InstaBound: Instance-Boundary Signed Distance Regression

Xinzi He[1,2] XH278@CORNELL.EDU
Ruining Deng[2] RUD4004@MED.CORNELL.EDU
Martin R. Prince[2] MAP2008@MED.CORNELL.EDU
Mert R. Sabuncu[2,3] MSABUNCU@CORNELL.EDU
[1] *Biomedical Engineering, Cornell University, Ithaca, NY, United States*
[2] *Radiology, Weill Cornell Medicine, New York City, NY, United States*
[3] *Electrical and Computer Engineering, Cornell Tech, New York City, NY, United States*

Editors: Under Review for MIDL 2026

Abstract

Accurate instance segmentation is a cornerstone of medical image analysis, essential for tasks ranging from quantifying nuclei in histopathology to measuring cysts, metastases, and other lesions in radiology. However, these tasks are fraught with challenges due to high object density, overlapping boundaries, and significant appearance heterogeneity. In this work, we present InstaBound, a novel deep learning framework that reformulates instance segmentation as an instance boundary Signed Distance Field (iSDF) regression problem. Unlike traditional methods that rely on binary masks or bounding boxes, InstaBound learns a continuous representation of object geometry, implicitly encoding boundary information to robustly separate touching instances. We introduce a dual-task objective function combining truncated-iSDF regression with a Dice loss to ensure both structural coherence and semantic accuracy. Our experiments on the PanNuke and NuInsSeg nuclei datasets demonstrate that InstaBound achieves state-of-the-art performance, significantly outperforming established baselines such as HoVer-Net, Micro-Net, and Mask R-CNN. Specifically, we report a multi-class Panoptic Quality (mPQ) improvement of 3.4% on the PanNuke. Our code and pre-trained models are available to facilitate further research.

Keywords: Nuclei Segmentation, Signed Distance Field, Instance Segmentation, Deep Learning, Computational Pathology, PanNuke, NuInsSeg.

1. Introduction

Instance segmentation is a challenging problem of modern medical image analysis (Ronneberger et al., 2015), serving as the bridge between raw pixel data and quantitative biological insights. By identifying and delineating each distinct object of interest, this task enables the extraction of morphometric features–such as size, shape, density, and spatial distribution–that are often the defining characteristics of disease pathology. In the domain of digital pathology, the precise segmentation of nuclei in Hematoxylin and Eosin (H&E) stained whole slide images is a critical prerequisite for tumor profiling (Graham et al., 2019). It allows for the automated computation of nuclear atypia, mitotic activity indices, and cellularity, which are standard biomarkers for cancer grading and prognosis. Beyond pathology, instance segmentation is equally vital in radiology. Figure 1 illustrates representative samples from the PanNuke (Gamper et al., 2019) and NuInsSeg (Mahbod et al., 2024) datasets used in this study.

