

Article

ScaleNet: An Imaris XTension for Deep-Learning-Based Per-Scale Quantification of Immune Infiltration in Whole-Mount Vitiligo Mouse Skin

Wenxuan Gao ^{1,2}, Xuyang Jiang ³ and Yucheng Hu ^{4,*} ¹ College of Life Sciences, Beijing Normal University, Beijing 100875, China; wenxuan_gao@foxmail.com² National Institute of Biological Sciences, Beijing 102206, China³ School of Mathematical Sciences, Capital Normal University, Beijing 100048, China; jrc604817556@163.com⁴ Academy for Multidisciplinary Studies, Capital Normal University, Beijing 100048, China

* Correspondence: hydrays@cnu.edu.cn

Abstract

Quantifying the spatial distribution of immune cells within intact skin tissue is essential for understanding diseases such as vitiligo, in which CD8⁺ T cells selectively destroy epidermal melanocytes within the discrete, parallelogram-shaped epidermal compartments of mouse tail skin, which we term scales. Existing workflows rely on manual region drawing, which is labor-intensive and operator-dependent. Here we present ScaleNet, a three-stage deep-learning pipeline for automated per-scale quantification of whole-mount immunofluorescent images, implemented as an Imaris XTension to enable seamless integration with existing 3D imaging workflows. ScaleNet (i) encodes a 3D confocal volume as a pseudo-RGB projection that preserves height information lost by standard maximum-intensity projection, (ii) applies two independently trained Detectron2 Mask R-CNN models—one for epidermal scales and one for hair follicles—with sliced inference (SAHI) to segment whole-mount images at full resolution, and (iii) maps the resulting 2D mask back into the Imaris 3D coordinate system to quantify user-defined Spot objects per scale. Applied to vitiligo mice imaging, ScaleNet produced per-scale counts of CD8⁺ T cells and DCT⁺ melanocytes, enabling unbiased spatial statistics in the tail epidermis, demonstrating that ScaleNet can provide the quantitative spatial resolution needed to dissect the micro-anatomical dynamics of autoimmune depigmentation.

Keywords: vitiligo; whole-mount immunofluorescence; Imaris; deep learning; instance segmentation



Academic Editor: Christopher Robinson

Received: 15 May 2026

Revised: 3 July 2026

Accepted: 7 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In histological and immunological research, investigators frequently need to quantify specific cell populations within their native tissue microenvironments. In vitiligo—an autoimmune skin disease driven by CD8⁺ T-cell-mediated destruction of melanocytes [1–3]—the key readouts are the density and spatial distribution of infiltrating effector T cells [4–6] and the extent of melanocyte loss [7,8], both of which must be measured within the anatomically relevant units of the skin [9,10].

In adult mouse tail epidermis, the interfollicular skin is organized into a regular, tessellated pattern of two interlocking compartments: pigmented, parallelogram-shaped scale regions (hereafter “scales” refers specifically to these epidermal compartments, not to image magnification or clinical instruments such as the Vitiligo Impact Scale [VIS-22])

and the thinner, unpigmented interscale regions that border them [11,12]. Each scale is associated with a triplet of hair follicles at its posterior margin, and the two compartments differ in keratin expression, proliferation kinetics, and stem-cell lineage [11,12]. Because melanocytes—and the autoreactive CD8⁺ T cells that target them in vitiligo—are concentrated within the scale compartment, per-scale quantification is the biologically meaningful unit at which to measure depigmentation and immune infiltration, rather than merely a convenient spatial grid. The pipeline described below targets the scale compartment specifically, subtracting hair-follicle openings from the detected mask (Section 2.7).

Whole-mount immunofluorescent imaging of the isolated epidermis combined with confocal microscopy [9,10,13] provides the necessary 3D structural information, but extracting per-scale quantitative data from the resulting volumes (typically $> 3500 \times 3000 \times 30$ voxels) is a bottleneck. A rich ecosystem of open-source bioimage analysis tools—including Fiji/ImageJ v20250302.2217 (National Institutes of Health, Bethesda, MD, USA) [14,15], CellProfiler v4.2.5 (Broad Institute, Cambridge, MA, USA) [16], ilastik v1.4.1 (Heidelberg University, Heidelberg, Germany) [17], and QuPath v0.6.0 (University of Edinburgh, Edinburgh, UK) [18]—and recent deep-learning segmentation frameworks such as U-Net [19,20], Cellpose [21,22], StarDist [23], and Mesmer [24] now achieve near-human accuracy on a wide range of cellular and tissue segmentation tasks [25,26], but none provides a turnkey solution for segmenting the scale-shaped anatomical units of mouse tail epidermis from whole-mount confocal stacks. Commercial platforms such as Imaris v10.1 (Oxford Instruments, Abingdon, UK) [27] offer excellent 3D rendering and interactive analysis, but do not natively integrate modern deep-learning instance segmentation, and manually drawing dozens of scale boundaries per image is labor-intensive and operator-dependent. Global intensity thresholding cannot separate adjacent scales because the scale boundaries are defined by sparse follicular landmarks rather than by differences in channel intensity.

Applying 2D deep-learning segmentation models to this problem raises two specific technical challenges. First, standard maximum-intensity projection (MIP) collapses the Z-axis into a single plane and discards the height information that distinguishes concave follicular openings from convex scale surfaces—precisely the morphological cue a model needs to delineate scale boundaries. Second, whole-mount tail epidermis images acquired at $20\times$ typically exceed 3500×3000 pixels, far larger than the 512×512 tile size on which contemporary segmentation backbones are trained.

Here we present ScaleNet, an Imaris XTension that addresses both challenges. ScaleNet contributes: (i) a pseudo-RGB projection that encodes the MIP, an L2-norm intensity projection, and a local height-deviation map into the three channels of a 2D RGB image, preserving 3D morphological context without modifying the model architecture; (ii) a dual-model segmentation strategy with independently trained Detectron2 [28] Mask R-CNN [29] networks (built on Faster R-CNN [30]) for scales and hair follicles, combined with Sliced Aided Hyper Inference (SAHI) [31] for large images and with morphological post-processing that yields non-overlapping per-scale masks; and (iii) a round-trip mapping that writes the 2D mask back into the Imaris 3D dataset so that user-defined Spot objects (e.g., CD8⁺ T cells, DCT⁺ melanocytes) can be counted per scale using native Imaris tools [27] and exported to a structured CSV table. The complete workflow is summarized in Figure 1.

