R2P-SAM: Root-to-Prompt Hair Instance Segmentation with Geometric Priors under Point Supervision

Haigen Hu[1], Yang Yu[1], Chenquan Gong[1], Huihuang Zhang[1], Su Ruan[2], and Yiping Su[3](✉)

[1] Zhejiang University of Technology, Hangzhou, China
[2] Laboratoire AIMS, Equipe Quantif,Université de Rouen Normandie
[3] The First People's Hospital of Lin'an District
syp4846@163.com

Abstract. Precise instance segmentation of individual hair strands is fundamental to automated trichological analysis, yet it remains challenging due to their extreme thinness, frequent entanglement, and complex background interference. Existing weakly supervised approaches that rely on bounding-box annotations often suffer from geometric ambiguity, while the standard Segment Anything Model (SAM) exhibits severe over-segmentation when applied without task-specific guidance. To address these limitations, we propose R2P-SAM, a cost-effective framework that exploits sparse root annotations to guide a frozen SAM backbone. The core of our method is a geometric prior calibration mechanism that distills orientation and centrality fields to generate accurate point prompts, enabling effective separation of overlapping hair strands. Extensive experiments on three diverse datasets demonstrate that the proposed R2P-SAM consistently outperforms state-of-the-art weakly supervised methods, achieving superior boundary adherence while substantially improving annotation efficiency. The code is available at https://github.com/YYleafsheep/R2P-SAM.

Keywords: Hair Segmentation · SAM Adaptation · Weakly Supervised Learning

1 Introduction

Recent advances in artificial intelligence have enabled effective automated analysis of complex medical data [30]. Globally prevalent hair loss disorders, such as androgenetic alopecia, require precise quantification to support objective assessment. Trichoscopy provides a non-invasive modality to assess scarring and non-scarring alopecias [3,9], enabling the extraction of quantitative metrics such as hair density, shaft diameter and terminal-to-vellus ratios [8]. A recent review highlights artificial intelligence in microscopic hair imaging for scalp disorders, covering acquisition, structural analysis, and downstream analysis. [4]. Accurate estimation of these metrics requires instance-level segmentation of individual hair

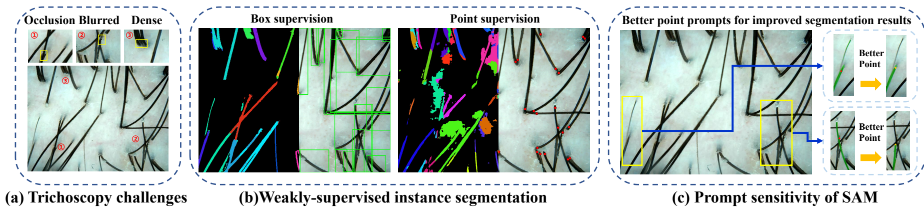


Fig. 1. Challenges in low-cost trichoscopy segmentation.

strands, which exceeds the capability of standard semantic segmentation. Existing methods [11,13,22,24] primarily focus on artifact removal or semantic segmentation using U-Net variants, which often merge adjacent strands and fail to separate instances. Hair strands are extremely thin, highly curved, and frequently overlapping (Fig. 1(a)). Although fully supervised frameworks such as Mask R-CNN [6] can address instance separation, they require dense, labor-intensive annotations. Weakly supervised bounding-box or point-based approaches often break strand continuity, producing fragmented predictions (Fig. 1(b)).