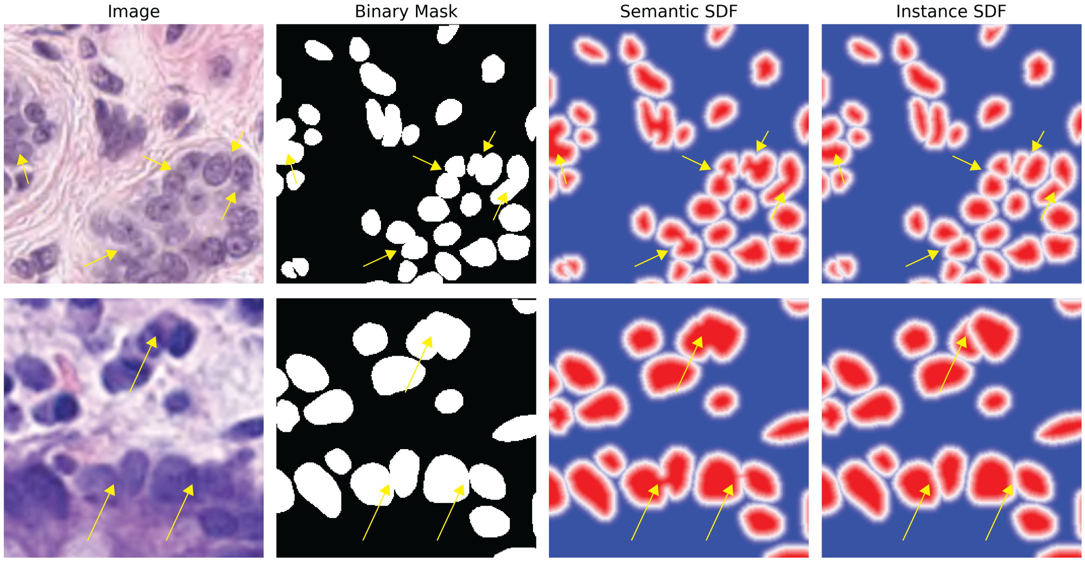

Figure 1: **Dataset Examples.** Sample images, ground truth masks, and computed semantic SDFs and instance SDFs from the PanNuke datasets. The iSDF representation encodes the distance to the nearest instance boundary, providing a robust signal for separating touching instances. Yellow arrows point to clustering nuclei, where semantic SDF clearly struggles to distinguish the instances.

Despite its importance, robust instance segmentation remains a formidable challenge. Biological structures often appear in dense, non-uniform clusters where boundaries between adjacent objects are ambiguous or entirely invisible. In high-grade tumors, nuclei are frequently packed together, creating large, overlapping aggregates. These scenarios expose the limitations of traditional computer vision techniques. Thresholding and standard watershed transforms are notoriously brittle, relying on heuristic parameters that fail to generalize across the extreme biological heterogeneity found in clinical data.

The advent of deep learning has shifted the paradigm, yet significant hurdles remain. Generic instance segmentation architectures adapted from natural image scenes, such as Mask R-CNN (He et al., 2017), employ a "detect-then-segment" strategy reliant on anchor boxes. While effective for sparse objects (e.g., pedestrians, cars), this approach falters in the dense environments typical of microscopy and medical imaging. When objects are heavily clustered, their bounding boxes overlap significantly, causing Non-Maximum Suppression (NMS)–the mechanism designed to filter duplicate detections–to erroneously suppress valid instances. Furthermore, the pixel-wise binary masks generated by these networks often lack the sub-pixel geometric precision required to separate touching boundaries, resulting in under-segmentation.

To address these issues, the field has seen the emergence of specialized architectures. Recent benchmarks have also highlighted the critical role of post-processing strategies, such as watershed, morphology, and skeletonization, in refining segmentation boundaries (Zhao et al., 2025). Notably, HoVer-Net (Graham et al., 2019) represented a significant leap forward by predicting horizontal and vertical distance maps relative to the center of mass of each nucleus. While highly effective, HoVer-Net necessitates a complex multi-branch architecture and an intricate post-processing pipeline to recombine these maps. Other approaches like StarDist (Schmidt et al., 2018) simplify the problem by assuming star-convex shapes, a geometric prior that holds for many nuclei but fails for irregular, concave structures often seen in disease states. While distance map regression was explored in earlier works like Naylor et al. (Naylor et al., 2019), these methods often struggled with optimization stability and boundary precision.

In this work, we introduce **InstaBound**, a framework that reformulates instance segmentation as instance-level Signed Distance Field (iSDF) regression. Avoiding the architectural complexity of methods like Mask R-CNN (He et al., 2017) or HoVer-Net (Graham et al., 2019), we demonstrate that a standard encoder-decoder trained to regress a *truncated* iSDF provides robust supervision. This representation implicitly encodes geometry—creating distinct "basins of attraction" for each object—allowing for the separation of touching instances using standard watershed techniques without complex heuristics.

Our contributions are:

- **Plug-and-Play Simplicity**: A unified, single-branch framework compatible with most encoder-decoders. It eliminates the need for anchor boxes, region proposals, or multi-directional maps while maintaining high performance.
- **Robust iSDF Regression**: By regressing a truncated iSDF, the model focuses on the immediate vicinity of boundaries, effectively resolving the "touching object" problem.
- **Dual-Task Learning**: We utilize a novel loss combining Huber loss (for structural coherence via iSDF) and Soft Dice loss (for semantic accuracy).
- **State-of-the-Art Performance**: We achieve superior Panoptic Quality (PQ) compared to established baselines on the PanNuke (Gamper et al., 2019) and NuInsSeg (Mahbod et al., 2024) datasets.

2. Methodology

We propose **InstaBound**, a unified framework that efficiently combines semantic classification and instance regression into a single output branch. An overview of our framework, including the training and inference stages, is illustrated in Figure 2. Unlike traditional approaches that separate these tasks into different heads (e.g., Mask R-CNN) or rely purely on regression, our method utilizes a "Logit-as-iSDF" paradigm. The network's raw output logits are trained to simultaneously represent the semantic probability distribution (relative values) and the geometric instance Signed Distance Field (absolute values).