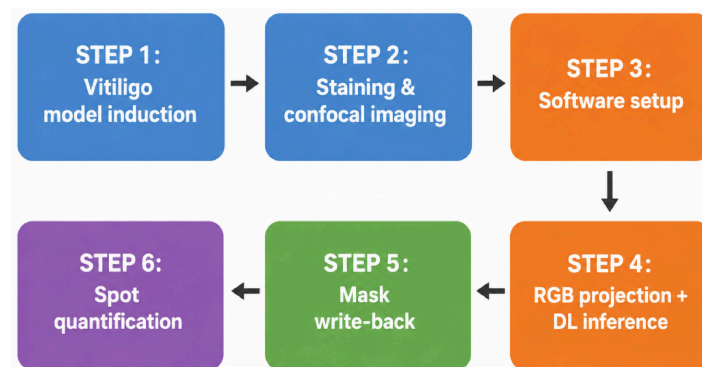


Figure 1. Overview of the ScaleNet workflow. The pipeline comprises six stages: vitiligo model induction and whole-mount staining (Sections 2.1 and 2.2), confocal imaging and Imaris export (Section 2.3), software and environment setup (Section 2.4), pseudo-RGB projection and deep-learning-based mask generation (Sections 2.5–2.7), mask write-back to the Imaris dataset (Section 2.8), and per-scale spot quantification (Section 2.9). Blue: wet-lab steps; orange: computational steps; green: Imaris write-back; purple: quantification output.

2. Materials and Methods

2.1. Mice and Vitiligo Induction

Mice were bred and kept in the specific-pathogen-free (SPF) animal facility of the National Institute of Biological Sciences (NIBS). All animal experimental procedures were performed in strict accordance with the NIBS Guide for the Care and Use of Laboratory Animals and approved by the Laboratory Animal Management Committee of NIBS, fully complying with relevant ethical guidelines and regulations. All mice used in experiments were socially housed under a 12 h light–dark cycle with free access to food and water, at 23–25 °C and 50–56% humidity.

Vitiligo was induced in mice as described previously [6,9,32]. On day 0, mice were intradermally injected with 2×10^5 B16F10 melanoma cells (ATCC, cat. CRL-6475, Manassas, VA, USA) into the right flank. On days 4 and 10, 10 mg/kg anti-mouse CD4 depletion antibody (clone GK1.5; Bio X Cell, cat. BE0003-1, Lebanon, NH, USA) was administered intraperitoneally to ablate CD4⁺ regulatory T cells. On day 12, the established tumors were surgically excised and the incisions closed with suture clips. To evaluate vitiligo induction, whole-mount immunofluorescent staining of tail skin was performed on day 33 to visualize epidermal melanocyte loss and CD8⁺ T cell infiltration.

2.2. Whole-Mount Immunofluorescent Staining

Whole-mount staining was performed essentially as previously described [10]. Hair shafts were first removed from dissected tails using Nair depilatory cream (Church & Dwight, Ewing, NJ, USA). Tail skin was collected, flattened, and cut into 1×1 cm squares. The tissue was incubated in 20 mM EDTA in PBS at 37 °C with shaking at 80 rpm for 1 h. The epidermis was peeled from the dermis in a posterior-to-anterior direction to retain hair follicles in the dermal layer; any residual follicles adhering to the epidermis were removed manually with fine-tipped forceps. The isolated epidermis was flattened, fixed in 4% paraformaldehyde in PBS (Sigma-Aldrich, cat. 158127, St. Louis, MO, USA) for 10 min, and washed three times for 15 min each in PBS. Samples were permeabilized in 0.3% hydrogen peroxide in methanol at −20 °C for 20 min, then blocked for 1 h at room temperature in PBS containing 2% normal donkey serum (Jackson ImmunoResearch, cat. 017-000-121, West Grove, PA, USA), 1% BSA (Sigma-Aldrich, cat. V900933, St. Louis, MO, USA), and 0.3% Triton X-100 (Sigma-Aldrich, cat. X100, St. Louis, MO, USA).

Samples were incubated overnight at 4 °C with primary antibodies: anti-mouse DCT (kindly provided by the laboratory of Ting Chen; 1:3000) and anti-mouse CD8a (clone

53-6.7; BioLegend, cat. 100712, San Diego, CA, USA; 1:300). Following primary incubation, samples were incubated with species-appropriate fluorescent secondary antibodies (Jackson ImmunoResearch, cat. 711-545-152/712-605-153, West Grove, PA, USA; 1:500) at room temperature for 1 h.

2.3. Confocal Imaging

Stained whole-mount epidermis was imaged on a Nikon AX confocal microscope with a 20× objective (Nikon Instruments Inc., Melville, NY, USA). 3D z-stacks spanning the full epidermal thickness were acquired and exported in .ims format for subsequent analysis in Imaris v10.1 (Oxford Instruments, Abingdon, UK).

2.4. ScaleNet Pipeline Overview and Installation

ScaleNet is implemented as three Imaris XTension Python scripts—XTScaleNetA.py, XTScaleNetB.py, and XTScaleNetC.py—that execute sequentially on an open .ims dataset. The three scripts and an accompanying ScaleNet directory (containing the deep-learning inference code and pretrained model weights) are placed in the Imaris XTensions directory; the paths to the Python interpreter and the XTensions directory are registered in File → Preferences → CustomTools in Imaris. The deep-learning code is run in an isolated Conda environment located at a fixed relative path .conda inside the plugin directory, so that no absolute paths need to be configured at install time. Core dependencies (numpy, opencv-python, pillow, scipy, sahi, torch, torchvision, detectron2) are installed with pip; the PyTorch v1.13 (PyTorch Foundation, San Francisco, CA, USA) and Detectron2 versions must match the local CUDA Toolkit v11.8 (NVIDIA Corporation, Santa Clara, CA, USA).