Recently, the Segment Anything Model (SAM) [10] introduced a promptable foundation model paradigm for general-purpose segmentation. Subsequent adaptations, such as MedSAM [20] and auto-prompting strategies [17,29], demonstrate that freezing the SAM backbone while customizing prompts can be effective in medical imaging scenarios. However, these approaches mainly target large, compact anatomical structures and remain ill-suited for objects with high-frequency topology. When applied to trichoscopic images, SAM is highly sensitive to domain shifts and prompt placement, often amplifying over-segmentation errors for thin, curvilinear, and densely entangled hair strands [2,19]. From a supervision perspective, point annotations offer a principled alternative to geometrically ambiguous bounding boxes for thin structures. In trichoscopy, the hair root serves as a unique anchor: it is spatially non-overlapping and remains identifiable even in regions of dense strand intersections. Nevertheless, hair roots are located at the structural extremities of elongated filaments, and naively prompting SAM at these marginal points introduces feature ambiguity due to insufficient local context. Moreover, such isolated prompts fail to constrain the global topology of hair strands in cluttered environments, resulting in incomplete or fragmented segmentations (Fig. 1(c)). Consequently, merely detecting hair roots is insufficient for effective SAM adaptation. Bridging this gap requires transforming sparse root signals into coherent, skeleton-aligned prompt trajectories that propagate structural priors along the strand, thereby enforcing topological continuity within the frozen backbone.

Motivated by these challenges, we propose R2P-SAM, a cost-effective framework that transforms sparse and unstable hair root annotations into high-fidelity hair strand instance masks. We freeze the SAM backbone and reformulate the instance segmentation task as a Root-to-Prompt generation task, enabling SAM adaptation with minimal annotation overhead. Our contributions are threefold:

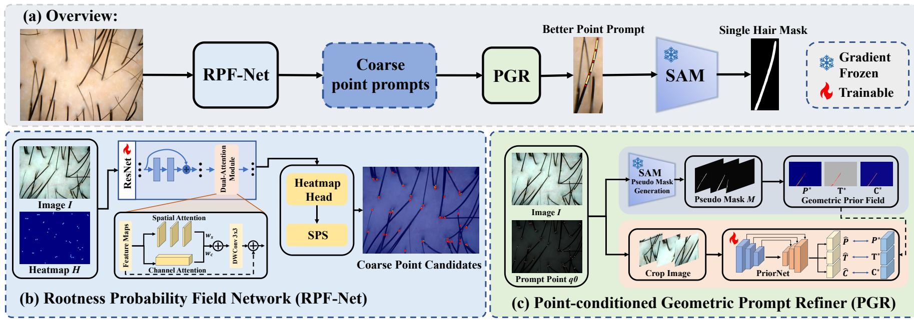


Fig. 2. R2P-SAM framework overview. (a) A two-stage pipeline combining root detection and geometric refinement feeds optimized prompts to a frozen SAM. (b) RPF-Net predicts rootness heatmaps via a ResNet encoder and dual-attention decoder. (c) PriorNet inputs the image and a Gaussian root query to predict geometric fields, guiding prompts to the stable centerline.

(i) We formulate hair strand instance segmentation as a Root-to-Prompt generation task, converting sparse root annotations into structurally meaningful prompts for frozen SAM. (ii) We propose a geometric calibration module that infers orientation- and centrality-aware priors to generate skeleton-aligned prompt trajectories, preserving topological continuity for thin, curvilinear structures in cluttered scenes. (iii) Extensive experiments on multiple trichoscopic datasets show that R2P-SAM outperforms state-of-the-art weakly supervised methods under dense strand overlap, while requiring only point-level annotations.

2 Method

Fig. 2 illustrates the overall framework of R2P-SAM, a coarse-to-fine pipeline for addressing prompt sensitivity. It comprises two stages: (1) Rootness Probability Field Learning for candidate detection, and (2) Geometric Prior-Guided Calibration for prompt alignment, enabling optimal queries to the frozen SAM.

2.1 Rootness Probability Field Learning

Trichoscopy hair strands are thin, curvilinear, and overlapping. Heuristic sampling based on edges or intensity often fails, drifting towards shafts or highlights rather than reliable roots. To mitigate this, we reformulate prompt generation as a density estimation problem.

