的问题，并认为有必要找到一个更合适的中间表示来弥合真实数据和合成数据之间的域差距。这种表示应该为 3D 头发重建提供足够的信息。此外，它应该是域不变的，并且可以很容易地从真实图像中得到。

为了解决上述问题，我们提出了 HairStep，一种具有发型感知和深度增强的单视角三维头发建模的混合表示。基于如何从真实图像中生成干净的方向图，我们通过沿着头发绘制对齐良好的密集二维向量曲线，我们为真实图形绘制发型图（即定向二维方向图）。有了这种帮助，我们就可以直接从输入的单视图图像中预测定向和干净的二维方向图。我们还需要一个额外的中间表示组件来为头发重建提供三维信息。受野外深度[5]的启发，我们为真实肖像的头发区域标注了相对深度信息。但是从稀疏和序数绘制中学习到的深度与合成深度有不可忽视的域差距。为了解决这个问题，我们提出了一种基于合成领域的先验知识的弱监督领域自适应解决方案。一旦我们得到了发型图和深度图，我们就把它们组合在一起形成 HairStep。然后将该混合表示输入一个网络，隐式学习三维头发模型的三维方向场和三维占用场。最后，可以从这两个领域合成三维链模型。高保真度结果如 Fig.1 所示。为了方便起见，我们将带“发型解释”的头发图像数据集命名为 HiSa，带有“深度解释”的数据集命名为 HiDa。

以往的方法主要是通过比较重建的三维头发的视觉质量和准备充分的用户研究来对真实输入进行评价。这种主观的测量可能会导致不公平的评价和有偏见的结论。NeuralHdHair [37]投影重建的三维

发型的生长方向，并与从真实图像中过滤后的二维方向图进行比较。这是一个值得注意的进展，但提取的方向图是有误差的和不准的。只评估了二维增长方向，而忽略三维信息。基于我们的解释，我们提出了新颖的和客观的指标来评价在真实图像上的单视图三维头发建模。我们渲染恢复三维头发模型以获得发型图和深度图，然后将它们与我们的基于真相得到的解释进行比较。在我们的真实数据集和合成的三维头发数据集 USC-HairSalon [10] 上进行的大量实验证明了我们的最新表示的优越性。

我们的工作的主要贡献如下：

- 我们首先重新考虑了单视图三维头发建模中合成数据和真实数据之间的显著领域差距问题，并提出了一种新的 HairStep。在此基础上，我们提供了一个全自动的单视角发型重建系统，实现了最先进的性能。
- 我们提供了两个数据集，即 HiSa 和 HiDa，用于解释 1250 个真实肖像图像的发型图和深度图。这为未来关于头发的理解、重建和编辑的研究打开了一扇大门。
- 我们仔细设计了一个框架，从真实的图像生成 HairStep。更重要的是，我们提出了一种弱监督域自适应的头发深度估计的解决方案。
- 基于我们的解释，我们引入了新颖的和公平的指标来评估单视图三维头发建模方法在真实图像上的性能。

2 相关工作

单视图 3D 头发建模。从单视图输入重建三维头发仍然是计算机视觉和图形学中一个开放的问题。与多视图头发建模 [19, 21, 40] 相比，单视图方法 [4, 10, 37, 46] 更有效和实用，因为多视图方法需要仔细调节的环境和复杂的硬件设置。开创性的基于单视图的方法 [2-4, 10] 通常首先基于数据库生成一个粗糙的头发模型，然后使用几何优化来近似目标发型。这些方法的有效性依赖于先验的质量，而对于具有挑战性的输入，性能不太令人满意。

近年来，随着深度学习的快速发展，人们提出了几种基于生成模型的 [26, 30, 37, 46] 方法。HairNet [46] 以方向图作为输入，缩小真实图像与合成数据之间的域差距，使网络能够使用大规模合成数据集对其进行训练。Hair-VAE [26] 采用变分自动编码器从单视图输入生成头发模型。Hair-GAN

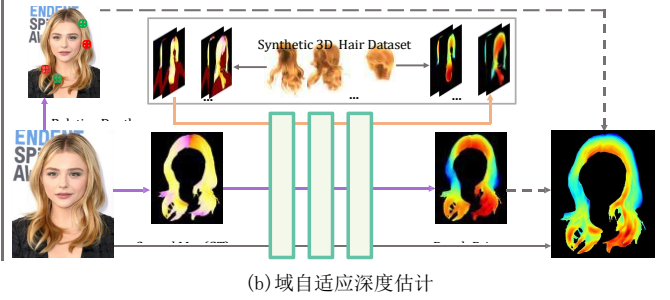
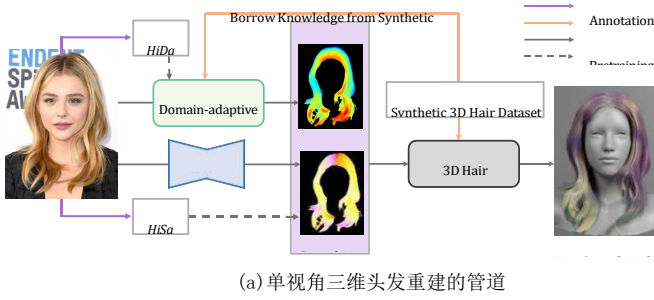
[41] 将基于 GAN 的方法引入到头发生成过程中。然而，用这些方法重建的头发模型往往过于粗糙和过度平滑，这主要是由于三维神经网络的能力有限。为了解决这个问题，NeuralHDHair [37] 提出了一种从粗到细的方式来获得高分辨率的 3D 定向场和占用场，使生长网能够生成像样的头发模型。

头发建模的方向图。由于头发发型固有的细长形状，使用二维方向图和/或三维方向场作为中间表示来指导建模过程是很直观的。现有的基于图像的头发建模方法通常将不同方向的 Gabor 滤波器应用于输入肖像，并计算局部二维方向，以遵循具有最大滤波响应 [23, 24] 的方向。然后将这些二维方向图转换为基于多视图校准信息 [9, 19, 20] 的三维方向场，或直接输入到神经网络中，作为预测三维目标发型 [37, 38, 41, 46] 的辅助输入。然而，基于图像滤波操作的二维方向图受到输入误差的影响，这可以通过额外的平滑或扩散过程来缓解，以降低精度 [19, 20] 为代价。更重要的是，这些二维方向图和三维方向场并不能从结构的角度区分毛根和尖端。解决这种方向性歧义需要额外的输入，如用户草图 [30] 和基于物理的示例 [9]，这可能很乏味，也可能不能很好地概括。一些针对二维头发图像生成的方法 [35] 也是基于方向图的。

深度图估计。许多使用先进技术的数据驱动方法 [8, 12, 17, 29] 在深度估计方面取得了令人信服的性能。然而，这些方法依赖于密集的深度标记 [6, 13, 14, 31]，这是发型图无法达到的。陈等人 [5] 通过解释稀疏点对之间的相对深度来消除密集深度标记，以帮助估计自然环境下的深度图。然而，目前还没有专门的工作来估计头发发线的深度图。大多数三维人脸或身体重建方法 [28, 34, 36] 只产生一个粗糙的头发区域深度图，这远远不足以实现高保真的头发建模。

3 HairStep 表示

通过基于学习的技术从单个图像中恢复三位头发的理想方法是训练一个网络，它可以将真实的图像变成基于真实图像的三维发型图。但是，真实的头发图像 [46] 获得基于真实图像得到的三维头发几何图形是困难和昂贵的。[26] 只能利用基于检索的方法 [10] 来创建伪三维头发模型。经过这些数据训练的网络不能生成与给定图像对齐的三维发型，因



为很难保证检索到的头发与输入图像对齐。由于真实头发渲染固有的困难，现有的[37, 38, 41, 46]方法以方向图代替头发图像作为输入，以缩小真实数据和合成数据之间的域差距。然而，由图像滤波器得到的方向图存在不确定的误差，远不是一个胜任的中间表示。因此，需要一个更好的方案来弥补这一显著的差距。

我们现在正式介绍了我们的新表示 HairStep 的单视图三维头发建模。我们的方法的概述如 Fig.2 所示。我们首先在 Sec.3.1 中给出了 HairStep 的定义。然后在 Sec.3.2 和 Sec.3.3 中描述如何从真实的图像中获得。我们在 Sec.4 描述了如何使用 HairStep 进行单视图三维头发建模。

3.1 定义

给定一个目标图像，我们将相应的表示 HairStep 定义为 $H = \{O, D\}$ ，其中 O 和 D 分别为发型图和深度图。发型图 O 表示为尺寸为 $W \times H \times 3$ 的 RGB 图像，其中 W 和 H 为目标图像的宽度和高度。发型图上某个像素 x 的颜色定义为

$$O(x) = (M(x), O_{2D}/2 + 0.5) \quad (1)$$

我们使用红色通道和二进制图 M 表示头发掩模。我们将像素 x 处头发生长的投影二维方向 O_{2D} 的单位向量进行归一化，并用绿色和蓝色通道表示这个生长方向。深度图 D 可以很容易地定义为 $W \times H \times 1$ 图，它表示头发与相机坐标中相机中心的最近距离。HairStep 的可视化例子如 Fig.1 所示。

与现有的表示形式的不同。现有的二维方向图使用具有两个模糊方向[23]的无向线来描述 180 度的像素级别的头发生长，而我们的发型图可以表示 360 度的方向（详情见 Fig.3(d-e)）。NeuralHdHair [37] 试图引入一个额外的亮度图来补充在真实图像中丢失的局部细节。不幸的是，在合成图像和真实图像的亮度之间存在着一个不可忽略的域差距。因为它与照明和材质等渲染场景高度相关。与亮度