2.5. Pseudo-RGB Projection (XTScaleNetA)

To overcome the loss of 3D spatial information inherent in traditional maximum-intensity projection (MIP) while remaining compatible with standard 2D deep-learning instance segmentation frameworks (which expect 3-channel RGB inputs), XTScaleNetA.py reads the full 3D stack from the DAPI channel and synthesizes a pseudo-RGB projection (*_RGB.png). The three channels are constructed as follows (Figure 2):

- Red (R): the standard MIP along the Z-axis, representing the maximum per-pixel intensity.
- Green (G): the L2-norm of the intensity along the Z-axis—the square root of the sum of squared intensities across all Z-slices—reflecting the overall tissue signal magnitude.
- Blue (B): a local height-deviation map, computed as the per-pixel difference between a large-radius Gaussian-blurred version of the argmax height map and the raw argmax height map. This channel encodes whether a pixel sits above or below its local baseline, highlighting the concavity of follicular openings and the convex curvature of scale interiors.

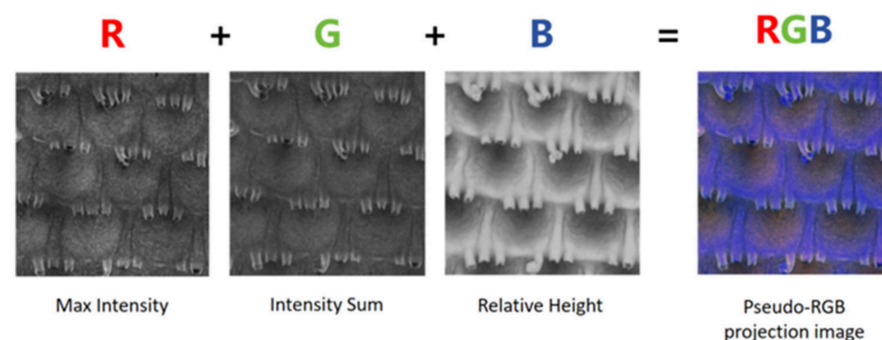


Figure 2. pseudo-RGB projection encoding 3D tissue morphology in a 2D image. (Left to right) Red channel: maximum intensity projection along the Z-axis. Green channel: L2-norm (square root of

summed squared intensities) across Z-slices. Blue channel: local height deviation, computed as the large-radius Gaussian blur of the argmax height map minus the raw height map. The composite RGB image (right) retains follicular and scale boundary information required for accurate deep-learning inference.

2.6. Deep-Learning Models: Training and Architecture

Two independent instance segmentation models were fine-tuned using Detectron2 [28] with a Mask R-CNN [29] head (built on Faster R-CNN [30]) on a ResNet-50-FPN backbone initialized from weights pre-trained on the MS COCO dataset [33]: one model for epidermal scale segmentation (model_scale.pth) and one for hair follicle (bulb) segmentation (model_bulb.pth). Fine-tuning from COCO pre-trained weights means the backbone already encodes general visual features—object boundaries, shape primitives, texture gradients—learned from over 100,000 diverse images; the tissue-specific fine-tuning stage adapts these priors to the morphology of mouse-tail epidermal scales and follicles rather than learning visual representations from scratch. Consequently, the amount of domain-specific annotated data required is substantially lower than for training from scratch, and the model is less susceptible to memorizing the particular imaging style of the training mice. Full-field confocal images were tiled into 522×522 px crops using the same SAHI sliding-window strategy applied at inference, yielding annotated tiles that directly match the model's input format. Annotations were performed manually on these tiles using the open-source CVAT annotation tool. For the scale model, 170 annotated tiles were obtained from 12 fields of view across three mice, comprising 1069 scale instances in total; 135 tiles (810 instances) were assigned to the training set and 35 tiles (259 instances) to the test set. For the bulb model, 181 tiles from the same 12 fields of view were annotated with three classes—bulge (hair follicle opening), stick (hair shaft), and unknown—comprising 3164 bulge instances in total; 145 tiles (2523 bulge instances) were used for training and 36 tiles (641 bulge instances) for testing. Only the bulge class is used by the inference pipeline; stick and unknown regions are subtracted from scale masks alongside bulge regions and do not affect per-scale quantification. The scale and bulb models used different held-out fields of view as test sets to assess generalization within each model. Because fine-tuning from COCO pre-trained weights substantially reduces the risk of mouse-specific style overfitting, an image-level field-of-view split was deemed appropriate. Fine-tuning ran for 3000 iterations on an NVIDIA GeForce RTX 3090 GPU. The full configuration is provided in config_scale.yaml and config_bulb.yaml (GitHub repository); key settings were: base learning rate 0.0025, batch size 4 images per iteration, momentum 0.9, weight decay 0.0001. The first two backbone stages were kept frozen (FREEZE_AT = 2), preserving the low-level COCO-derived features and restricting adaptation to the higher-level feature layers. The learning rate followed a WarmupMultiStepLR schedule: linear warm-up over the first 1000 iterations (initial factor 0.001), then decay by 0.1 at iterations 1500 and 2500. Data augmentation comprised random horizontal flip, random crop to 90% of the image (relative range), and multi-scale resizing with the short side randomly sampled from {400, 500, 600} pixels per iteration; no additional offline augmentation was applied. The random seed was not fixed (SEED = -1); reproducibility is ensured through the publicly available configuration files and pretrained weights rather than a fixed seed. Detection confidence thresholds were set to 0.8 for both models; this value was determined by empirical inspection on held-out training images and hardcoded in the inference script (run_scalenet.py) prior to any evaluation on the test set. All models were implemented in PyTorch [34].

2.7. Tiled Inference (Sahi) and Mask Post-Processing

Because whole-mount images typically exceed 3500×3000 pixels—far larger than the 512×512 tiles on which the models were trained—the pipeline performs tiled inference using SAHI [31]. The input image is partitioned into overlapping 512×512 crops; each crop is inferred independently; and the resulting instance masks are merged with non-maximum suppression. The scale model output is converted into a soft distance-based mask using dual thresholds (foreground 0.30, boundary 0.12).