2.1. InstaBound Network and Loss Function

Network Architecture Our backbone is a customized deep U-Net (Ronneberger et al., 2015) trained from scratch. The network employs a 6-level encoder-decoder structure with strided convolutions for downsampling and nearest-neighbor interpolation for upsampling.

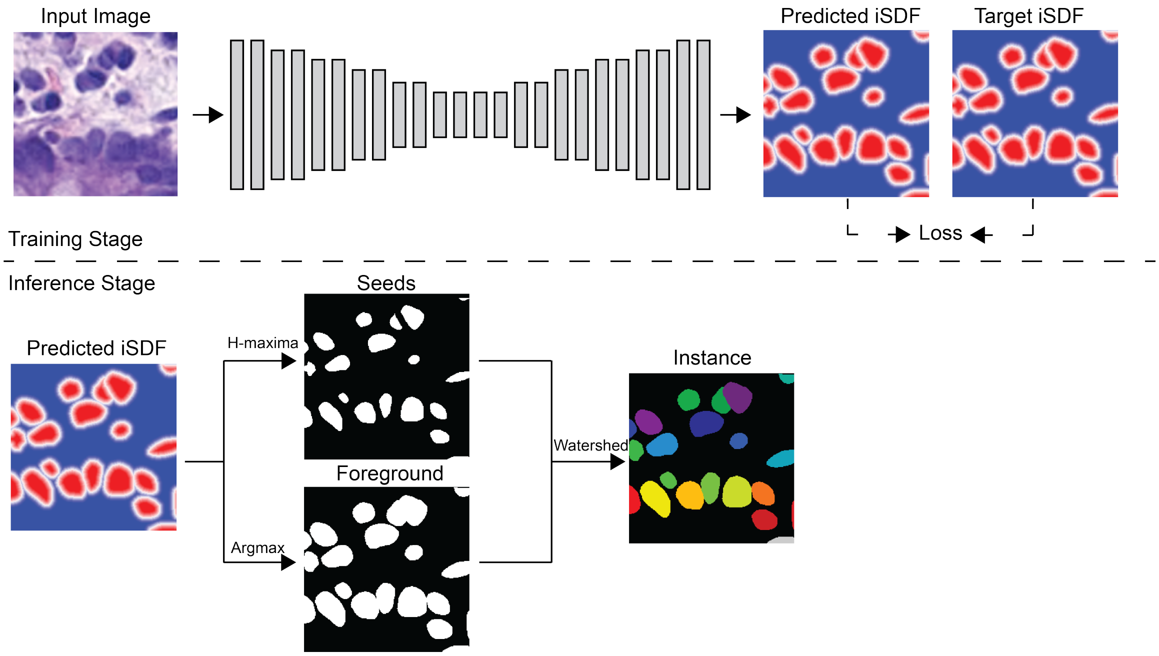

Figure 2: **InstaBound Framework Overview.** The Training Stage (top) regresses truncated instance Signed Distance Fields (iSDFs) directly from the image, supervised by a dual-task objective. The Inference Stage (bottom) utilizes the predicted iSDF for seed detection (via H-maxima) and the foreground mask for marker-controlled watershed segmentation to robustly delineate individual instances.

The network terminates in a single 1 × 1 convolutional head with C output channels, where $C = K + 1$ (K tissue classes plus one background class). Crucially, the output logits $\mathbf{L} \in \mathbb{R}^{C \times H \times W}$ is supervised directly with truncated iSDFs.

Dual-Purpose Logits and Loss We interpret the output tensor $\mathbf{L}$ in two complementary ways during training, optimized via a unified objective function:

$$L_{total} = L_{dice} + \lambda_{reg} L_{huber} \tag{1}$$

where $\lambda_{reg} = 0.01$ is a weighting factor balancing the two tasks. The components are defined as follows:

1. **Semantic Classification (L_{dice}):** We apply the softmax function along the channel dimension to obtain class probabilities: $P_{class} = \text{Softmax}(\mathbf{iSDF}/\mathbf{T})$, where $iSDF$ is instance Signed Distance Fields and T is softmax temperature. L_{dice} is the **Soft Dice Loss** computed on these probabilities to enforce semantic segmentation accuracy, ensuring the dominant channel corresponds to the correct tissue type.