Ground Truth Generation. Given an image $I \in \mathbb{R}^{H \times W \times 3}$ and sparse root annotations $P = \{p_k\}_{k=1}^{K}$, we construct a ground truth heatmap H^*. Unlike standard crowd counting that sums Gaussian responses, which leads to merged, flat responses in dense regions (Fig. 3(b)), we employ pixel-wise maximization

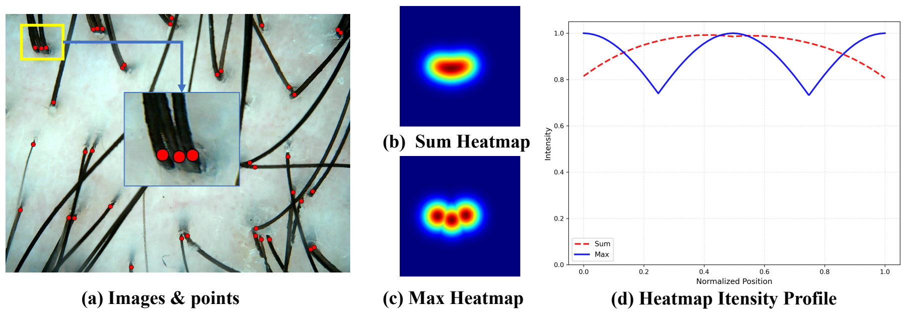


Fig. 3. Heatmap generation analysis. (a) Annotated input. (b) Summation merges overlapping spots, obscuring distinct peaks. (c) Maximization strategy preserves distinct peaks. (d) Intensity profiles confirm the maximization strategy separates adjacent roots, facilitating Spatial Peak Suppression.

to preserve peak distinctness for instance separation (Fig. 3(c)):

$$H^*(x) = \max_k \exp\left(-\frac{\|x - p_k\|_2^2}{2\sigma_g^2}\right), \tag{1}$$

where x denotes pixel coordinates. This ensures distinct topology even for adjacent roots (Fig. 3(d)).

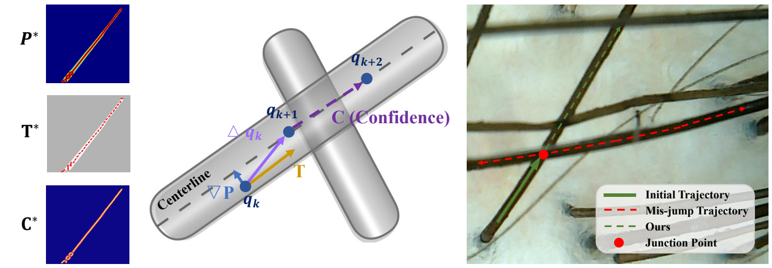
Network and Loss. To regress the pixel-wise probability field $\hat{H} = f_\theta(I)$ from raw images, we instantiate an encoder-decoder architecture termed ResNet50-HairNet. It utilizes a pre-trained ResNet50 to extract semantic features, coupled with a decoder to restore spatial resolution. To address class imbalance and mitigate false positives arising from strand crossings, optimizing a compound objective: $\mathcal{L}^{(H)} = \mathcal{L}_{reg} + \lambda_{shape}\mathcal{L}_{shape} + \lambda_{cnt}\mathcal{L}_{cnt}$. This objective combines focal-weighted regression $\mathcal{L}_{reg}$ for intensity matching, a shape consistency term $\mathcal{L}_{shape}$ penalizing non-root activations, and a global count constraint $\mathcal{L}_{cnt}$.

Inference via SPS. Instead of global thresholding, we propose Spatial Peak Suppression (SPS). SPS applies a sliding-window maximum filter to $\hat{H}$, retaining only local maxima that exceed a confidence threshold. This effectively disentangles adhered responses in dense areas, yielding discrete candidate points Q_{init}.

2.2 Geometric Prior-Guided Prompt Calibration

Initial points Q_{init} may deviate from the medial axis due to domain shifts. Direct prompting with off-center points in frozen SAM often results in mask merging. (Fig. 1(c)). To address this, we distill geometric knowledge from SAM's coarse pseudo-masks into a PriorNet, predicting geometric fields to guide prompts towards the stable centerline.