图相比，我们的头发深度图只包含几何信息，这有助于缩小合成图像和真实图像的域差距。

3.2 链状图的提取

为了实现基于学习的单视角三维头发建模，需要首先从合成的三维头发数据和真实图像中提取 HairStep 进行训练和测试。对于合成数据，我们可以很容易地从成熟渲染技术[18]的辅助下的三维链模型中获得发型图和深度图。但通过现有的方法从真实图像中提取发型图是不可行的。因此，我们使用一种基于学习的方法，并解释一个数据集 HiSa 来提供监督。

HiSa 数据集。我们从网上收集了 1250 张不同发型的清晰肖像图像。有关发型、性别和种族的统计数据见补充资料。我们首先雇佣艺术家来标注沿着图像上的头发，从头发根部到头发末端的密集的二维方向向量曲线（见 Fig.3(b) 中的例子）。平均而言，每一张头发图片需要花费大约 1 个小时的时间才能画出大约 300 个矢量曲线。

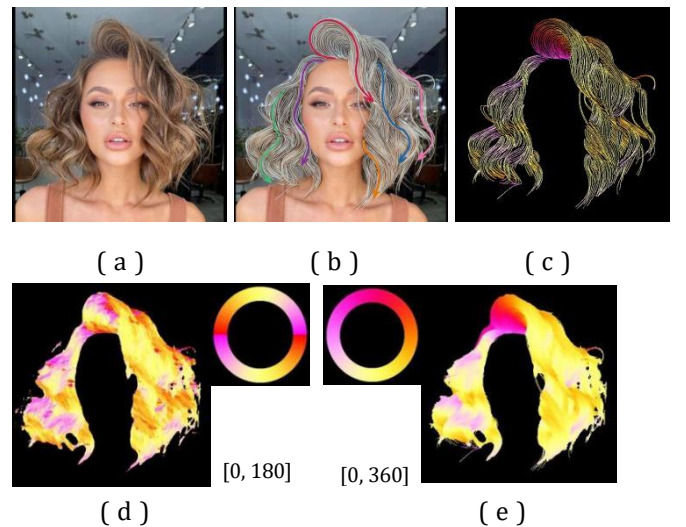


Figure 3 从矢量图中获得发型图。(a) 肖像图像 (b) 带解释的矢量图。(c) 彩色的发型图。(d) 由 Gabor 过滤器提取的方向图。(e) 我们的发型图。

一旦我们获得了头发区域的密集发图,我们就将它们转换为由等式的定义着色的 [Eq.\(1\)](#), 就像 [Fig.3\(c\)](#)所示的那样。最后,我们在头发图内插入彩色笔画,以获得给定图像的真实发型图([Fig.3 \(e\)](#))。由于密集的注释,这些孔很容易被可忽略的细节损失填满。与 Gabor 滤波器提取的无向方向图 ([Fig.3\(d\)](#))相比,我们的发型图是干净的,可以不带歧义地表示生长方向。

发型图预测。我们将从真实图像中提取发型图作为一个图像到图像的转换任务。我们发现,简单地使用一个 U-Net [\[25\]](#)已经可以取得令人满意的结果。根据标准设置,我们使用像素级 L1 损失和感知损失来对抗基于真实图像的真实发型图 O , 其表述为

$$L_{strand} = \frac{1}{C \cdot \sum M} \|\hat{O} - O\|_1 + \alpha \cdot \frac{1}{W_j H_j C_j} \|\varphi_j(\hat{O}) - \varphi_j(O)\|_2^2 \quad (2)$$

其中, \hat{O} 表示预测的发型图, C 表示方向图的通道数。函数 $\varphi_j(\cdot)$ 表示预先训练好的 VGG-19 [\[32\]](#) 的前 j 层,我们将 j 设为 35。 W_j 、 H_j 和 C_j 表示 φ_j 的输出特征的形状。

3.3 域自适应深度估计

从真实的图像中获得头发的深度并不简单,因为我们不能直接获得深度解释。受自然环境下深度图 [\[5\]](#) 的启发,我们将真实图像的头发区域的相对深度标注为弱标签。然而,只有受顺序深度信息的限制,网络往往才会产生非自然的深度图。弱标签学习深度与训练中使用的合成深度之间存在明显的领域差距,导致将训练模型应用于实域时,泛化性较差。根据流行的基于伪标签 [\[7,15,33,42,44,45\]](#) 的域自适应框架,我们提出了一种域自适应深度估计方法,以减小真实和合成数据的差距(见 [Fig.2](#))。

HiDa 数据集。我们在 HiSa 的 1250 幅肖像中,对每幅图像的头发区域中随机选择的像素对进行了深度关系分析。与每幅图像只选择一对的自然环境深度不同,我们平均为每幅图像分析超过 140 对,这可以给出更准确和更密集的预测。我们首先根据

头发和面部面积的比值在头发区域内生成超像素。然后,我们从所有相邻的超像素中随机抽取像素对,最终生成总共 1250 张真实图像的 177,074 个像素对。一对点中的两点分别被染成红色和蓝色。QA 程序的设计目的是通过每次在图像上显示一对来分析顺序深度,并询问“一对采样像素中的哪个点看起来离你更近,红点,蓝点,还是难分辨?”,就与 [\[5\]](#) 相同。12 名训练有素的工人被邀请进行分析,并将其分成三组,以确保每一对选定的组都被不同的组分析了三次。最后收集到 129,079 个有效答案(所有组都给出了一定的相对深度,即红色或蓝色,并彼此一致)。我们的采样平均需要 4.6 秒来决定,三组同意相对深度的时间为 72.9%。

学习深度图。我们通过 [\[5\]](#) 直接学习输入图像 I 和通过沙漏网络 [\[22\]](#) 输出的头发区域密度深度图 D_r 之间的映射,这是由我们分析的弱监督。为了使用深度的顺序标签来训练网络,我们需要一个损失函数来促进预测的深度图与基于真实描绘的图关系相一致。我们发现,在 [\[16,39,43\]](#) 中使用的边际排名损失在我们的任务中很有效:

$$L_{rank} = \frac{1}{N} \sum_{i=1}^N \max(0, (D_r(p_1^i) - D_r(p_2^i)) \cdot r^i + \epsilon). \quad (3)$$

其中 p_1^i 和 p_2^i 是 i_{th} 个对 p^i 像素的注释, r^i 是真实值标签,如果 p_1^i 更接近,则设置为 1, 否则为 -1。 N 表示图像中采样对的总数。 ϵ 被设置为 0.05,这就控制了 p_1^i 和 p_2^i 中的深度值的差异。

领域自适应。虽然顺序标签可以提供局部深度变化,但它是一个弱约束,引入了模糊性,导致了不均匀的解。预测的深度图通常是不自然的,并且充满了锯齿状的伪影(见 [Fig.5](#) 中的侧视图)。将这种深度应用于头发建模通常会导致粗糙和有误差的三维形状。为了解决上述问题,我们提出了一种弱监督域自适应的头发深度估计解决方案。我们相信从合成领域借鉴的知识有助于提高学习深度的质量。

用顺序标签训练的网络不能感知绝对的位置、大小和深度范围。与用于三维头发建模训练的合成深度图相比, $Depth_{syn}$ 有较大的领域差距。为了给出合成域的约束,我们首先训练一个网络深度系统,通过最小化预测与真实图像之间的 L1 距离,从而合成数据集上的发型图中预测深度图。然后将真实图像输入到 $Depth_{syn}$ 中,查询伪标签 \mathcal{D} 作为深度先验。需要注意的是,直接将此伪深度图应用于三维头发

建模是不合理的，因为将发型图作为输入并不能为网络提供足够的三维信息。在深度先验和相对深度标注的弱标签的共同监督下，我们预测了良好的深度图，它不仅外观自然，而且保持了深度排序的局部关系。域自适应深度估计的损失函数由两部分组成，即对伪标签的 L1 损失和在等式中定义的排名损失 (3)：

$$L_{depth} = \beta \cdot ||D_r - D||_1 + L_{rank} \quad (4)$$

4 单个视角的三维头发造型

给定一个单视图肖像图像的 HairStep 表示，我们进一步将其恢复到一个高级的 3D 头发模型。在本节中，我们首先说明 3D 隐式头发表示，然后描述头发造型重建的过程。

4.1 3D 头发表示

继 NeuralHdHair [37] 被认为是最先进的单视图头发建模之后，我们使用隐式占用场和方向场来表示标准头皮的规范空间中的三维头发模型。占用域内的点的值将被分配为 1，则设置为 0。定向场中的一个点的属性被定义为头发生长的单位三维方向。头发体积外的点的方向被定义为零向量。我们使用与[26]相同的方法来提取面部的头发。在训练过程中，我们采样大量的点，形成一个离散的占用场。抽样策略遵循[27]，在网格表面周围随机采样并在边界框内均匀采样，比例为 1: 1。对于方向场，我们计算每个模型超过 10k 条线的密点的单位三维方向。

4.2 发型生成

为了从 HairStep 中生成三维支架，我们首先按照吴等人 [37] 描述的方法训练一个神经网络 NeuralHdHair*。以我们的 HairStep 作为输入，该网络可以预测代表目标三维头发模型的隐式占用场和方向场。然后，我们采用在[30]中生长的方法从标准头皮的发根中合成头发。

神经元 HdHair[37]的代码尚未发布，我们自己的实现神经元 alHdHair*保留了神经元 HdHair 的主要管道和完全丢失的功能，但与原来的 NeuralHdHair 有两个主要的不同之处。首先，我们不使用亮度映射的子模块。亮度有可能为头发重建提供更多的线索，但由于它与光线高度相关，因此在合成图像和真实图像之间存在明显的域差距。我