2. **Geometric Regression (L_{huber}):** The raw values of the logits are treated as predicted iSDFs. For a pixel belonging to class k, $\mathbf{L}_k(x)$ predicts the truncated distance

to the boundary. L_{huber} is the **Huber Loss** applied directly to these logits against the ground truth truncated iSDF. This regression task encourages the network to learn the internal shape and peaks of the distance field, which are critical for separating touching instances.

2.2. Post-processing

Our inference strategy leverages the dual nature of the logits to robustly separate instances. We rely on the semantic classification (Softmax-Argmax) to define the "where" (the instance mask) and the geometric regression (Raw Logits) to define the "what" (individual instance separation). The process proceeds as follows:

1. **Semantic Masking**: We first compute the softmax probabilities and take the argmax to determine the dominant class k^* for each pixel, yielding semantic masks.

2. **iSDF Extraction**: For pixels classified as class k, we extract the raw logit values from channel k. This creates a composite iSDF map Φ_{pred}. Note that we do not use a zero-level set thresholding ($\Phi > 0$) for masking; instead, we strictly rely on the Softmax-Argmax decision, which is robust to calibration errors in the regression magnitude.

3. **Seed Detection**: We smooth Φ_{pred} with a Gaussian kernel ($\sigma = 1$) to reduce texture noise. We then apply an **h-maxima** transform (Van der Walt et al., 2014), which isolates significant local maxima by suppressing peaks with a prominence lower than $h = 2$[1]. This is achieved by performing a morphological reconstruction by dilation of Φ_{smooth} from the marker image defined as $\Phi_{smooth} - h$, effectively filtering out minor fluctuations while preserving the main instance centroids. The remaining regional maxima serve as robust seeds for the watershed.

4. **Marker-Controlled Watershed**: We perform a watershed transform on the inverted iSDF ($-\Phi_{pred}$), using the detected seeds as sources and the semantic masks from Step 1 as the bounding control region. This effectively partitions the connected semantic components into individual instances.

3. Experiments

3.1. Datasets and Pre-processing

PanNuke Dataset (Gamper et al., 2019): This large-scale dataset contains over 7,000 patches of 256x256 pixels across 19 tissue types (e.g., Breast, Kidney, Colon). It includes multi-class instance annotations for 5 nuclei types (Neoplastic, Inflammatory, Connective, Dead, Non-Neoplastic). We utilize the official 3-fold cross-validation training setting.

NuInsSeg Dataset (Mahbod et al., 2024): A diverse collection of H&E stained tissue images from 31 organs, with (binary) nucleus instance annotations. For this dataset, we conducted a 5-fold cross-validation using the same code and random seed as for the NuInsSeg dataset to ensure consistent data splitting.

---

[1] Hyper-parameters σ and h were chosen based on common practice and not optimized.

Augmentation : We applied heavy data augmentation during training, including random rotations $(90°, 180°, 270°)$, horizontal/vertical flips, and color jittering (brightness and contrast) to simulate staining variations.

3.2. Implementation Details

All models were implemented in PyTorch 2.0 (Paszke et al., 2019). Training was conducted on a single NVIDIA A100 (40GB) GPU. We used the AdamW optimizer (Loshchilov and Hutter, 2019) with an initial learning rate of 5×10^{-5} and a cosine annealing schedule with warm up 10 epoch. The batch size was set to 10. Models were trained for 500 epochs. The channel number of InstaBound is set to 64 which is the same as 2d nnUNet (Isensee et al., 2021). For loss hyperparameters, we set temperature to 1 and clamp to 10. In post processing, we use $h = 2$ and $sigma = 1$.

3.3. Evaluation Metrics

We employ Panoptic Quality (PQ) (Kirillov et al., 2019) as our primary metric, which provides a unified measure of instance segmentation performance. PQ is defined as the product of Segmentation Quality (SQ) and Recognition Quality (RQ), i.e., $PQ = SQ \times RQ$. Specifically, the "Recognition Quality" (RQ) corresponds to the per-object F_1-score of class c in image i. The "Segmentation Quality" (SQ) is the average Intersection-over-Union (IoU) computed on the matched pairs of ground truth and predicted instances of this class. We report both multi-class PQ (mPQ) and binary PQ (bPQ). While bPQ treats all nuclei as a single foreground class, mPQ averages performance across all distinct nuclei types, serving as a more rigorous metric that penalizes misclassification. Our method excels particularly in mPQ, outperforming baselines where it matters most for clinical sub-typing. We present the Aggregated Jaccard Index (AJI) (Kumar et al., 2017) and Dice coefficient (Dice, 1945), which are standard metrics reported on the binary NuInsSeg Dataset.