Two post-processing steps then isolate clean per-scale regions. First, because hair follicles frequently contain non-specific CD8⁺ T cell signals that would otherwise contaminate scale quantification, the hair follicle mask is subtracted from the scale mask (scale – bulb). Second, an area-based filter discards connected components whose area deviates by more than two standard deviations (2σ) from the mean scale area—this removes fragmented or clipped scales near the image border. The filtered mask is then subjected to morphological erosion followed by dilation (implemented in OpenCV [35]) to guarantee that each scale is a fully separated connected component (Figure 3). The final mask is written to *_RGB_final_mask.png in the output directory.

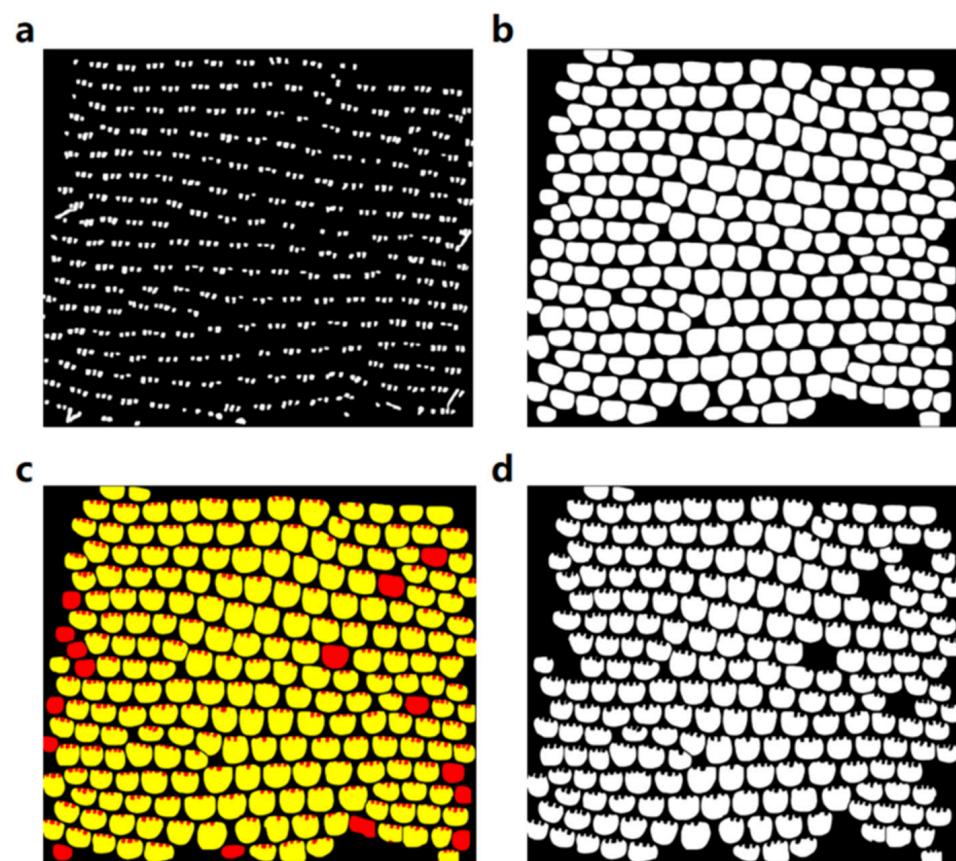


Figure 3. Mask post-processing pipeline. (a) Hair follicle (bulb) mask from the bulb segmentation model. (b) Raw scale mask from the scale segmentation model. (c) Scale mask after subtraction of the bulb mask (scale – bulb). (d) Final mask after 2σ area filtering followed by morphological erosion and dilation (saved as *_RGB_final_mask.png); yellow/red highlighted regions indicate connected components removed because their area falls below the mean minus two standard deviations.

2.8. Mask Write-Back to the Imaris Dataset (XTScaleNetB)

XTScaleNetB.py loads *_RGB_final_mask.png, flips the Y-axis to match the Imaris coordinate convention, and writes the mask into the bottom Z-slice of Channel 0 of the current dataset as a 16-bit image with mask intensity 60,000 (Figure 4a). The user then

creates Surface objects from Channel 0 using the Imaris Surfaces wizard, and applies the native Split tool to separate the mask into per-scale Surfaces (Figure 4b). Because the mask has already been topologically cleaned (Section 2.7), each scale becomes an independently addressable Surface that can be manually reviewed or deleted.

