

Ultra-high Resolution Image Segmentation

Sakura

School of Earth Sciences, Zhejiang University

August 14, 2026

1. Motivation
2. Problem Formulation
3. Patch-Based Methods
4. Multi-Scale Architectures
5. Foundation Models
6. Remote Sensing
7. Evaluation
8. Takeaways

Ultra-high resolution (UHR) images appear across:

- Remote sensing and Earth observation
- Whole-slide pathology
- Aerial and satellite orthophotos
- Gigapixel photography and urban mapping

Goal: dense pixel-wise labels at native resolution.

TODO: add teaser figure under the blog assets folder.

Given $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, predict $\mathbf{Y} \in \mathbb{L}^{H \times W}$.

Key challenges:

1. GPU memory limits at full resolution
2. Context vs. detail trade-off
3. Expensive pixel-level annotation
4. Seam-free fusion across tiles

Sliding-window / tile-based inference remains the practical baseline.

- Train on fixed-size crops; infer with overlap
- U-Net-style fully convolutional designs (Ronneberger et al., 2015)
- Weighted blending at tile boundaries

TODO: pipeline diagram and overlap-blending illustration.

Combine local detail with global semantics:

- Feature pyramids and encoder–decoder variants
- Coarse-to-fine refinement
- Cross-scale attention and memory-efficient context modeling
- Hierarchical self-supervision for gigapixel images (Chen et al., 2022)

TODO: architecture comparison figure.

Segment-anything style models enable flexible prompting (Kirillov et al., 2023).

Open questions for UHR:

- Tiling vs. sparse attention for inference
- Prompt design at scale (points, boxes, text)
- Adaptation without full-resolution fine-tuning

TODO: SAM-on-tiles workflow figure.

Primary domain for UHR segmentation:

- Land-cover / land-use mapping
- Building footprint and road extraction
- Disaster damage assessment

Related survey: self-supervised learning in RS (Wang et al., 2022).

TODO: benchmark table and qualitative results.

- Metrics: mIoU, F1, boundary-aware scores
- Tiled validation vs. full-scene test
- Efficiency: latency, memory, throughput on gigapixel inputs

TODO: evaluation protocol checklist slide.

1. Patch-based inference is strong but not always sufficient.
2. Design axes: context aggregation, fusion, annotation efficiency.
3. Foundation models change the prompting paradigm but not the memory bottleneck.
4. Remote sensing remains a key testbed for UHR segmentation.

TODO: refine as content is added.

Thank you!

bili_sakura@zju.edu.cn