4. Results

4.1. Comparative Analysis on PanNuke

We compared InstaBound against four strong baselines: Semantic U-Net (with watershed), Mask R-CNN (He et al., 2017), Micro-Net (Raza et al., 2019), and HoVer-Net (Graham et al., 2019). Table 1 summarizes the results on PanNuke. Figure 3 provides a visual comparison of InstaBound against U-Net and our SemanBound ablation, described below.

Quantitative results in Table 1 demonstrate that InstaBound sets a new state-of-the-art on the PanNuke dataset, achieving an average mPQ of 0.4787. Notably, our method surpasses HoVer-Net, a specialized multi-branch architecture, by a margin of 1.6% in mPQ. This improvement is driven by a superior Segmentation Quality (SQ), suggesting that our continuous iSDF regression yields more precise boundary adherence than the horizontal/vertical distance maps used in HoVer-Net.

A closer inspection of tissue-specific performance reveals significant gains in challenging, high-density tissues. For instance, in "Colon" and "Kidney" tissues, which are characterized by densely packed glandular structures or nuclei often cluster, InstaBound improves mPQ by approximately 7.8% and 17.4% respectively over HoVer-Net. These results validate our

Table 1: **Average mPQ and bPQ across three dataset splits.** We compare InstaBound against state-of-the-art methods (Left|) and internal baselines (|Right). InstaBound achieves the **best** or second-best multi-class Panoptic Quality (mPQ), setting a new SOTA.

Tissue Type	**DIST**		**Mask-RCNN**		**Micro-Net**		**HoVer-Net**		**U-Net**		**SemanBound**		**InstaBound**	
	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ	mPQ	bPQ
Adrenal Gland	0.3442	0.5603	0.3470	0.5546	0.4153	0.6440	0.4812	0.6962	0.4673	0.6225	0.4612	0.6197	**0.4929**	0.6698
Bile Duct	0.3614	0.5384	0.3536	0.5567	0.4124	0.6232	**0.4714**	0.6696	0.4101	0.5717	0.4117	0.5882	0.4668	0.6434
Bladder	0.4463	0.5625	0.5065	0.6049	0.5357	0.6488	0.5792	0.7031	0.4925	0.6210	0.4965	0.6206	**0.5832**	0.7019
Breast	0.3790	0.5466	0.3882	0.5574	0.4407	0.6029	0.4902	0.6470	0.4429	0.5384	0.4442	0.5790	**0.4957**	0.6330
Cervix	0.3371	0.5309	0.3402	0.5483	0.3795	0.6101	0.4438	0.6652	0.4272	0.5705	0.4253	0.6027	**0.4761**	0.6560
Colon	0.2989	0.4508	0.3122	0.4603	0.3414	0.4972	0.4095	0.5575	0.3654	0.4582	0.3716	0.4874	**0.4416**	0.5557
Esophagus	0.3942	0.5295	0.4311	0.5691	0.4668	0.6011	0.5085	0.6427	0.4315	0.5213	0.4439	0.5555	**0.5285**	0.6392
Head & Neck	0.3177	0.4764	0.3946	0.5457	0.3668	0.5242	0.4530	0.6331	0.3843	0.5317	0.4002	0.5492	**0.4777**	0.6240
Kidney	0.3339	0.5727	0.3553	0.5092	0.4165	0.6321	0.4424	0.6836	0.4369	0.5909	0.4291	0.5803	**0.5192**	0.6719
Liver	0.3441	0.5818	0.4103	0.6085	0.4365	0.6666	0.4974	0.7248	0.4842	0.6491	0.4670	0.6432	**0.5032**	0.6945
Lung	0.2809	0.4978	0.3182	0.5134	0.3370	0.5588	**0.4004**	0.6302	0.3644	0.5303	0.3636	0.5523	0.3977	0.5931
Ovarian	0.3789	0.5289	0.4337	0.5784	0.4387	0.6013	0.4863	0.6309	0.4017	0.4725	0.4136	0.5354	**0.5147**	0.6425
Pancreatic	0.3395	0.5343	0.3624	0.5460	0.4041	0.6074	**0.4600**	0.6491	0.4160	0.5576	0.4314	0.5928	0.4567	0.6404
Prostate	0.3810	0.5442	0.3959	0.5789	0.4341	0.6049	0.5101	0.6615	0.4254	0.5280	0.4433	0.5728	**0.5217**	0.6588
Skin	0.2627	0.5080	0.2665	0.5021	0.3223	0.5817	0.3429	0.6234	0.3133	0.4779	0.3375	0.5403	**0.3828**	0.6084
Stomach	0.3369	0.5553	0.3684	0.5976	0.3872	0.6293	**0.4726**	0.6886	0.4300	0.5972	0.4286	0.6243	0.4468	0.6744
Testis	0.3278	0.5548	0.3512	0.5420	0.4088	0.6300	0.4754	0.6890	0.4198	0.5508	0.4612	0.6129	**0.4931**	0.6519
Thyroid	0.2574	0.5596	0.3037	0.5712	0.3712	0.6555	0.4315	0.6983	0.3660	0.5756	0.3828	0.6078	**0.4368**	0.6678
Uterus	0.3487	0.5246	0.3683	0.5589	0.3965	0.5821	0.4393	0.6393	0.3228	0.4620	0.3902	0.5408	**0.4610**	0.6285
Average	0.3406	0.5346	0.3688	0.5528	0.4059	0.6053	0.4629	0.6596	0.4106	0.5488	0.4212	0.5792	**0.4787**	0.6450