们试图将亮度图应用到 NeuralHdHair*上，但它只能带来微小的改进。其次，我们放弃了 NeuralHdHair 的生长网，因为我们的工作关注的是重建结果的质量，而不是效率，而生长网的设计是为了加速从 3D 隐含领域到发丝的转换。与[37]中报道的传统头发生长算法相比，它保持了相同的生长性能。

5 实验

5.1 数据集

我们在 USC-HairSalon[10]上训练了所提出的方法，这是一个公开访问的 3D 发型数据库，由 343 个不同风格的合成发型模型组成。我们遵循[46]来增强三维头发数据，并为每个头发选择 3 个随机视图来生成相应的发型图和深度图，进而形成我们的 HairStep。对于我们的真实数据集 HiSa 和 HiDa，我们使用 1054 张分辨率为 512×512 的图像进行训练，196 张图像用于测试。在训练过程中，我们通过随机旋转、缩放、平移和水平翻转来增强图像和解释。

5.2 指标评价

基于 HiSa 和 HiDa，我们提出了两个新颖和公平的指标，即 HairSale 和 HairRide，以评估单视图计算的三维头发建模结果。我们渲染重建的 3D 头发模型以获得发型图 O_r 或深度图 D_r ，然后将它们与我们的真实图像解释 O_{gt} 和 D_{gt} 进行比较。此外，这两个指标也可以应用于 HairStep 的评价。

HairSale. 我们首先计算了在渲染发型图上的增长方向的平均角度误差，其范围从 0 到 180。我们将 HairSale 定义为：

$$\text{HairSale} = \frac{1}{K} \sum_{x^i \in U} \arccos \left(V(O_r(x^i)) \cdot V(O_{gt}(x^i)) \right), \quad (5)$$

其中 U 是渲染的区域和基础区域的相交区域。 K 是 $U \cdot V(O_r(x^i))$ 中的像素的总数，将链映射 O_r 的像素 x^i 处的颜色转换为表示增长方向的单位向量。

HairRide. HairSale 只测试二维空间的匹配程度。我们还需要一个度量的 HairRida 来测量 HiDa 上的相对深度精度，其定义为：

$$\text{HairRida} = \frac{1}{Q} \sum \max \left(0, r^i - \text{sign}(D_r(p_1^i) - D_r(p_2^i)) \right)$$

(6)

注意,我们还计算了渲染相交区域的 HairRida。此外,我们还提供了 IoU 的统计数据以供参考。对于合成数据的评价,我们遵循[37]使用方向场的 L2 误差来计算占用场的精度。

5.3 提取 HairStep 的评价

我们首先评估了我们的从真实图像的 HairStep 提取方法的有效性。我们发现,简单地应用一个 U-Net 已经可以产生干净发型图,而 Gabor 滤波器则会产生不确定的误差(见 Fig.4)。根据我们预测的链图计算出的 HairSale 是 12.3。由于 Gabor 滤波器只能产生无向方向图,我们将发型图转换为无向图,以定量计算 HairSale,以进行公平比较。我们的结果和 Gabor 的 HairSale 分别是 14.2 和 18.4,我们的方法表现得好 22.8%。值得一提的是,无定向的误差比定向的大,因为模糊的双向定向导致更糟糕的测量。

我们使用两个指标来评估深度估计: HairRida 和针对伪标签的 L1 误差(w/ or w/o 标准化),以测量预测深度和合成先验之间的差异。我们比较了域自适应方法 $Depth_{DA}$ 与合成域 $Depth_{pseudo}$ 的伪标签伪,以及 $Depth_{weak}$ 标记顺序弱监督方法的结果。HairRida 的 $Depth_{pseudo}$ 、 $Depth_{weak}$ 和 $Depth_{DA}$ 分别为 80.47%、85.17% 和 85.20%。针对 $Depth_{weak}$ 和 $Depth_{DA}$ 的伪标签(w/ or w/o 标准化)的 L1 误差为 0.2470/3.125 和 0.1768/0.1188。由深度图转换而来的点云与不同视图的定性比较也如 Fig.5 所示。定量和定性给出了相同的结论,即我们的 $Depth_{DA}$ 更有能力平衡深度的局部细节和全局形状与合成先验的相似性。但是 $Depth_{weak}$ 是不自然的,充满了锯齿状的人工制品。 $Depth_{pseudo}$ 受到平面几何形状的影响,因为发型图不能提供强大的三维提示。

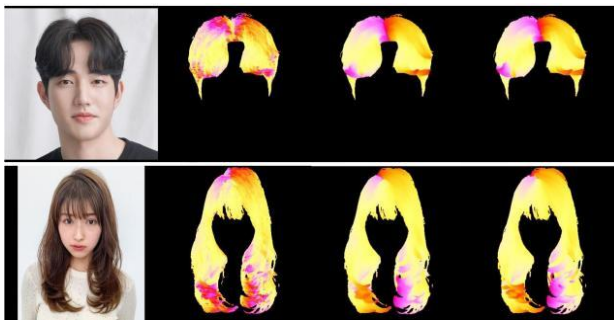


Figure 4 对方向/发型图的定性比较。(a)输入图像;(b)来自 Gabor 滤波器的无向方向映射(c)发型图;(d)基于真实的图像。

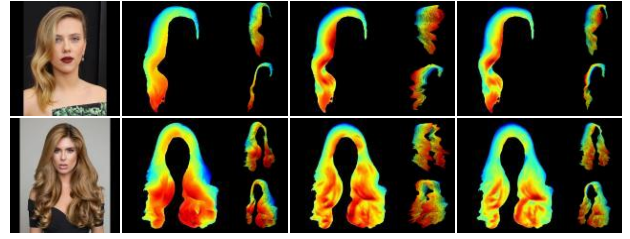


Figure 5.关于深度估算的定性比较。



Figure 6.与以前的方法[37,38,46]的比较。



Figure 7.定性评价结果。从左到右分别是:输入图像, NeuralHDHair*的结果,使用我们的发型图表示的结果,以及完整方法(Full)的结果。

比较表示。为了评估我们的 HairStep 的效果,我们将其与发型图图和三种不同框架上的方向图和现有的方向图[46]进行了比较,即 NeuralHDHair*、DynamicHair[38]和 HaerNet[46]。合成数据和真实数据的定量比较分别见 Tab.1 和 Tab.2。如标签页中所示 Tab.1,我们的表示比这三种方法合成的数据更好。由于 HairNet 只能输出头发丝,我们遵循[46]在 Tab.1 中报告平方距离误差。在 Tab.2 中是对 HiSa 和 HiDa 的评价。证明了使用我们的发型图比使用之前的方向图[46]能实现更好的毛发生长对齐,因为之前的方向图存在方向模糊和图像误差。HairNet 和 DynamicHair 的泛化能力受到全局特征的使用的限制。因此,直接将深度信息连接到输入中似乎没有帮助。通过完整的 HairStep, NeuralHDHair*的深度精度有明显的提高。通过对 Fig.7 中所示的定性比较,得出了相同的结论,其中 HairStep 在深度上表现最好,并保持了与发型图一样的头发生长的精细

排列。仅应用方向图才会导致不需要的工件。请注意, [Tab.2](#) 中的 HairNet 和 DynamicHair 的深度精度是基于低 IoU, 这与 NeuralHDHair* 不同。此外, 我们对 39 个用户随机抽取的 10 个用户的例子进行了用户研究, 从 3 种表示中获得 NeuralHDHair* 的重建结果。64.87% 的人选择 HairStep 的结果最好, 选择发型图的占 21.28% 和选择无向图的占 13.85%

Method	Orien. err. ↓	Occ. acc. ↑
NeuralHDHair* (Orientation map)	0.1324	82.59%
NeuralHDHair* (Strand map)	0.0722 (-41.7%)	84.18%
NeuralHDHair* (HairStep)	0.0658 (-50.3%)	86.77%
DynamicHair (Orientation map)	0.1352	78.19%
DynamicHair (Strand map)	0.1185 (-12.4%)	79.62%
DynamicHair (HairStep)	0.1174 (-13.2%)	79.78%
HairNet (Orientation map)	0.02349	/
HairNet (Strand map)	0.02206 (-6.1%)	/
HairNet (HairStep)	0.02184 (-7.0%)	/

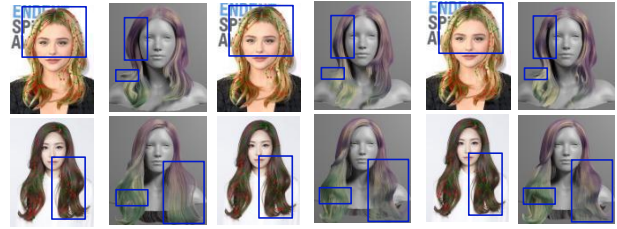
Table 1. 使用不同中间表示的 USC-HairSalon 数据集对 NeuralHDHair*, DynamicHair [38] HairNet [46] 进行定量比较

Method	IoU ↑	HairSale ↓	HairRida ↑
NeuralHDHair* (Orientation map)	77.56%	19.6	70.67%
NeuralHDHair* (Strand map)	77.6%	16 (-18.4%)	72.37%
NeuralHDHair* (HairStep)	77.22%	16.36 (-16.5%)	76.79%
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%
HairNet (Orientation map)	57.15%	31.97	75.65%
HairNet (Strand map)	57.48%	28.6 (-10.5%)	74.81%
HairNet (HairStep)	57.01%	27.68 (-13.4%)	74.97%