Geometric Prior Fields Definition. Let M be the pseudo-mask generated by SAM. We extract its skeleton S and define three supervision fields: *(1) Feasible*



(a) Geometric Prior Field (b) Mechanism of Field-Guided Update (c) Trajectory Comparison at Junctions

Fig. 4. Geometric prior-guided prompt calibration. (a) Defined geometric fields (P^*, T^*, C^*). (b) Schematic of the field-guided update mechanism. (c) Trajectory comparison: our confidence-weighted update (solid) successfully navigates junctions, whereas unweighted prompts (dotted) diverge.

Region P^* is a soft gate defining the valid centerline neighborhood: $P^*(x) = \exp(-d_S(x)^2/2\sigma_p^2) \cdot 1[x \in M]$, where $d_S(x)$ is the distance to S. *(2) Tangent Field* T^* resolves 180° sign ambiguity. We perform Topological Sorting on the skeleton graph $G = (V, E)$, executing BFS traversal from the root v_{root} to define a consistent root-to-tip flow t_s. *(3) Consistency Field* C^* weights direction reliability ($C^*(x) = \mathrm{prox}(x) \cdot \mathrm{coh}(x)$), where $\mathrm{prox}(x) = \exp(-d_S(x)^2/2\sigma_c^2) \cdot \mathbf{1}[x \in M]$, and $\mathrm{coh}(x) = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2 + \epsilon)$ is the linearity ratio of eigenvalues from local skeleton PCA, suppressed at junctions by $(1 - \gamma)$ (Fig. 4(a)).

PriorNet and Quality-Aware Loss. We train a lightweight U-Net (PriorNet) to predict these fields using point-conditioned input $X = \mathrm{Concat}(I, G(q_0))$, where $G(q_0)$ is a Gaussian heatmap of the query root that guides learning toward relevant geometry. Since pseudo-masks contain noise, introducing a soft isolation weight $q \in [0, 1]$ derived from geometric consistency. The training objective is:

$$\mathcal{L}_{prior} = q\left(\lambda_P \mathcal{L}_{BCE}(\hat{P}, P^*) + \lambda_C \mathcal{L}_{L1}(\hat{C}, C^*) + \lambda_T \mathcal{L}_c(\hat{T}, T^*) + \lambda_{neg}\mathcal{L}_{neg}\right) \quad (2)$$

Where $\mathcal{L}_c$ minimizes the cosine distance between predicted and ground-truth tangents, masked by region confidence, while $\mathcal{L}_{neg}$ suppresses activation at neighboring (negative) roots to enhance separability.

Field-Guided Prompt Refinement. During inference, we iteratively update prompt p_k using the predicted fields $(\hat{P}, \hat{T}, \hat{C})$. The update combines tangential flow and centerline attraction (Fig. 4(b)):

$$p_{k+1} = \mathit{\Pi}_{\mathcal{A}}\left(p_k + \eta\left(\underbrace{\hat{C}(p_k)\hat{T}(p_k)}_{\text{Reliable Flow}} + \beta \underbrace{\nabla \hat{P}(p_k)}_{\text{Centerline Attraction}}\right)\right) \quad (3)$$

where $\mathit{\Pi}_{\mathcal{A}}$ projects the point into the feasible region $\mathcal{A} = \{x \mid \hat{P}(x) > \tau\}$. As shown in Fig. 4(c), the consistency term $\hat{C}$ acts as a gate, dampening updates at ambiguous junctions to prevent mis-jump errors.