hypothesis that the instance-aware iSDF provides a more robust signal for separating touching objects than binary masks or bounding boxes. While HoVer-Net remains competitive in binary PQ (bPQ) - essentially a detection task - InstaBound's superior multi-class performance (mPQ) highlights its efficacy in the more clinically relevant task of simultaneously segmenting and classifying nuclei types. As visualized in Figure 3, baseline methods such as U-Net and the semantic-only ablation (SemanBound) struggle significantly with dense cellular aggregates, frequently merging touching nuclei into single objects. In contrast, InstaBound predicts sharp, instance-specific valleys in the distance field. This geometric fidelity allows the marker-controlled watershed to successfully delineate individual nuclei in crowded regions (e.g., Colon and Kidney tissues), directly translating to higher Recognition Quality (RQ) and minimized under-segmentation errors.

4.2. Comparative Analysis on NuInsSeg

The results on the NuInsSeg dataset (Table 2) demonstrate that InstaBound achieves state-of-the-art performance across all evaluated metrics. It is important to note that the baselines presented, particularly the Residual Attention U-Net (Mahbod et al., 2024) and Dual Decoder U-Net (Mahbod et al., 2022), represent the current state-of-the-art on this benchmark as reported in recent literature.

Our method matches the highest pixel-level Dice score (0.814) achieved by Residual Attention U-Net. However, the true strength of InstaBound lies in its instance-level performance. We establish a new State-of-the-art for instance-aware metrics, outperforming the strongest competitor (Dual Decoder U-Net) by significant margins of 5.9% in AJI and 7.01% in PQ. This discrepancy between the marginal gain in Dice and the substantial boost

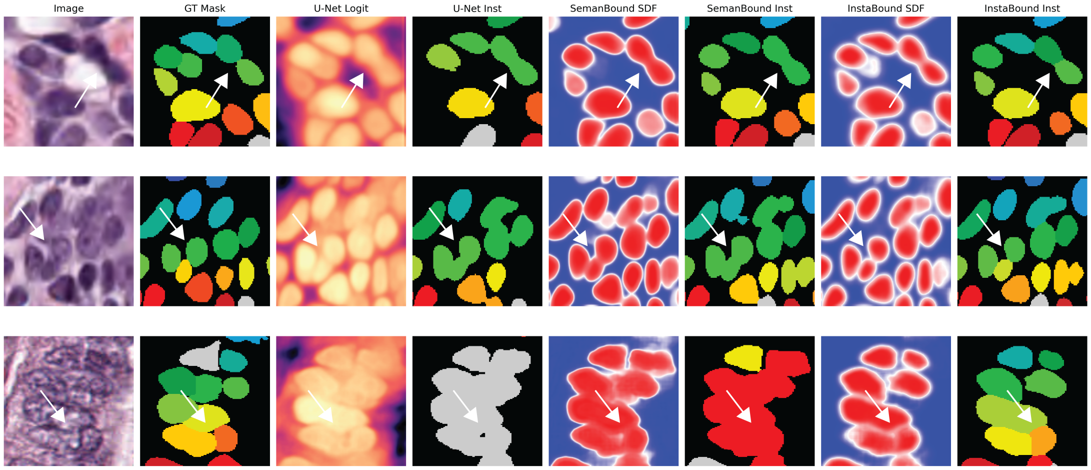

Figure 3: **Visual Comparison.** Comparison of InstaBound against U-Net and SemanBound. Columns show (left to right): Input Image, Ground Truth, U-Net (Logit/Instance), SemanBound (SDF/Instance), and InstaBound (SDF/Instance). White arrows show InstaBound clearly separates touching nuclei that are merged by other methods.

in AJI/PQ underscores the core contribution of our work: while standard U-Nets can accurately classify pixels (high Dice), they often fail to separate adjacent instances, merging them into single blobs. InstaBound's truncated iSDF regression effectively resolves these topological errors, preserving individual instance identities without compromising semantic accuracy.