Table 2. 神经元 HDHair*、DynamicHair [38] 和 HairNet [46] 的不同中间表示形式的 HiSa 和 HiDa 的定量比较。

Method	IoU ↑	HairSale ↓	HairRida ↑
C0	77.75%	16.03(-18.2%)	73.57%
C1	77.11%	16.54(-15.6%)	75.8%
Full	77.22%	16.36(-16.5%)	76.79%

Table 3. 关于深度估算的定量消融研究。



C0 C1 C2

Figure 8. 定性消融研究结果。每对从左到右: 输入具有 HairRida 可视化的图像, 其中绿色/红线表示对两个终点相对深度的预测为右 (错误) 和重建的三维发链模型(见 [Sec.5.5](#) 为详细说明)。

5.5. 消融研究

为了更好地研究每种设计的深度估计对最终结果的影响, 我们的表示法有三种构型:

- C0: 发型图+ $Depth_{pseudo}$ 。
- C1: 发型图+ $Depth_{weak}$ 。
- Full: 发型图+ $Depth_{DA}$ 。

定量比较报告见表 [Tab.3](#), 定性结果如 [Fig.8](#) 所示。我们的完整表现是深度精度和头发生长的最佳结果。C0 具有深度的平面几何形状。同时, C1 可以产生具有良好的深度精度的结果, 但在头发生长的排列上得到了比完整表示方法(Full)表示相对较大的差异。

6 结论

在这项工作中, 我们重新思考了单视图三维头发建模的整体解决方案, 认为一个适当的中间表示来弥补合成数据和真实数据之间的域差距是必要的。为此, 我们提出了一种新的三维头发表示 HairStep, 它由发型图和深度图组成, 以缩小现有的域间隙。我们还收集了两个数据集, 即 HiSa 和 HiDa, 我们还手动解释了发型图和深度图。这些数据集不仅允许训练我们基于学习的方法, 还引入公平和客观的指标来评估单视图三维头发建模的性能。在不同的例子上进行的大量实验证明了我们的新表示的有效性。我们的方法可能会在一些罕见和复杂的发型上失败, 因为三维网络在当前的数量和多样性有限的合成数据集上基本上被过度拟合。

致谢。这项工作得到了 NSFC 和 No. 62293482 基金的部分支持, 基础研究项目 No.HZQB-KCZYZ-2021067 河套深圳香港 S&T 科技合作区, 国家重点 R&D 研发计划项目 No. 2018YFB1800800, 深圳优秀人才培养基金 202002, 广东省研究项目 No. 2017ZT07X152 和 No. 2019CX01X104, 广东省未来情报网络重点实验室(资助项目 No.2022B12120100 01)和深圳大数据和人工智能关键实验室(No.ZDSY S201707251409055). 它也得到了深圳通用项目的部分支持 JCYJ20220530143604010.

参考

- [1] Yongtang Bao and Yue Qi. A survey of image-based techniques for hair modeling. *IEEE Access*, 6:18670–18684, 2018. 1 (包永堂, 齐越. 对基于图像的头发表建模技术的调查. 计算机视觉和模式识别会议. 6:18670-18684, 2018, 1)
- [2] Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. Autohair: Fully automatic hair modeling from a single image. *ACM Transactions on Graphics*, 35(4), 2016. 1, 2 (柴梦磊, 邵天家, 吴洪志, 翁艳林, 周坤. 自动头发: 全自动的头发表建模从单个图像. ACM 图像处理, 35(4), 2016. 1, 2)
- [3] Menglei Chai, Lvdi Wang, Yanlin Weng, Xiaogang Jin, and Kun Zhou. Dynamic hair manipulation in images and videos. *ACM Transactions on Graphics (TOG)*, 32(4):1–8, 2013. 1, 2 (柴梦蕾, 王绿迪, 翁彦霖, 金小刚, 周坤. 在图像和视频中的动态头发操作. ACM 图像处理 (TOG), 32(4):1–8, 2013. 1, 2)
- [4] Weifeng Chen, Zhao Fu, Dawei Yang, and Jia Deng. Single-image depth perception in the wild. *Advances in neural information processing systems*, 29, 2016. 2, 3, 5, 11 (柴梦蕾, 王绿迪, 翁彦霖, 金小刚, 周坤. 单视图的头发表建模的肖像操作. ACM 图像处理 (TOG), 32(4):1–8, 2013. 1, 2)
- [5] Weifeng Chen, Zhao Fu, Dawei Yang, and Jia Deng. Single-image depth perception in the wild. *Advances in neural information processing systems*, 29, 2016. 2, 3, 5, 11 (陈伟峰, 赵甫, 杨大伟, 贾登. 在自然环境的单图像深度感知. 神经信息处理系统的研究进展. 32(4):1–8, 2013. 1, 2)
- [6] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision*, pages 2650–2658, 2015. 3 (大卫艾根, 罗伯费格斯. 使用通用的多尺度卷积体系结构预测深度, 表面法线和语义标签. IEEE 计算机视觉国际会议, 第 2650-2658 页, 2015. 3)
- [7] Hehe Fan, Xiaojun Chang, Wanyue Zhang, Yi Cheng, Ying Sun, and Mohan Kankanhalli. Self-supervised global- local structure modeling for point cloud domain adaptation with reliable voted pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6377–6386, 2022. 5 (范荷荷, 张晓军, 张万岳, 易毅, 孙英, 坎坎哈利. 基于可靠的投票伪标签的点云域自适应的自监督全局局部构建建模. IEEE/CVF 计算机视觉和模式识别会议. 第 6377-6386 页, 2022. 5)
- [8] Derek Hoiem, Alexei A Efros, and Martial Hebert. Automatic photo pop-up. In *ACM SIGGRAPH*, pages 577–584, 2005. 3 (德里克·霍伊姆, 阿列克谢·埃弗罗斯, 军事赫伯特. 自动弹出照片. ACM 计算机绘图专业组. 第 577-584 页, 2005. 3)
- [9] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. Robust hair capture using simulated examples. *ACM Transactions on Graphics (TOG)*, 33(4):1–10, 2014. 1, 3 (胡立文, 马崇, 罗林杰, 李浩. 使用模拟的例子进行稳定的头发捕捉, ACM 图像处理 (TOG). 33(4):1–10, 2014. 1, 3)
- [10] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (ToG)*, 34(4):1–9, 2015. 1, 2, 3, 6 (胡立文, 马崇, 罗林杰, 李浩. 使用发型数据库进行单视发型建. ACM 图形处理 (TOG). 34(4):1-9, 2015. 1, 2, 3, 6)
- [11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 11 (伊索拉, 朱军燕, 周廷辉和阿列克谢. 使用有条件的对抗网络的图像到图像的转换. IEEE 关于计算机视觉和模式识别的会议. 第 1125-1134 页, 2017. 11)
- [12] Kevin Karsch, Ce Liu, and Sing Bing Kang. Depth transfer: Depth extraction from video using non-parametric sampling. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2144–2158, 2014. 3 (凯文卡尔施, 刘茜, 康辛冰. 深度转移: 使用非参数采样从视频中提取深度. IEEE 关于模式分析和机器智能的处理. 36(11):2144-2158, 2014. 3)
- [13] Lubor Ladicky, Jianbo Shi, and Marc Pollefeys. Pulling

- things out of perspective. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 89–96, 2014. 3 (吕博拉狄克, 石建波, 麦西·波莱菲. 透视提取. 在 IEEE 关于计算机视觉和模式识别的会议的会议. 第 89-96 页, 2014.3)
- [14] Bo Li, Chunhua Shen, Yuchao Dai, Anton Van Den Hengel, and Mingyi He. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical crfs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1119–1127, 2015. 3 (李博, 沈春华, 戴宇超, 范登恒哲, 何明义. 利用深度特征和分层立体流场回归对单眼图像进行深度和表面正态估计. 在 IEEE 关于计算机视觉和模式识别的会议的会议. 第 1119-1127 页, 2015.3)
- [15] Jian Liang, Dapeng Hu, and Jiashi Feng. Domain adaptation with auxiliary target domain-oriented classifier. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16632–16642, 2021. 5 (梁吉安, 胡大鹏, 冯家诗. 基于辅助目标域导向分类器的域自适应. 在 IEEE/CVF 计算机视觉和模式识别会议. 第 16632-16642 页, 2021.5)
- [16] Kwan-Yee Lin and Guanxiang Wang. Hallucinated-iqa: No-reference image quality assessment via adversarial learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 732–741, 2018. 5 (林颖怡, 王冠翔. 通过对抗性学习对图像质量进行无定性评估. IEEE 关于计算机视觉和模式识别会议, 第 732-741 页, 2018.5)
- [17] Fayao Liu, Chunhua Shen, and Guosheng Lin. Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5162–5170, 2015. 3 (刘法姚, 沈春华, 林国胜. 用于从单个图像中进行深度估计的深度卷积神经网络. 在 IEEE 关于计算机视觉和模式识别会议, 第 5162-5170 页, 2015.3)
- [18] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7708–7717, 2019. 4 (刘世臣, 李天业, 陈伟凯, 郝浩. 软光栅化器: 一种用于基于图像的 3D 推理的可微渲染器. IEEE/CVF 计算机视觉国际会议论文集, 第 7708-7717 页, 2019.4)
- [19] 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. 1, 2, 3 (罗林杰, 李浩, 西尔万·维斯, 马克·保利和辛凯维奇. 使用方向域的多视图头发捕捉. 在 2012 年 IEEE 计算机视觉与模式识别会议. 第 1490-1497 页, 2012.1, 2, 3)
- [20] Linjie Luo, Cha Zhang, Zhengyou Zhang, and Szymon Rusinkiewicz. Wide-baseline hair capture using strand-based refinement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 265–272, 2013. 3 (罗林杰, 张茶, 张正友, 西尔万·宽基线的头发捕获使用基于链状基线的细化. IEEE 计算机视觉和模式识别会议, 第 265-272 页, 2013.3)
- [21] Giljoo Nam, Chenglei Wu, Min H Kim, and Yaser Sheikh. Strand-accurate multi-view hair capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 155–164, 2019. 2 (南吉珠, 吴成雷, 金敏, 谢赫. 精确发丝的多视图的头发捕捉. IEEE/CVF 计算机视觉和模式识别会议. 第 155-164 页, 2019.2)
- [22] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hour-glass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016. 5 (亚历杭德罗·纽维尔, 杨凯宇, 贾登. 用于人体姿态估计的堆叠沙漏网络, 欧洲计算机视觉会议. 第 483-499 页, 施普林格, 2016.5)
- [23] Sylvain Paris, Hector M Briceno, and Francois X Sillion. Capture of hair geometry from multiple images. *ACM transactions on graphics (TOG)*, 23(3):712–719, 2004. 2, 3, 4 (西尔万巴黎, 赫克托·布里克诺(法国). 从多个图像中捕获头发的几何图形. ACM 图形处理 (TOG), 23(3):712-719, 2004.2, 3, 4)