3 Experiments

3.1 Experimental Protocol

Datasets. We validated R2P-SAM on three datasets covering varying hair densities. *(1) Long-Hair Dataset (Private)*: Collected via a 50× trichoscopic magnification, comprising 400 images with a resolution of 2560 × 1920. It features dense, untrimmed long hair with frequent crossovers (3,942 instances, 12,575 root keypoints). *(2) FDU_HairFollicle* [5]: A public dataset with 1,652 images (1280 × 1024, 70×) of sparse trimmed scalp (avg. ∼12 strands/image). We re-annotated a random subset of 500 images with instance annotations for validation and test. *(3) HFD Dataset* [21]: Contains 711 medium-density images (640 × 640, avg. ∼18 strands/image). We randomly re-annotated 50 images with instance masks for quantitative evaluation.
Implementation Details. Implemented in PyTorch on an RTX A6000 GPU. Long-Hair images were resized to 1280 × 960, while FDU and HFD retained original resolutions. We split the Long-Hair dataset 6:1:1 (train/val/test), and the others 5:1:1. Models were trained for 80 epochs with a batch size of 8, using the AdamW optimizer with a cosine annealing schedule. Both learning rate and weight decay were set to 1×10^{-4} following a 10-epoch warm-up. Augmentations included flips and brightness changes.
Evaluation Metrics. We evaluate segmentation performance using *Precision (Prec.)*, *Recall (Rec.)*, and *F1-score (F1)*. For COCO-style instance segmentation evaluation, we adopt AP_{50} and AP_{75}, following the MS COCO evaluation protocol [16].Additionally, for the ablation study on root detection, we introduce *Mean Distance (Dist)* to measure the average Euclidean distance (in pixels) between predicted root points and ground truth.

3.2 Comparison with Weakly Supervised Approaches

We evaluated R2P-SAM against two categories: standard SAM (AMG mode)[10]; and weakly supervised methods utilizing box (DiscoBox [12], BoxLevelset [14], Box2Mask [15]) and point (SAPNet [28],P2Seg [27], PointSup [1]) supervision. Table 1 shows the results on the Long-Hair, FDU[5], and HFD[21] datasets.
Performance on Tangled Strands (Long-Hair). As shown in Table 1, traditional weakly supervised methods degrade significantly on dense, untrimmed data. Box-supervised approaches yield suboptimal AP_{50} due to the geometric ambiguity of boxes for thin structures, where single boxes often cover multiple intersecting strands. While point-supervised methods like PointSup alleviate this, R2P-SAM leads significantly, reaching 46.9% AP_{50} and surpassing PointSup by +10.7%. This is corroborated in the top row of Fig. 5, where our root-driven topology effectively disentangles overlapping instances better than box signals.
Robustness to Background Noise (FDU). The FDU results reveal limitations of standard prompts on sparse data. SAM suffers severe over-segmentation, with precision collapsing to 24.8%, as it misclassifies scalp textures without task-specific guidance. While PointSup maintains competitive recall, it struggles with

Table 1. Quantitative comparison across three datasets. Best results are in **bold**.

Method	Long-Hair				FDU				HFD			
	Prec.	Rec.	AP_{50}	AP_{75}	Prec.	Rec.	AP_{50}	AP_{75}	Prec.	Rec.	AP_{50}	AP_{75}
SAM (baseline)	0.732	0.281	0.238	0.184	0.248	**0.826**	0.360	0.237	0.406	0.495	0.270	0.152
DiscoBox [12]	0.198	0.488	0.260	0.043	0.082	0.494	0.230	0.070	0.086	0.467	0.240	0.056
BoxLevelset [14]	0.180	0.444	0.236	0.043	0.052	0.330	0.123	0.006	0.400	0.121	0.224	0.052
Box2Mask [15]	0.112	0.255	0.119	0.030	0.101	0.599	0.420	0.001	0.055	0.288	0.092	0.009
SAPNet [28]	0.461	0.350	0.271	0.156	0.590	0.410	0.326	0.182	0.398	0.121	0.135	0.018
P2Seg [27]	0.162	0.139	0.053	0.006	0.102	0.102	0.047	0.006	0.082	0.082	0.032	0.005
PointSup [1]	0.477	0.493	0.362	0.098	0.540	0.592	0.492	0.104	0.295	0.391	0.224	0.036
Ours	**0.740**	**0.612**	**0.469**	**0.235**	**0.751**	0.655	**0.500**	**0.310**	**0.730**	**0.500**	**0.401**	**0.158**