4.3. Ablation Studies

To validate the sensitivity of our method to its key hyperparameters, we performed a grid search over the iSDF clamping value (c) and the Softmax temperature in Dice Loss (T). Table 3 summarizes the results.

We observe a clear trade-off with the clamping value c. Extremely small values (e.g., $c = 1$) excessively truncate the distance field, limiting the gradient flow to a narrow band around the boundary and reducing the network's receptive field for geometric context. This results in poorer performance. Conversely, a moderate clamp ($c = 10$) is optimal, allowing the network to learn a robust, informative gradient within the object interior while ignoring irrelevant background variations far from the instance.

Regarding the temperature T, smaller values (e.g., $T = 0.1$) yield the best raw Dice score by approximating a smoother, more convex loss landscape. However, we opt for $T = 1.0$ in our final model as it offers a better balance between performance (only 1% drop in Dice) and training stability. High temperatures (e.g., $T = 10$) flatten iSDF too aggressively, leading to vanishing gradients and unstable optimization, as evidenced by the sharp drop in performance.

Table 2: **Comparative Performance on NuInsSeg Dataset.** Comparison of methods trained and evaluated on the NuInsSeg dataset. Comparison models and results are adapted from (Mahbod et al., 2024). Best scores are in **bold**, and second best are <u>underlined</u>.

Model	**Dice**	**AJI**	**PQ**
Shallow U-Net (Ronneberger et al., 2015)	0.788	0.505	0.427
Deep U-Net (Ronneberger et al., 2015)	0.797	0.494	0.404
Attention U-Net (Oktay et al., 2018)	<u>0.805</u>	0.457	0.364
Residual Attention U-Net (He et al., 2016; Oktay et al., 2018)	**0.814**	0.462	0.369
Two-stage U-Net (Mahbod et al., 2019)	0.766	0.528	0.472
Dual Decoder U-Net (Mahbod et al., 2022)	0.794	<u>0.559</u>	<u>0.513</u>
SemanBound (Ours)	0.791	0.543	<u>0.513</u>
InstaBound (Ours)	**0.814**	**0.592**	**0.549**

Impact of Instance-Aware SDF We investigated the necessity of modeling instance-specific boundaries by comparing InstaBound against standard U-Net and our baseline "SemanBound" (which regresses SDF from the semantic label map, treating touching instances as a single object). As shown in Table 1, both U-Net and SemanBound suffer a significant performance drop compared to InstaBound (average mPQ 0.4106 and 0.4212 vs 0.4787). In Table 2, InstaBound achieved average PQ 0.549 which outperforms SemanBound 0.513. The degradation is most severe in dense tissues where nuclei often touch; without the instance-aware SDF valleys (local minima) between objects, the post-processing fails to separate them, leading to severe under-segmentation. The "Dead" cell class, which often appears in necrotic clusters, showed particularly poor recovery in the baselines, highlighting the critical role of our instance-based formulation.

Table 3: **Ablation Study on Hyperparameters.** We report Dice scores for varying iSDF clamping distances (c) and Soft Dice temperatures (T). The best performance is observed with $c = 10$ and $T = 10.0$.

Clamp (c)	**Temperature (T)**		
	0.1	**1.0**	**10.0**
1	0.4073	0.3766	0.3687
2	0.3967	0.3951	0.3861
5	<u>0.4075</u>	0.4041	0.3913
10	**0.4168**	0.4065	0.3812
∞	0.3849	0.3818	0.3477

Table 4: **Post-processing Hyperparameter Search.** We report the mean Panoptic Quality (mPQ) for varying Gaussian smoothing sigma (σ) and h-maxima prominence (h). The best performance is observed with $\sigma = 3.0$ and $h = 1.0$.