- [24] 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 Trans. Graph.*, 27(3):30, 2008. 3 (西尔万,张外,奥列格,贾罗斯,马图西克,兹威克,和杜兰德.头发照相馆:真实发型的几何和光度采集.ISO4 标准期刊.27(3):30,2008.3)
- [25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4 (奥拉夫朗内伯格,菲利普费舍尔和托马斯布罗克斯.U-net:用于生物医学图像分割的卷积网络.医学图像计算和计算机辅助干预国际会议.第 234-241 页,施普林格,2015.4)
- [26] Shunsuke Saito, Liwen Hu, Chongyang Ma, Hikaru Ibayashi, Linjie Luo, and Hao Li. 3d hair synthesis using volumetric variational autoencoders. *ACM Transactions on Graphics (TOG)*, 37(6):1–12, 2018. 2, 3, 5 (齐藤顺助,胡立文,马崇,桥光,罗林杰,李浩.使用体积变分自动编码器进行三维头发合成.ACM 图形处理 (TOG).37(6):1–12, 2018.2,3,5)
- [27] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2304–2314, 2019. 5 (齐藤俊助,曾黄,夏敏敏,森岛茂,金泽安珠,李浩.Pifu:像素对齐的隐式功能,用于高分辨率覆盖的人类数字化.IEEE/CVF 计算机视觉国际会议.第 2304-2314 页,2019.5)
- [28] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 84–93, 2020. 3 (齐藤顺助,西蒙,萨拉吉,俊俊.Pifuhd:多级像素对齐隐式功能的高分辨率三维人类数字化.IEEE/CVF 计算机视觉和模式识别会议,第 84-93 页,2020.3)
- [29] Ashutosh Saxena, Min Sun, and Andrew Y Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2008. 3 (阿舒托什萨克森纳,孙敏和安德鲁.制作 3d:从一个单一的静态图像中学习 3D 场景结构.31(5):824–840,2008.3)
- [30] 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. 2, 3, 6 (沈元,张长庚,傅洪波,周坤,郑友谊.基于深度草图的 3d 头发建模.IEEE 关于可视化和计算机图形学的会议.27(7):3250-3263,2020.2, 3, 6)
- [31] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *European conference on computer vision*, pages 746–760. Springer, 2012. 3 (内森西尔伯曼,德里克霍耶姆,普什米特科利,和罗伯费格斯.室内分割和支持从 rgb-d 图像的推断.在欧洲计算机视觉会议,第 746-760 页,施普林格,2012.3)
- [32] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4 (凯伦·西蒙尼扬和安德鲁·泽塞耳曼.大规模图像识别的非常深的卷积网络,arxiv:1409,1556,2014.4)
- [33] Liangchen Song, Yonghao Xu, Lefei Zhang, Bo Du, Qian Zhang, and Xinggang Wang. Learning from synthetic images via active pseudo-labeling. *IEEE Transactions on Image Processing*, 29:6452–6465, 2020. 5 (宋良辰,徐永好,张乐飞,杜波,张谦,王兴刚.通过主动伪标记从合成图像中学习.IEEE 图像处理,29:6452-6465,2020.5)
- [34] Zhuo Su, Lan Xu, Zerong Zheng, Tao Yu, Yebin Liu, and Lu Fang. Robustfusion: Human volumetric capture with data-driven visual cues using a rgb-d camera. In *European Conference on Computer Vision*, pages 246–264. Springer, 2020. 3 (卓苏,兰旭,郑泽荣,陶宇,刘叶斌,陆芳.稳健融合:使用 rgb-d 相机进行数据分割视觉线索的人类体积捕获,欧洲计算机视觉会议.第 246-264 页,施普林格,2020.3)

- [35] Zhentao Tan, Menglei Chai, Dongdong Chen, Jing Liao, Qi Chu, Lu Yuan, Sergey Tulyakov, and Nenghai Yu. Michigan: multi-input-conditioned hair image generation for portrait editing. *ACM Transactions on Graphics (TOG)*, 39(4):95–1, 2020. 3 (谭振涛, 柴梦蕾, 陈冬冬, 景辽, 齐楚, 陆远, 谢尔盖·图利亚科夫, 余能海. 多输入条件的头发图像生成的肖像编辑. ACM 图形处理(TOG).39(4):95-1,2020.3)
- [36] Sicong Tang, Feitong Tan, Kelvin Cheng, Zhaoyang Li, Siyu Zhu, and Ping Tan. A neural network for detailed human depth estimation from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7750–7759, 2019. 3 (唐思聪, 谭费通, 程开尔文, 李朝阳, 朱思雨, 谭萍等. 一种从单一图像中详细估计人体深度的神经网络. IEEE/CVF 计算机视觉国际会议, 第 7750-7759 页, 2019.3)
- [37] Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi 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*, pages 1526–1535, 2022. 2, 3, 4, 5, 6, 7, 11 (吴友, 叶一凡, 杨灵晨, 傅宏波, 周坤, 郑友谊. 神经毛发: 使用隐式神经表征从单一图像自动高保真头发建模. IEEE/CVF 计算机视觉和模式识别会议, 第 1526-1535 页, 2022.2, 3, 4, 5, 6, 7, 11)
- [38] Lingchen Yang, Zefeng Shi, Youyi Zheng, and Kun Zhou. Dynamic hair modeling from monocular videos using deep neural networks. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019. 2, 3, 7, 8 (杨灵晨, 石泽锋, 郑友谊, 周坤. 利用深度神经网络对单眼视频进行动态头发建模. ACM 图形处理(TOG), 38(6):1-12, 2019.2,3,7,8)
- [39] Xiaoshan Yang, Tianzhu Zhang, Changsheng Xu, Shuicheng Yan, M Shamim Hossain, and Ahmed Ghoneim. Deep relative attributes. *IEEE Transactions on Multimedia*, 18(9):1832–1842, 2016. 5 (杨萧山, 张天柱, 许长生, 燕水成, 沙敏侯赛因, 高敏. 深度相对属性. IEEE 多媒体学报. 18(9):1832-1842, 2016.5)
- [40] Meng Zhang, Pan Wu, Hongzhi Wu, Yanlin Weng, Youyi Zheng, and Kun Zhou. Modeling hair from an rgb-d camera. *ACM Transactions on Graphics (TOG)*, 37(6):1–10, 2018. 2 (张孟, 潘武, 吴洪志, 翁艳林, 郑友谊, 周坤. 用 rgb-d 相机拍摄的头发模型. ACM 图形处理(TOG).37(6):1–10, 2018.2)
- [41] Meng Zhang and Youyi Zheng. Hair-GAN: Recovering 3D hair structure from a single image using generative adversarial networks. *Visual Informatics*, 3(2):102–112, 2019. 2, 3 (张孟, 郑友谊. 头发移植: 使用生成式对抗网络从单个图像中恢复 3D 头发结构. 视觉信息学. 3(2):102-112, 2019.2,3)
- [42] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12414–12424, 2021. 5 (张潘, 张波, 张婷, 陈东, 王勇, 方文. 领域自适应语义分割的典型伪标签去噪和目标结构学习. 发表 IEEE/CVF 计算机视觉和模式识别会议, 第 12414-12424, 2021.5)
- [43] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3096–3105, 2019. 5 (张文龙, 刘一豪, 赵董, 于乔. 具有图像超分辨率的生成对抗网络. IEEE/CVF 计算机视觉国际会议, 第 3096-3105 页, 2019.5)
- [44] Shanshan Zhao, Huan Fu, Mingming Gong, and Dacheng Tao. Geometry-aware symmetric domain adaptation for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9788–9798, 2019. 5 (赵珊珊, 桓福, 龚明明, 陶大成. 用于单眼深度估计的几何感知对称域自适应. IEEE/CVF 计算机视觉和模式识别会议, 第 9788-9798 页, 2019.5)
- [45] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks. In *Proceedings of the European conference on computer vision*

(ECCV), pages 767–783, 2018. 5 (郑传霞,张达家,蔡剑飞.T2net:用于解决单幅图像深度估计任务的合成到现实的转换.欧洲计算机视觉会议(ECCV),第 767-783 页,2018.5)