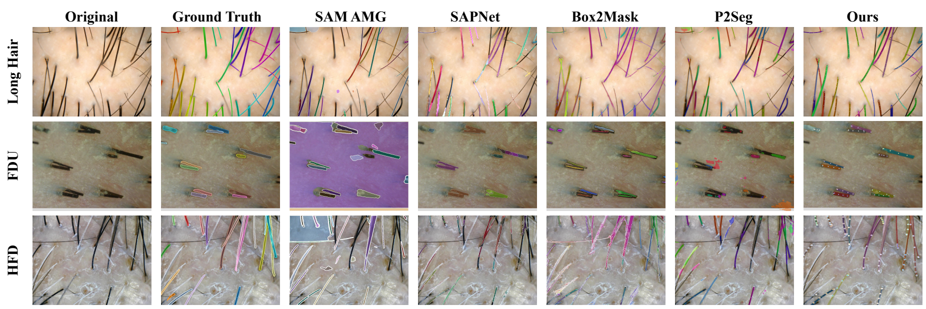


Fig. 5. Qualitative comparison across three datasets. R2P-SAM shows superior instance separation and boundary adherence over baselines.

54.0% precision, and methods like P2Seg fail to converge. Similarly, Box2Mask yields near-zero AP_{75} as box supervision fails to capture precise boundaries of thin strands. In contrast, R2P-SAM maintains 75.1% precision and 31.0% AP_{75}. Our framework outperforms the baselines, validating geometric prior calibration as essential for high-fidelity segmentation in noise.

Evaluation on Complex Structures (HFD). Finally, HFD results illustrate performance on highly curved structures. This dataset is challenging, evidenced by lower scores across all methods. Despite sample complexity, R2P-SAM maintains a lead with 40.1% AP_{50} over SAM's 27.0%. This relative improvement suggests that the inclusion of root-based topological constraints provides a stable guidance signal that appearance features alone cannot offer. As shown in the bottom of Fig. 5, our method successfully tracks intricate strand interactions and curvatures, whereas appearance-based baselines struggle to maintain continuity.

3.3 Ablation studies

Ablation of Backbone Architecture for RPF-Net. To determine the optimal feature extractor for RPF-Net, we benchmarked the six architectures detailed in Table 2. Quantitatively, ResNet50 [7] leads by achieving the highest F1-score of 0.795 and Recall of 0.823, alongside minimal localization error of 4.03

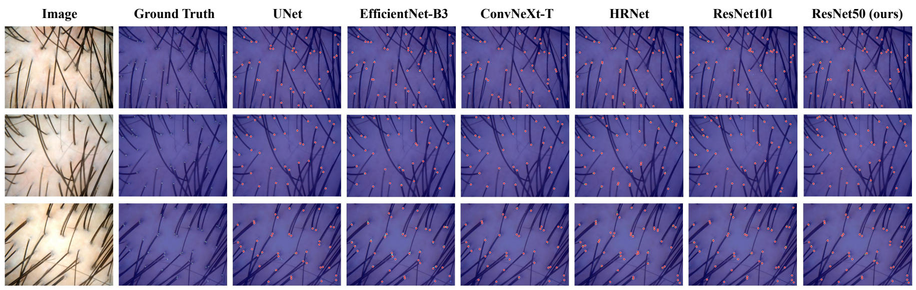


Fig. 6. Qualitative comparison of heatmap regressors. ResNet50 shows sharper peaks and better alignment with Ground Truth.

Table 2. Backbone comparisons.

Network	Prec.	Rec.	F1	Dist.
U-Net[23]	0.764	0.804	0.784	4.37
ResNet50[7]	0.770	**0.823**	**0.795**	**4.03**
EffNetB3[25]	**0.815**	0.750	0.781	6.28
ConvNeXt-T[18]	0.753	0.769	0.761	4.60
HRNet-W18[26]	0.701	0.733	0.717	6.72
ResNet101[7]	0.728	0.791	0.758	4.38

Table 3. Component ablations.