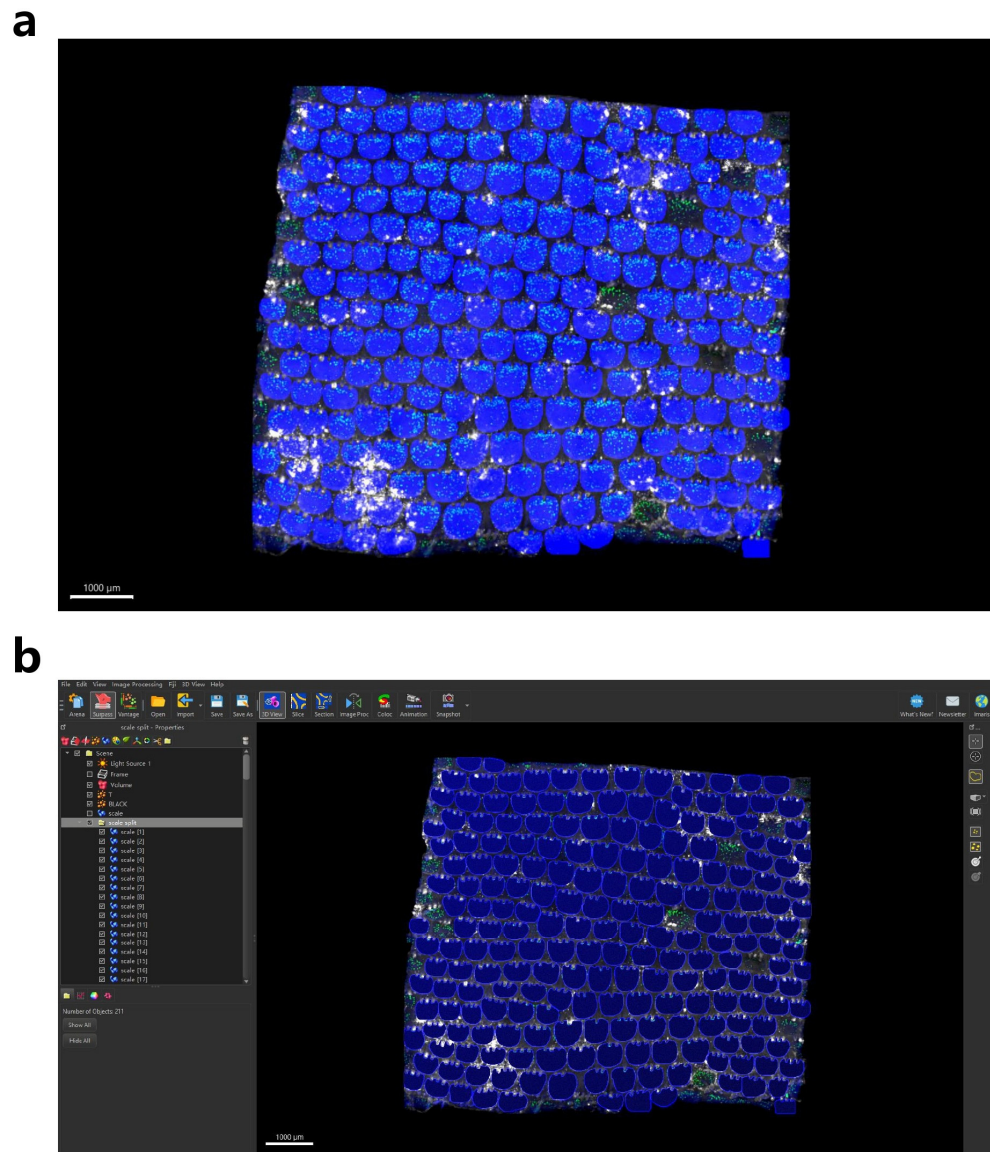


Figure 4. Mask write-back and surface splitting in Imaris. (a) The final mask is imported by XTScaleNetB and written as 16-bit pixel values (intensity = 60,000) into the bottom Z-slice of Channel 0, confirming correct spatial alignment within the Imaris volume. (b) Surface objects generated from the mask using the Surfaces creation wizard followed by the native Split tool; each connected scale region becomes an individually addressable Surface.

2.9. Per-Scale Spot Quantification (XTScaleNetC)

Target cell populations—for example, CD8⁺ T cells and DCT⁺ melanocytes—are detected in Imaris as Spot objects using the standard Spots wizard prior to running XTScaleNetC.py. With one Surface object (or a folder containing multiple Surfaces) and one or more target Spot objects selected in the Surpass tree, XTScaleNetC.py maps each Spot's physical XYZ coordinates to voxel indices using the dataset extent and voxel size. For each selected Surface it generates a 3D binary voxel mask and performs 2D connected-components labelling (OpenCV) on the projected mask to partition the Surface into individual scale Parts. The number of Spots of each category falling within each Part is

counted, and the results are exported as a CSV table with columns Surface_Name, Part_Id, Part_Voxels, Surface_Center_X, Surface_Center_Y, and one count column per Spot category (Figure 5).

Surface_Name	Part_Voxels	Surface_Center_X	Surface_Center_Y	black	T
scale 1	19,868	764.550080531508	813.6915139923495	119	17
scale 2	21,189	555.8076360375667	840.6604842135071	101	1
scale 3	20,637	646.777244754567	675.587827688133	127	4
scale 4	20,334	750.9800334415265	514.8834956230943	117	16

Figure 5. Example CSV output from XTScaleNetC. Each row represents one scale region (Part), reporting the parent Surface name, voxel count (Part_Voxels), centroid coordinates (Surface_Center_X, Surface_Center_Y), and the count of each Spot category within that region.

2.10. Model Evaluation and Runtime Benchmarking

Scale and bulb segmentation models were evaluated on the held-out test set using standard COCO-style instance segmentation metrics: mask average precision averaged across IoU thresholds 0.50–0.95 (AP), and at fixed thresholds of 0.50 (AP₅₀) and 0.75 (AP₇₅). End-to-end pipeline runtime was measured on a representative whole-mount image (3557 × 3122 pixels) on an NVIDIA GeForce RTX 3090 GPU and, for comparison, on an AMD Ryzen 9 5900X CPU.

3. Results

3.1. Pseudo-RGB Projection Preserves 3D Morphology for 2D Segmentation

Whole-mount staining of day-33 vitiligo mouse tail epidermis revealed extensive loss of DCT⁺ melanocytes and dense CD8⁺ T cell infiltration across the scale pattern, in clear contrast to control animals in which melanocytes were intact and T cell numbers were low (Figure 6). The pseudo-RGB projection constructed in Section 2.5 successfully compressed this 3D structure into a 2D image in which scale boundaries are visually resolved: follicle openings appear as dark concavities surrounded by bright rings in the blue height-deviation channel, and the combination of MIP (red), L2-norm (green), and height deviation (blue) encodes the complementary features needed for instance segmentation (Figure 2, right).