- [46] Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. Hairnet: Single-view hair reconstruction using convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 235–251, 2018. 2, 3, 6, 7, 8, 11 (周毅,胡立文,军兴,陈伟凯,韩伟,辛通,郝浩.头发网络:利用卷积神经网络进行单视图的头发重建.欧洲计算机视觉会议(ECCV).第 235-251 页,2018.2, 3, 6, 7, 8, 11)

- 补充材料 -

A. 数据集

数据分布统计。为了构建 HiSa 和 HiDa, 我们从互联网上收集了 1250 张不同发型的清晰肖像图像, 其中 80% 为女性, 20% 为男性。我们根据其发梢的位置, 将收集到的发型分为短、中、长三类。如果头发末端在嘴上方, 发型将被归类为短。如果头发的末端在肩膀以下, 那么这个发型就属于长类。否则, 它是中间的。我们收集了 300 个短发, 300 个中毛和 650 个长发。对于卷曲类型, 直数、波浪数、卷曲数分别为 210、620、420。更多关于注释和深度采样的例子如图 Fig. 9 所示。

B. 实施细节

我们将在本节中描述我们的网络和关于 HairStep 提取和 3D 头发重建的训练细节。

Hairstep 提取。我们在 [11] 中使用相同的 U-Net, 从分辨率为 512×512 的真实图像中提取发型图。该网络由一个八层编码器和一个八层解码器组成, 其中每一层下采样/上采样为 2 倍, 对称层之间采用跳过连接。我们建议读者参考 [11] 的详细设计。培训在 1 NVIDIA RTX3090Ti 卡上使用 16 批处理 50 个周期, 持续约 12 小时。学习率为 0.0003。在训练过程中, 减重的 α 被设置为 0.1。

我们在 [5] 中使用相同的沙漏网络来估计分辨率为 512×512 的真实图像的深度图。沙漏网络由四个堆栈组成, 它由一系列的卷积、下采样、上采样和跳过连接组成。详情请参考 [5]。该网络在 2 个 NVIDIA RTX3090Ti 卡上进行 100 个周期的批量训练, 持续约 6 小时。学习率为 0.0003, 减重 β 设置为 0.1。

三维头发重建。我们使用与 [37] 中的 IRHaieNet 相同的结构, 其中, 我们首先通过 U-Net 结合 VIFu 从输入大小为 $256 \times 256 \times 256$ 的表示中提取 $96 \times 128 \times 128 \times 64$ 特征体, 然后使用两个 MILPs 查询粗糙的 3D 占用字段和方向字段。对于精细模块, 我们将亮度映射替换为调整大小为 1024×1024 的输入表示, 并通过沙漏网络和两个 MLPs 提取高分辨率的占用场和方向场。有关网络设计的详情, 请参考 [37]。我们遵循 [46], 将身体掩模结合到发型图/方向图的掩模通道上, 而不是引入一

个新的通道。请注意, 我们的 HairStep 比方向图和发型图有更多的一个深度通道。因此, 当使用 HairStep 时, 编码器的第一层有 4 个通道, 而当以发型图或方向图作为输入时, 有 3 个通道。培训是在一个 NVIDIA RTX3090Ti 卡上使用 2 批处理 100 个周期, 大约 5-6 天。学习速率被初始化设置为 0.0001, 并在第 60 阶段衰减了 0.1 倍。

C. 后视图

Fig. S10 中显示了两个后视图的例子, 不可见部分往往光滑, 但仍然合理。这是因为三维头发数据集提供了形状先验。

D. 故障案例

如结论中所述, 由于现有的三维头发数据集的数量和多样性有限, 我们的方法可能会在一些罕见和复杂的发型上失败。例如, 如 Fig. S12 所示, 我们的方法不适用于带有辫子 (左) 和复杂卷曲图案 (右) 的发型。

E. 更多比较

知觉损失。我们认为感知损失在发型图预测中是必要的。虽然它不能提供明显的定量改进 (w/ 14.2 v.s. w/o 14.1), 它带来了视觉上更清晰的局部特征 (Fig. S11)。同时, 我们在没有感知损失的方法下, 做了一个额外的三维头发重建实验。我们发现其 HairSale 和 HairRida (16.51 和 75.3%) 比使用感知损失 (16.36 和 76.79%) 更差。

对不同表征的定性比较。图中显示了对不同表示形式的更多的定性比较。Fig. S13 中, 使用 HairStep 的效果最好。

深度消融术的定性比较。更多的深度消融的定性比较如 Fig. S14 中所示, 我们的完整模型在深度上达到了最好的精度。

F. 用户研究

我们对 10 个随机抽取的涉及 39 个用户的例子进行了用户研究, 从三种表示中获得 NeuralHDAir* 的重建结果。64.87% 的人选择 HairStep 的结果最好, 21.28% 和 13.85%。Fig. S15 和 Fig. S16 为每个示例提供了 3 种不同的表示。

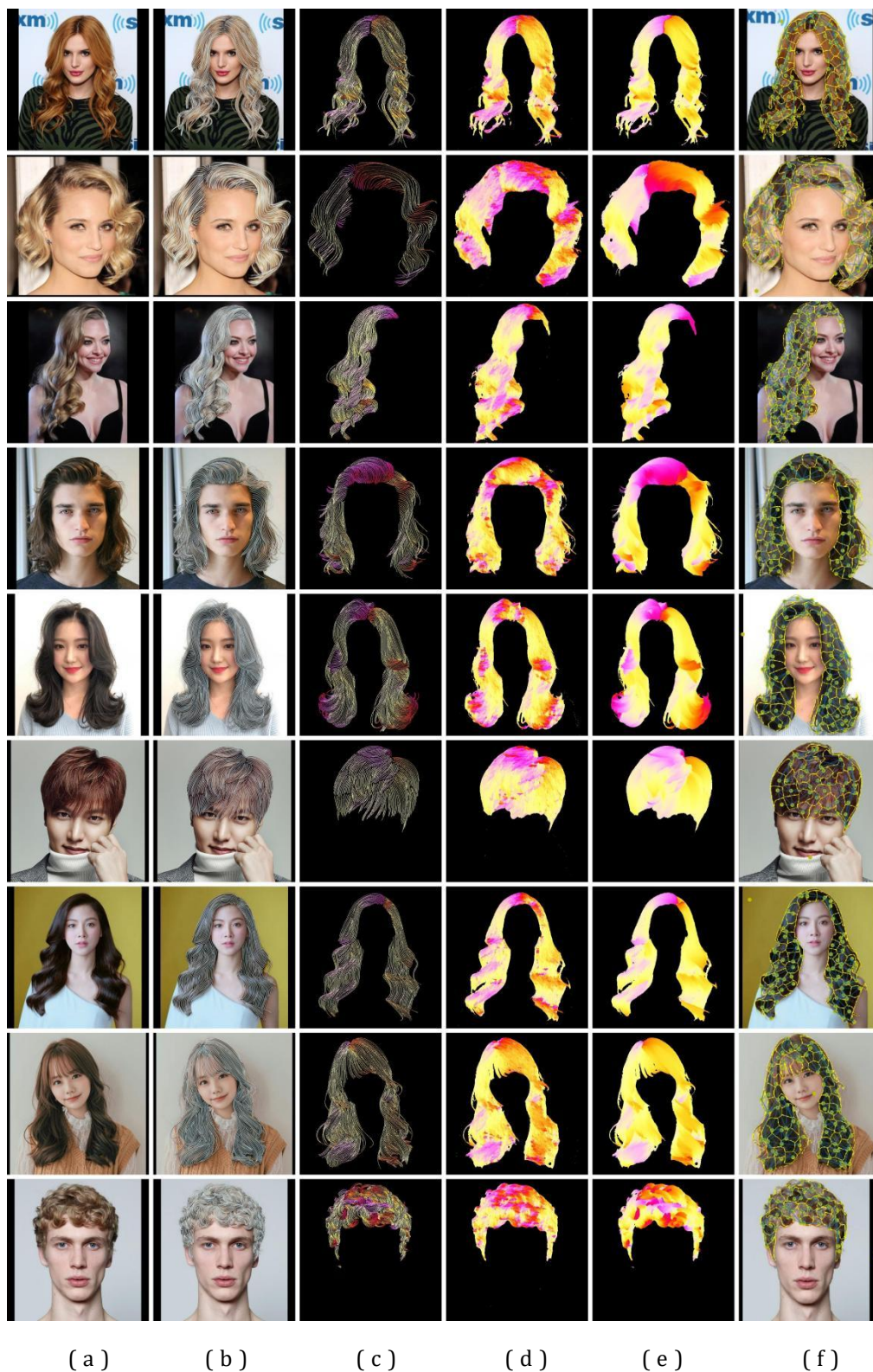


Figure S9: 更多的发型图注释和深度采样的例子。从左到右: (a) 收集的图像, 由艺术家绘制的 (b) 笔画 (c) 彩色笔画, 来自 Gabor 滤波器, (d) 无方向图, (e) 发型图, (f) 超像素的深度对采样。