P	T	C	Prec.	Rec.	AP_{50}	AP_{75}
			0.666	0.602	0.415	0.188
✓			0.652	0.543	0.373	0.170
	✓		0.719	0.609	0.450	0.229
		✓	0.676	0.574	0.401	0.190
✓	✓		0.713	0.591	0.439	0.220
✓		✓	0.674	0.570	0.395	0.189
	✓	✓	0.732	**0.613**	0.465	0.234
✓	✓	✓	**0.740**	0.612	**0.469**	**0.235**

px. As visually corroborated in Fig. 6, this performance stems from its ability to resolve fine-grained features. While lightweight models like EfficientNet-B3 [25] achieve high precision by suppressing background noise, they suffer from a significant drop in recall to 0.750 due to missing thin strands. Conversely, earlier architectures like U-Net [23] produce diffuse heatmaps that degrade centroid accuracy. In contrast, ResNet50 strikes the best balance, generating compact, high-contrast Gaussian peaks that align with hair roots to ensure robust prompt generation. Consequently, it is adopted as our default backbone.

Ablation of Calibration Components. Table 3 analyzes the contributions of our geometric calibration modules: Feasible Region (P), Tangent Field (T), and Consistency Field (C). Using only P confines prompts near hair structures but yields suboptimal boundary adherence due to the lack of directional guidance. Integrating T boosts AP_{75} from 0.188 to 0.229. This confirms the necessity of orientation guidance. Without T, updates lack consistent root-to-tip flow and fail to traverse the centerline. Furthermore, C filters ambiguity at dense crossings, improving Precision from 0.732 to 0.740. Without this gating, updates may drift to intersecting strands at unreliable junctions. The full model achieves peak metrics with an AP_{75} of 0.235, confirming that combining spatial constraints, directional guidance, and junction reliability ensures precise calibration.

4 Conclusion

We propose R2P-SAM, a SAM-adapted framework for point-supervised hair strand instance segmentation. Root-driven prompting and geometric prior calibration help R2P-SAM reduce strand entanglement and background ambiguity, yielding consistent gains over weakly supervised baselines on three trichoscopic datasets, including public datasets. A main limitation is the lack of fully supervised upper-bound comparisons because dense annotations are costly. We will release test code, checkpoints, and selected annotated test samples for reproducibility. Future work will explore whether R2P-SAM can support downstream trichoscopic measurements, including hair density, shaft diameter, and terminal-to-vellus ratios, and further assess its utility for quantitative hair image analysis.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62373324 and U24A20242), in part by the Zhejiang Province Leading Geese Plan (Grant Nos. 2026C02A1221 and 2025C02160), and in part by the National Key Research and Development Program of China (Grant No. 2024YFC3306902).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cheng, B., Parkhi, O., Kirillov, A.: Pointly-supervised instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2617–2626 (2022)
2. Fan, K., Liang, L., Li, H., Situ, W., Zhao, W., Li, G.: Research on medical image segmentation based on sam and its future prospects. Bioengineering **12**(6), 608 (2025)
3. Fernández-Domper, L., Ballesteros-Redondo, M., Vañó-Galván, S.: [translated article] trichoscopy: an update. Actas dermo-sifiliograficas **114**(4), T327–T333 (2023)
4. Gong, C., Su, Y., Ruan, S., Hu, H.: Artificial intelligence in microscopic hair imaging for scalp disorders: From image acquisition to clinical decisions. Medical Image Analysis p. 104017 (2026)
5. Gu, X., Zhang, X., Fang, H., Wu, W., Lin, J., Yang, K.: A new dataset for hair follicle recognition and classification in robot-aided hair transplantation. In: International Conference on Artificial Neural Networks. pp. 38–49. Springer (2023)
6. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988 (2017)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Hoffmann, R.: Trichoscan: a novel tool for the analysis of hair growth in vivo. In: Journal of Investigative Dermatology Symposium Proceedings. vol. 8, pp. 109–115. Elsevier (2003)