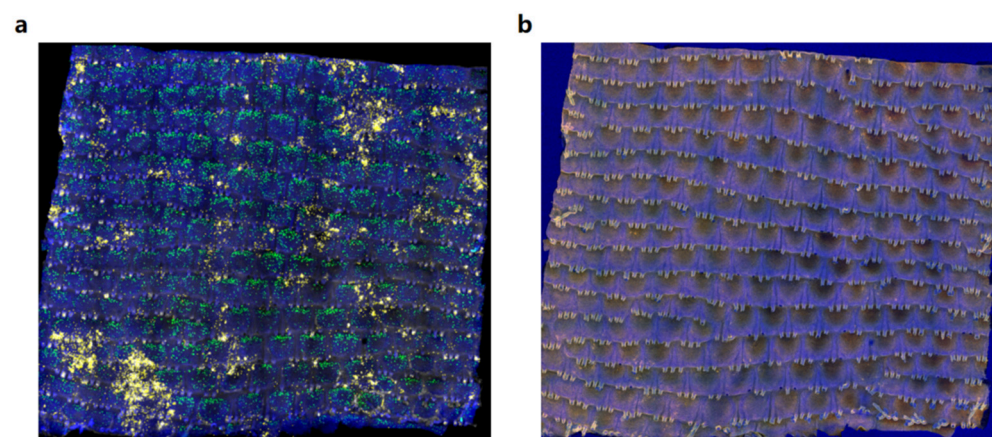


Figure 6. Whole-mount immunofluorescent imaging of vitiligo mouse tail epidermis. Representative confocal images (20× objective, Nikon AX) acquired on day 33. (a) Control mouse with intact DCT⁺ melanocytes and minimal CD8⁺ T cell infiltration. (b) Vitiligo mouse showing extensive melanocyte loss and dense CD8⁺ T cell infiltration across the epidermal scale pattern.

3.2. Scale and Bulb Segmentation Accuracy

On the held-out test set, the scale segmentation model achieved a mask AP of 75.3%, an AP₅₀ of 97.6%, and an AP₇₅ of 94.2%, indicating that detected scale instances match ground-truth annotations with high boundary fidelity. The bulb (hair follicle) segmentation model achieved a mask AP of 65.4%, an AP₅₀ of 97.7%, and an AP₇₅ of 80.2%. The lower overall AP of the bulb model is consistent with the smaller size and weaker boundary contrast of individual follicle openings relative to full scales, while the AP₅₀ of 97.7% indicates that follicles are reliably detected even when boundary localization is less precise—a regime that is acceptable for the pipeline, because the follicle masks are used to subtract contaminating regions from scale masks rather than as a primary quantitative output. We acknowledge that AP estimates from a small test set may be sensitive to the particular split; this is noted as a limitation and would be resolved by expanding the annotated dataset.

3.3. Tiled Inference Runtime on Whole-Mount Images

End-to-end inference on a representative 3557×3122 -pixel whole-mount image completed in 6.7 s on an NVIDIA GeForce RTX 3090 GPU and in 3.78 min on an AMD Ryzen 9 5900X CPU. Because SAHI tiling parallelizes across the 512×512 crops of the input, runtime scales approximately linearly with image area and is dominated by the Mask R-CNN forward pass; no out-of-memory failures were observed on the tested GPU for images up to ~42 megapixels.

3.4. Mask Write-Back and Surface Splitting Reproduce Scale Geometry in Imaris

Running XTScaleNetB.py successfully inserted the final mask into the bottom Z-slice of Channel 0 at intensity 60,000 (Figure 4a), and the Imaris Surfaces Split tool partitioned the imported mask into per-scale Surface objects without manual boundary editing (Figure 4b). Because the mask was topologically cleaned upstream by morphological erosion and dilation (Figure 3d), adjacent scales were never merged into a single Surface. Scales that were clipped at the image border or otherwise incompletely segmented were easily deleted by hand from the Surpass tree prior to quantification.

3.5. Per-Scale Quantification of CD8⁺ T Cells and DCT⁺ Melanocytes in Day-33 Vitiligo Mice

XTScaleNetC.py produced, for each whole-mount dataset, a CSV table in which every row corresponds to a single scale region (Part) and reports the parent Surface name, voxel count (Part_Voxels), Surface centroid coordinates, and the count of each Spot category within that region (Figure 5). In day-33 vitiligo animals, CD8⁺ T cells were distributed across essentially all segmented scales, with per-scale counts substantially above control baselines, while DCT⁺ melanocyte counts were markedly reduced—a pattern consistent with the whole-mount images (Figure 6) and now quantifiable at single-scale resolution across the full tail epidermis. Regions with zero Spots in all categories were confined to the image periphery where scale coverage was partial, as expected.

4. Discussion

ScaleNet addresses a specific bottleneck in the analysis of whole-mount vitiligo skin: the transition from 3D fluorescent volumes to per-scale quantitative tables. In contrast to general-purpose bioimage analysis pipelines such as CellProfiler [16], ilastik [17], and QuPath [18]—which lack a tissue-specific scale segmentation model—ScaleNet combines a geometrically motivated pseudo-RGB projection with two independently trained Mask R-CNN models [28,29] and SAHI-based tiled inference [31] to achieve high instance segmentation accuracy (scale AP₅₀ 97.6%, bulb AP₅₀ 97.7%) on images far larger than the model input size, while keeping end-to-end inference under ten seconds per image on

a single GPU. Closing the loop through Imaris—by writing the mask back into the 3D dataset and building Surface objects—lets the user perform per-scale Spot counts using the platform’s native, well-validated machinery rather than a bespoke coordinate system, and preserves the option for manual curation before quantification.

Several aspects of the design were chosen deliberately. First, training two single-class models rather than one multi-class model keeps each model’s loss surface well-conditioned on a modest training set (175 images, 72 scale and 597 bulb instances) and lets the scale and bulb thresholds be tuned independently for downstream subtraction. Second, encoding height information in the blue channel of a pseudo-RGB image (Section 2.5) preserves the three-channel input format expected by pretrained Detectron2 backbones, so ImageNet-initialized weights remain usable; an alternative would be to retrain a one-channel model on MIP alone, but preliminary experiments confirmed that follicle-versus-scale discrimination degrades when height cues are removed. Third, the 2σ area filter and erosion/dilation steps are essential for making Imaris’ native Split tool produce per-scale Surfaces without manual intervention; without them, adjacent scales tend to merge into a single Surface.

The workflow imposes a small number of operational requirements. §2.5 (XTScaleNetA) and §2.8 (XTScaleNetB) must be run consecutively on the same dataset without resolution changes or cropping, because the mask is written back at the dataset’s native voxel grid. Before §2.9 (XTScaleNetC), both Surfaces and Spots must be selected in the Surpass tree; omitting either selection is the most common runtime error in our experience.