Figure.S10 带有前后视图的例子

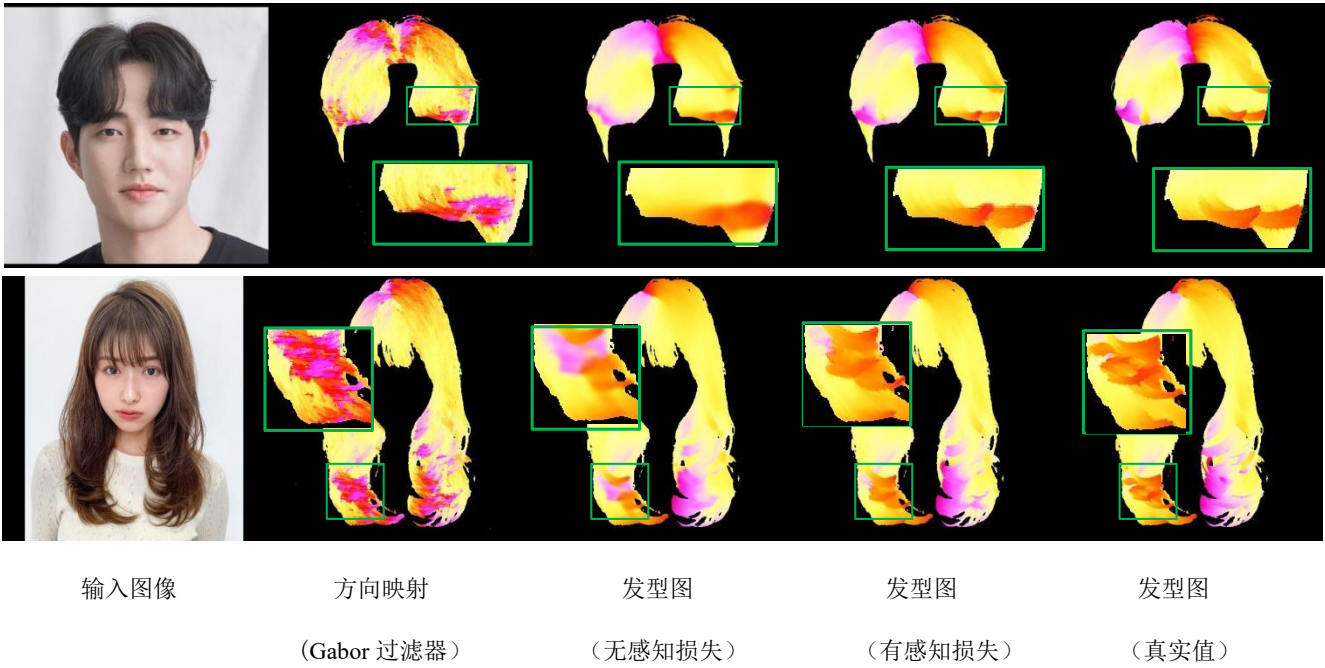


Figure.S11 对方向/发型图的定性比较

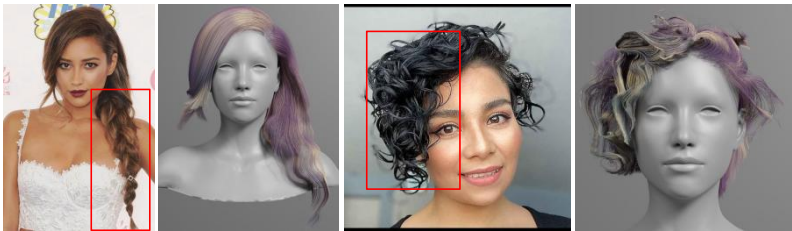


Figure.S12 故障案例

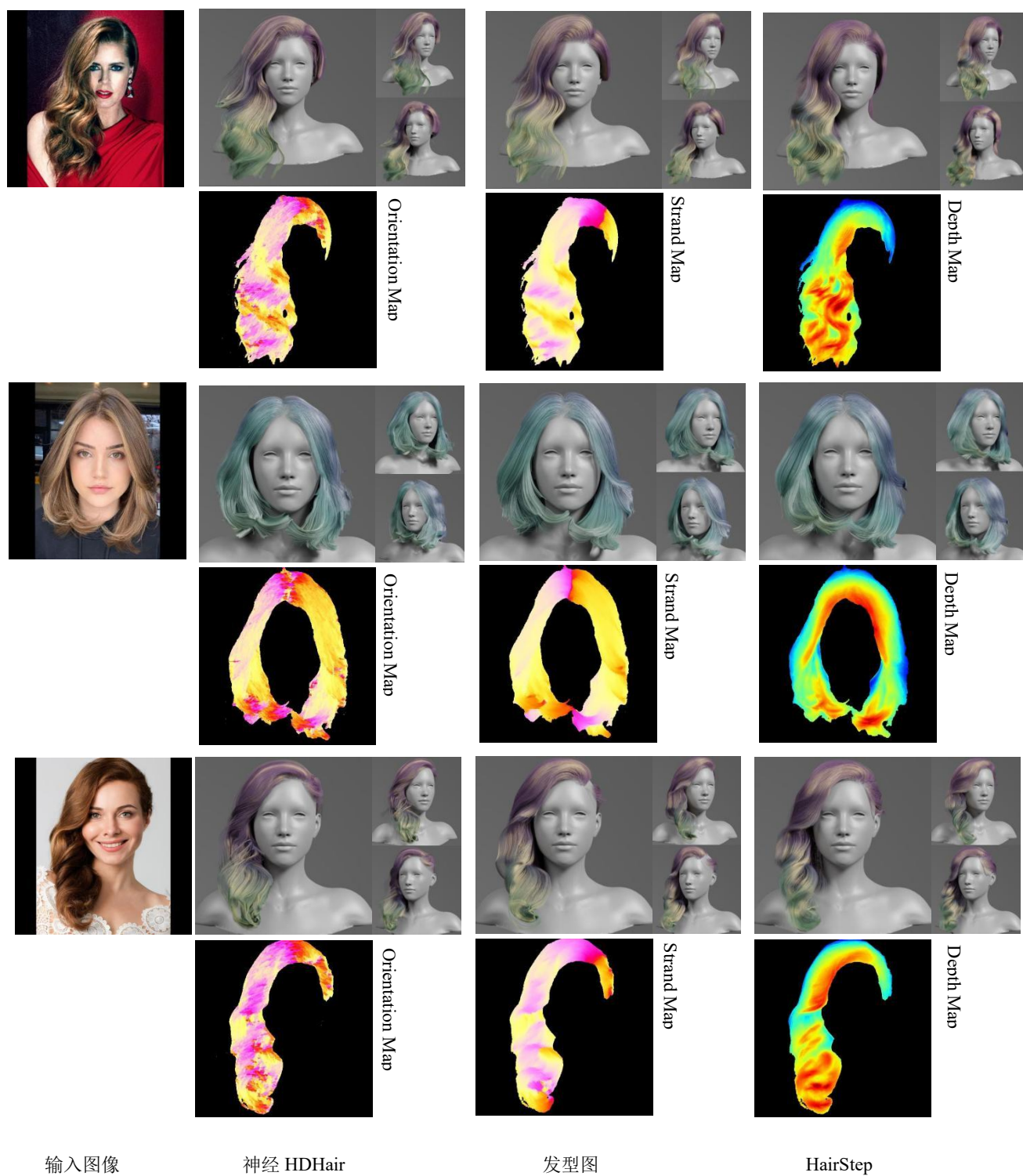
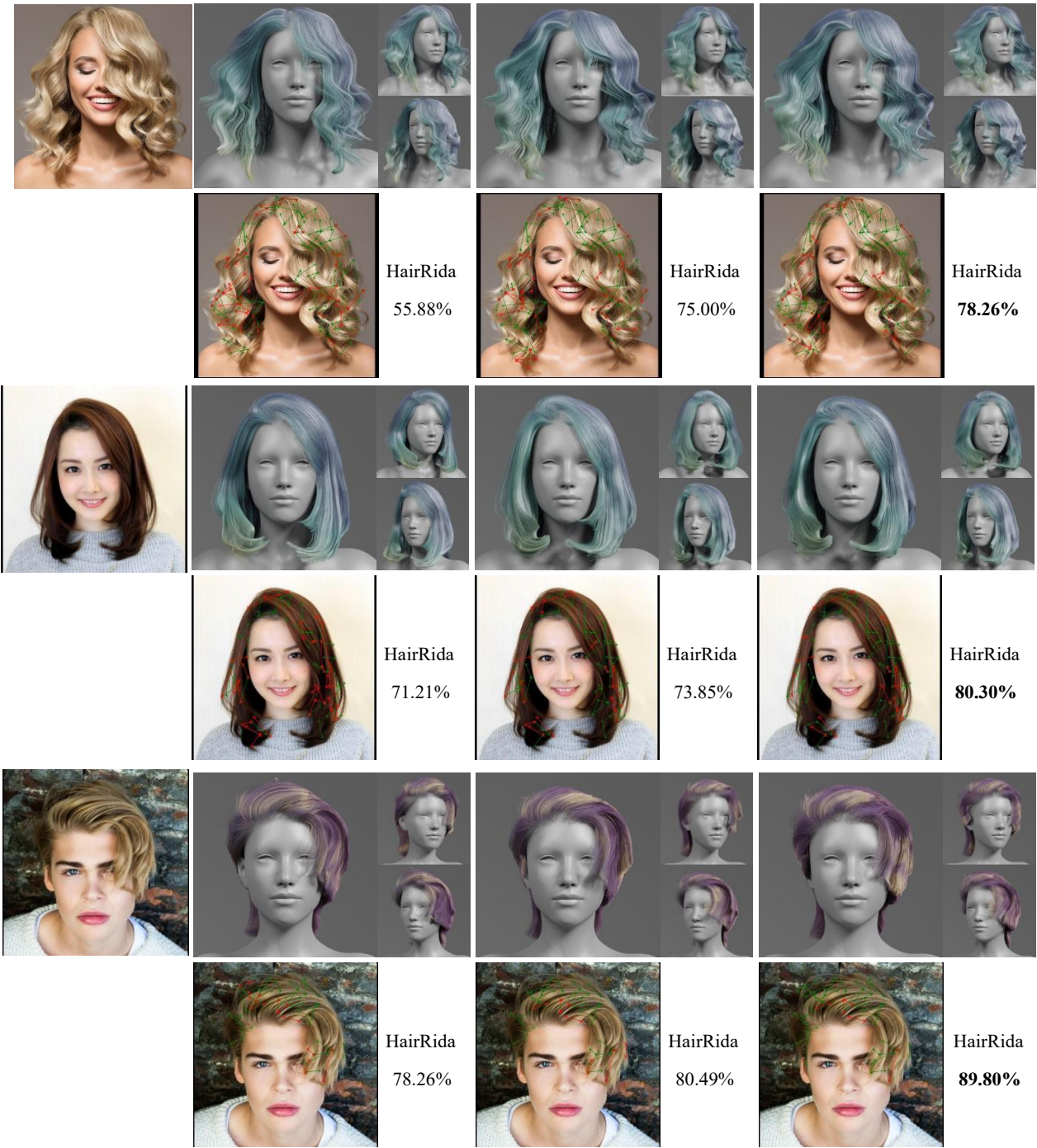