9. Khare, S., Behera, B., Ding, D.D., Lallas, A., Chauhan, P., Enechukwu, N.A., Sławińska, M., Akay, B.N., Ankad, B.S., Bhat, Y.J., et al.: Dermoscopy of hair and scalp disorders (trichoscopy) in skin of color–a systematic review by the international dermoscopy society "imaging in skin of color" task force. Dermatology Practical & Conceptual **13**(4 Suppl 1), e2023310S (2023)
10. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
11. Lama, N., Kasmi, R., Hagerty, J.R., Stanley, R.J., Young, R., Miinch, J., Nepal, J., Nambisan, A., Stoecker, W.V.: Chimeranet: U-net for hair detection in dermoscopic skin lesion images. Journal of Digital Imaging **36**(2), 526–535 (2023)
12. Lan, S., Yu, Z., Choy, C., Radhakrishnan, S., Liu, G., Zhu, Y., Davis, L.S., Anandkumar, A.: Discobox: Weakly supervised instance segmentation and semantic correspondence from box supervision. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3406–3416 (2021)
13. Li, W., Raj, A.N.J., Tjahjadi, T., Zhuang, Z.: Digital hair removal by deep learning for skin lesion segmentation. Pattern Recognition **117**, 107994 (2021)
14. Li, W., Liu, W., Zhu, J., Cui, M., Hua, X.S., Zhang, L.: Box-supervised instance segmentation with level set evolution. In: European conference on computer vision. pp. 1–18. Springer (2022)
15. Li, W., Liu, W., Zhu, J., Cui, M., Yu, R., Hua, X., Zhang, L.: Box2mask: Box-supervised instance segmentation via level-set evolution. IEEE transactions on pattern analysis and machine intelligence **46**(7), 5157–5173 (2024)
16. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
17. Liu, X., Shi, G., Wang, R., Lai, Y., Zhang, J., Han, W., Lei, M., Li, M., Zhou, X., Wu, Y., et al.: Segment any tissue: One-shot reference guided training-free automatic point prompting for medical image segmentation. Medical Image Analysis **102**, 103550 (2025)
18. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)
19. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature communications **15**(1), 654 (2024)
20. Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y.: Segment anything model for medical image analysis: an experimental study. Medical Image Analysis **89**, 102918 (2023)
21. Mobal, M.: Hair follicle_indi dataset (apr 2024), visited on 2026-01-10
22. Nguyen, N.H., Lee, T.K., Atkins, M.S.: Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel. In: Medical Imaging 2010: Image Processing. vol. 7623, pp. 1436–1443. SPIE (2010)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
24. Talavera-Martinez, L., Bibiloni, P., Gonzalez-Hidalgo, M.: Hair segmentation and removal in dermoscopic images using deep learning. IEEE Access **9**, 2694–2704 (2020)
25. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)

26. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence **43**(10), 3349–3364 (2020)
27. Wang, Z., Yu, X., Han, X., Yu, W., Huang, Z., Jiao, J., Han, Z.: P2seg: Pointly-supervised segmentation via mutual distillation. arXiv preprint arXiv:2401.09709 (2024)
28. Wei, Z., Chen, P., Yu, X., Li, G., Jiao, J., Han, Z.: Semantic-aware sam for point-prompted instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3585–3594 (2024)
29. Xie, B., Tang, H., Cai, D., Yan, Y., Agam, G.: Self-prompt sam: Medical image segmentation via automatic prompt sam adaptation. arXiv preprint arXiv:2502.00630 (2025)
30. Zhang, H., Hu, H., Zhou, D., Cao, B., Zhang, X.: Feature diversity learning for a compact cnn module. IEEE Transactions on Emerging Topics in Computational Intelligence (2025)