Limitations and generalizability. The segmentation models were fine-tuned on images from three mice imaged at $20\times$ on a Nikon AX confocal microscope; cross-mouse generalizability has not been formally tested (a mouse-level holdout was infeasible at $n = 3$) and is an acknowledged limitation; application to other tissues, species, magnifications, or imaging systems will require retraining or fine-tuning on representative annotated data, following workflows established for domain-specific bioimage models [22,24,26]. As a practical guide, successful transfer to a new tissue type will generally require: (1) confocal z-stacks with sufficient axial resolution to distinguish anatomical subunits in the pseudo-RGB height channel; (2) a minimum of ~50–100 annotated instances per target class; and (3) re-tuning of the detection threshold and the GAUSSIAN_RADIUS parameter to match the spatial scale of the new tissue. The pseudo-RGB projection assumes that the tissue of interest occupies a single focal plane, so strongly curved or multi-layered samples may require adjustment of the Gaussian-blur radius in the blue channel (controlled by the GAUSSIAN_RADIUS constant in run_scalenet.py). The plugin has been validated on Imaris v10.1 (Oxford Instruments, Abingdon, UK); compatibility with substantially older or future versions of the Imaris XTension API is not guaranteed. Finally, the bulb model’s moderate AP75 (80.2%) indicates that follicle boundary localization is imperfect—this is acceptable for subtraction-based scale cleanup but may warrant retraining if follicles themselves become the primary quantitative target.

A formal manual-versus-automated count comparison was not performed, for two reasons. First, the AP metrics in Section 3.2 already provide a direct human–AI comparison at the scale masking step—ScaleNet’s primary contribution—since ground-truth masks were drawn by a human annotator and AP quantifies model agreement with those boundaries. Second, per-scale cell counts depend not only on scale masks but also on the user’s Spots detection settings in Imaris (threshold, estimated cell diameter, background subtraction); different operators applying different parameters to the same dataset can produce substantially different counts independent of mask quality. Because cell identification is explicitly outside the scope of ScaleNet—the pipeline accepts any user-defined Spot objects as input—a composite count comparison would conflate masking accuracy with operator-level variability in Spot detection and would be difficult to interpret.

5. Conclusions

ScaleNet demonstrates that per-scale quantification of whole-mount vitiligo skin can be automated by pairing a 3D-preserving pseudo-RGB projection with dual Mask R-CNN segmentation models, and that the result can be round-tripped back into Imaris for 3D-aware cell counting using the platform's native tools. The workflow produces a reproducible, per-scale CSV table of immune infiltration and melanocyte loss across the full tail epidermis, replacing manual region drawing and global thresholding with a largely automated, openly implemented pipeline (Spot object creation and optional manual scale curation remain user-directed steps). The same strategy—lossy 2D projection augmented with complementary height information, plus dual-model instance segmentation with SAHI tiling—should be adaptable to other whole-mount tissue analyses in which anatomical subunits must be resolved at scale.

Author Contributions: Conceptualization, Y.H. and W.G.; methodology, Y.H., X.J. and W.G.; software, Y.H. and X.J.; writing—review and editing, Y.H., X.J. and W.G.; funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant number 12271375) and Beijing Natural Science Foundation (grant number Z210003).

Institutional Review Board Statement: All animal procedures were approved by the Laboratory Animal Management Committee of NIBS (approval code: NIBS2025M0016 and approval date: 3 October 2025) and were in compliance with all relevant ethical regulations.

Informed Consent Statement: Not applicable.

Data Availability Statement: The ScaleNet source code, pretrained model weights (model_scale.pth, model_bulb.pth), configuration files, and example outputs are openly available at https://github.com/hydrays/ScaleNet_plugin (accessed on 2 July 2026). The annotated training and test tiles, train/test split files (split_scale.json, split_bulb.json), Detectron2 configuration files (config_scale.yaml, config_bulb.yaml), Conda environment file, and an inference example are openly available at <https://doi.org/10.5281/zenodo.21000662> (accessed on 2 July 2026).

Acknowledgments: We thank Pei'e Li for her assistance with image labeling and mask annotation. During the preparation of this study, the author(s) used the “brainstorming skill” in Claude Code 2.1.2 for the purposes of paper drafting. This gives us a outline of the paper and we revise and extend based on it to get the current manuscript. The authors have reviewed and edited the output and take full responsibility for the content of this publication. Our code, data and figures are all developed without or with little GenAI involved. So mostly we used AI for paper drafting and grammar checking.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Frisoli, M.L.; Essien, K.; Harris, J.E. Vitiligo: Mechanisms of pathogenesis and treatment. *Annu. Rev. Immunol.* **2020**, *38*, 621–648. [[CrossRef](#)] [[PubMed](#)]
2. Rashighi, M.; Harris, J.E. Vitiligo pathogenesis and emerging treatments. *Dermatol. Clin.* **2017**, *35*, 257–265. [[CrossRef](#)] [[PubMed](#)]
3. Picardo, M.; Dell'Anna, M.L.; Ezzedine, K.; Hamzavi, I.; Harris, J.E.; Parsad, D.; Taieb, A. Vitiligo. *Nat. Rev. Dis. Primers* **2015**, *1*, 15011. [[CrossRef](#)] [[PubMed](#)]
4. Riding, R.L.; Harris, J.E. The role of memory CD8+ T cells in vitiligo. *J. Immunol.* **2019**, *203*, 11–19. [[CrossRef](#)] [[PubMed](#)]
5. Boniface, K.; Seneschal, J.; Picardo, M.; Taïeb, A. Vitiligo: Focus on clinical aspects, immunopathogenesis, and therapy. *Clin. Rev. Allergy Immunol.* **2018**, *54*, 52–67. [[PubMed](#)]