Figure.S13 对不同的表示进行更多的定性比较。从左到右：分别为输入图像，NeuralHDHair*的结果，基于发型图的结果，表示 HairStep 的结果。重建结果下还显示了 Gabor 滤波器的方向图、预测的链图和深度图。



输入图像

C0

C1

Full

Figure.S14 更多关于深度消融的定性比较。从左到右分别为输入图像，C0、C1 的结果和 Full 方法。我们还在每个重建结果下面可视化了 HairRida，其中绿色/红线表示正确/错误的预测。

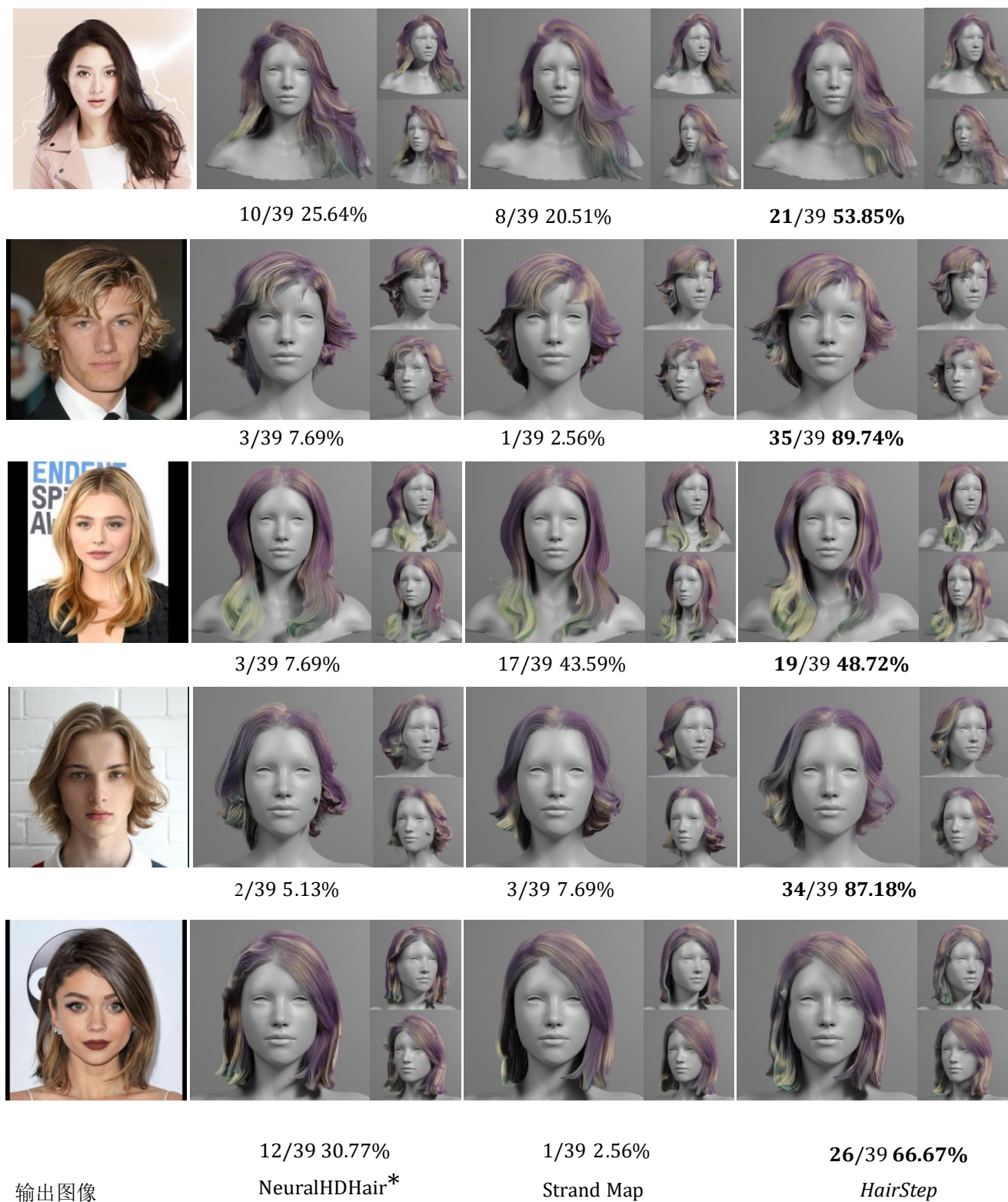


Figure S15 用户研究示例子。从左到右分别为：输入图像，NeuralHDHair*的结果，使用发型图表示的结果，以及完整方法(Full)的结果。我们还为每个例子提供了 3 种不同表示的统计数据。

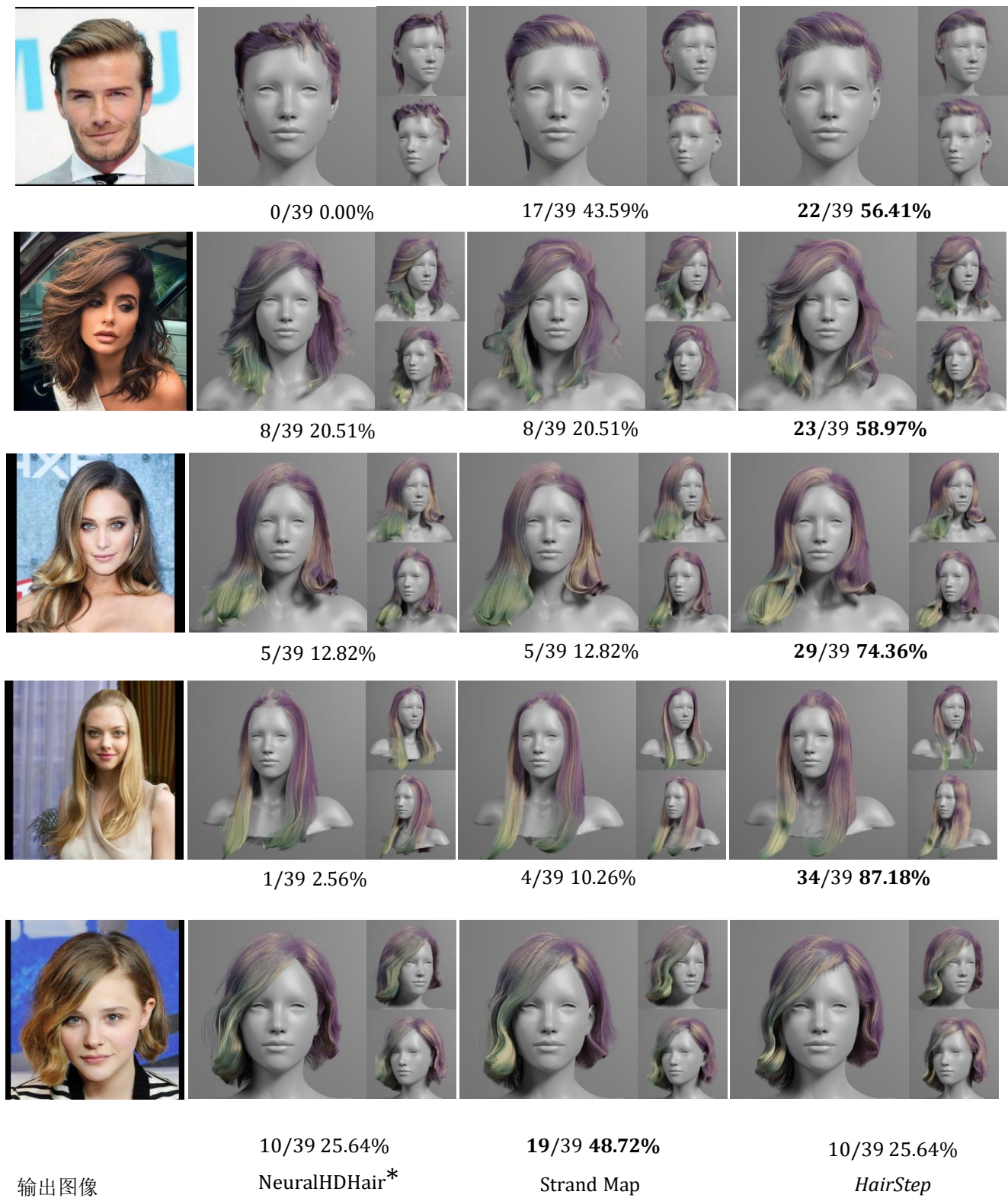


Figure S16 用户研究示例。从左到右分别为：输入图像，NeuralHdHair*的结果，使用发型图表示的结果，以及完整方法（Full）的结果。我们还为每个例子提供了 3 种不同表示的统计数据。