h-maxima (h)	**Sigma (σ)**				
	0.1	**1.0**	**2.0**	**3.0**	**4.0**
1.0	0.4391	0.4640	0.4718	**0.4750**	0.4731
2.0	0.4497	0.4705	<u>0.4734</u>	0.4722	0.4682
3.0	0.4514	0.4707	0.4712	0.4686	0.4603
4.0	0.4499	0.4685	0.4674	0.4621	0.4470
5.0	0.4477	0.4657	0.4622	0.4493	0.4340

Post-processing Hyperparameters In addition, we analyze the impact of the post-processing parameters of the watershed. Table 4 presents the mPQ scores across a range of Gaussian smoothing sigmas (σ) and h-maxima prominence values (h). The results indicate that moderate smoothing ($\sigma = 3.0$) coupled with a sensitive peak detection ($h = 1.0$) provides optimal performance, effectively suppressing noise while preserving the centroids of the instance.

5. Discussion and Conclusion

The experimental results presented in this work confirm that reformulating instance segmentation as an instance-aware Signed Distance Field (iSDF) regression is a highly effective strategy for computational pathology. By regressing a continuous geometric signal, InstaBound provides a richer and more spatially consistent supervision than traditional binary masks, effectively resolving the topological ambiguities inherent in dense cellular aggregates. This is evidenced by our state-of-the-art performance on both the PanNuke and NuInsSeg datasets (Gamper et al., 2019; Mahbod et al., 2024), where our method demonstrated superior object-level separation (mPQ, AJI or PQ) compared to established baselines like HoVer-Net (Graham et al., 2019), Dual Decoder U-Net (Mahbod et al., 2022) and Mask R-CNN (He et al., 2017).

While our "Logit-as-iSDF" paradigm significantly simplifies the architectural complexity of instance segmentation, the framework's performance currently relies on the selection of watershed post-processing parameters (e.g., h-maxima prominence). Although our ablation studies indicate a relatively stable range of optimal values, this dependency remains a limitation compared to fully end-to-end approaches.

In conclusion, InstaBound offers a robust, plug-and-play solution for dense instance segmentation, addressing the critical challenge of touching objects with high geometric fidelity. Future work will focus on eliminating the heuristic post-processing step by integrating differentiable watershed modules directly into the network architecture. Furthermore, we aim to extend this methodology to 3D volumetric data, exploring its potential for tasks such as cyst segmentation in MRI for ADPKD management.

References

Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26 (3):297–302, 1945.

Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benet, Ali Khuram, and Nasir Rajpoot. Pannuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. In *European Congress on Digital Pathology*, volume 11435 of *Lecture Notes in Computer Science*, pages 11–19. Springer, 2019. doi: 10.1007/978-3-030-23937-4_2.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019. doi: 10.1016/j.media.2019.101563.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9404–9413, 2019.

Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Sethu Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Amirreza Mahbod, Rupert Ecker, and Isabella Ellinger. A two-stage u-net algorithm for segmentation of nuclei in hé-stained tissues. In *European Congress on Digital Pathology*, pages 75–82. Springer, 2019.

Amirreza Mahbod, Gerald Schaefer, Georg Dorffner, Sepideh Hatamikia, Rupert Ecker, and Isabella Ellinger. A dual decoder u-net-based model for nuclei instance segmentation in hematoxylin and eosin-stained histological images. *Frontiers in Medicine*, 9:978146, 2022. doi: 10.3389/fmed.2022.978146.

Amirreza Mahbod, Christine Polak, Katharina Feldmann, Rumsha Khan, Katharina Gelles, Georg Dorffner, Ramona Woitek, Sepideh Hatamikia, and Isabella Ellinger. Nuinsseg:

A fully annotated dataset for nuclei instance segmentation in h&e-stained histological images. volume 11, page 295. Nature Publishing Group, 2024. doi: 10.1038/s41597-024-03117-2.

Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019. doi: 10.1109/TMI.2018.2865709.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Matt Higgins, Kazunari Kazuno, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, volume 32, 2019.

Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, Simon Graham, David Epstein, Stella Pelengaris, Michael Khan, and Nasir M Rajpoot. Micro-net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, 52:160–173, 2019. doi: 10.1016/j.media.2018.12.003.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, pages 265–273. Springer, 2018.

Stefan Van der Walt, Johannes L Schr"onberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

Leiyue Zhao, Yuechen Yang, Yanfan Zhu, Haichun Yang, Yuankai Huo, Paul D Simonson, Kenji Ikemura, Mert R Sabuncu, Yihe Yang, and Ruining Deng. Dymorph-b2i: Dynamic and morphology-guided binary-to-instance segmentation for renal pathology. *arXiv preprint arXiv:2508.15208*, 2025.