6. Harris, J.E.; Harris, T.H.; Weninger, W.; Wherry, E.J.; Hunter, C.A.; Turka, L.A. A mouse model of vitiligo with focused epidermal depigmentation requires IFN- γ for autoreactive CD8+ T-cell accumulation in the skin. *J. Investig. Dermatol.* **2012**, *132*, 1869–1876. [CrossRef] [PubMed]
7. Lin, J.Y.; Fisher, D.E. Melanocyte biology and skin pigmentation. *Nature* **2007**, *445*, 843–850. [CrossRef] [PubMed]
8. Mort, R.L.; Jackson, I.J.; Patton, E.E. The melanocyte lineage in development and disease. *Development* **2015**, *142*, 620–632. [CrossRef] [PubMed]
9. Chen, D.; Xu, Z.; Cui, J.; Chen, T. A mouse model of vitiligo based on endogenous auto-reactive CD8+ T cell targeting skin melanocyte. *Cell Regen.* **2022**, *11*, 31. [CrossRef] [PubMed]
10. Xu, Z.; Chen, D.; Hu, Y.; Jiang, K.; Huang, H.-W.; Du, Y.; Wu, W.; Wang, J.; Sui, J.; Wang, W.; et al. Anatomically distinct fibroblast subsets determine skin autoimmune patterns. *Nature* **2022**, *601*, 118–124. [CrossRef] [PubMed]
11. Schepeler, T.; Page, M.E.; Jensen, K.B. Heterogeneity and plasticity of epidermal stem cells. *Development* **2014**, *141*, 2559–2567. [CrossRef] [PubMed]
12. Gomez, C.; Chua, W.; Miremadi, A.; Quist, S.; Headon, D.J.; Watt, F.M. The interfollicular epidermis of adult mouse tail comprises two distinct cell lineages that are differentially regulated by Wnt, Edaradd, and Lrig1. *Stem Cell Rep.* **2013**, *1*, 19–27. [CrossRef]
13. Rempel, P.; Deschene, E.R.; Zito, G.; Gonzalez, D.G.; Saotome, I.; Haberman, A.; Greco, V. Live imaging of stem cell and progeny behaviour in physiological hair-follicle regeneration. *Nature* **2012**, *487*, 496–499. [CrossRef] [PubMed]
14. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682. [CrossRef] [PubMed]
15. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef] [PubMed]
16. McQuin, C.; Goodman, A.; Chernyshev, V.; Kametsky, L.; Cimini, B.; Karhohs, K.W.; Doan, M.; Ding, L.; Rafelski, S.M.; Thirstrup, D.; et al. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **2018**, *16*, e2005970. [CrossRef] [PubMed]
17. Berg, S.; Kutra, D.; Kroeger, T.; Straehle, C.; Kausler, B.X.; Haubold, C.; Schiegg, M.; Ales, J.; Beier, T.; Rudy, M.; et al. ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **2019**, *16*, 1226–1232. [CrossRef] [PubMed]
18. Bankhead, P.; Loughrey, M.B.; Fernández, J.A.; Dombrowski, Y.; Mcart, D.G.; Dunne, P.D.; Mcquaid, S.; Gray, R.T.; Murray, L.J.; Coleman, H.; et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **2017**, *7*, 16878. [CrossRef] [PubMed]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015, Munich, Germany, 5–9 October 2015; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351*, pp. 234–241.
20. Falk, T.; Mai, D.; Bensch, R.; Çiçek, Ö.; Abdulkadir, A.; Marrakchi, Y.; Böhm, A.; Deubner, J.; Jäckel, Z.; Seiwald, K.; et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **2019**, *16*, 67–70. Erratum in *Nat. Methods* **2019**, *16*, 351. [CrossRef] [PubMed]
21. Stringer, C.; Wang, T.; Michaelos, M.; Pachitariu, M. Cellpose: A generalist algorithm for cellular segmentation. *Nat. Methods* **2021**, *18*, 100–106. [PubMed]
22. Pachitariu, M.; Stringer, C. Cellpose 2.0: How to train your own model. *Nat. Methods* **2022**, *19*, 1634–1641. [CrossRef] [PubMed]
23. Schmidt, U.; Weigert, M.; Broaddus, C.; Myers, G. Cell detection with star-convex polygons. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2018, Granada, Spain, 16–20 September 2018; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11071*, pp. 265–273.
24. Greenwald, N.F.; Miller, G.; Moen, E.; Kong, A.; Kagel, A.; Dougherty, T.; Fullaway, C.C.; McIntosh, B.J.; Leow, K.X.; Schwartz, M.S.; et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **2022**, *40*, 555–565. [PubMed]
25. Moen, E.; Bannon, D.; Kudo, T.; Graf, W.; Covert, M.; Van Valen, D. Deep learning for cellular image analysis. *Nat. Methods* **2019**, *16*, 1233–1246. [CrossRef] [PubMed]
26. Caicedo, J.C.; Roth, J.; Goodman, A.; Becker, T.; Karhohs, K.W.; Broisin, M.; Molnar, C.; McQuin, C.; Singh, S.; Theis, F.J.; et al. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry A* **2019**, *95*, 952–965. [CrossRef]
27. Bitplane. *Imaris Reference Manual*, version 10.1; Oxford Instruments: Abingdon, UK, 2023.
28. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; Girshick, R. Detectron2. GitHub. 2019. Available online: <https://github.com/facebookresearch/detectron2> (accessed on 1 May 2026).
29. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: New York, NY, USA, 2017; pp. 2980–2988*.
30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]

31. Akyon, F.C.; Altinuc, S.O.; Temizel, A. Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection. In *Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022*; IEEE: New York, NY, USA, 2022; pp. 966–970. [[CrossRef](#)]
32. Riding, R.L.; Richmond, J.M.; Harris, J.E. Mouse model for human vitiligo. *Curr. Protoc. Immunol.* **2019**, *124*, e63. [[PubMed](#)]
33. 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 *Proceedings of the European Conference on Computer Vision (ECCV) 2014, Zurich, Switzerland, 6–12 September 2014*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014; Volume 8693, pp. 740–755.
34. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, 8–14 December 2019*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32, pp. 8024–8035.
35. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 120–125.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.