Swarm-Based Decoding for Pixel-Based Detection

Ciro Russo[1,2][0009-0002-8751-8605], Alessandro Bria[2][0000-0002-2895-6544], and Claudio Marrocco[1,2][0000-0003-0840-7350]

[1] Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy
[2] EUT+, European University of Technology, European Union, Italy

Abstract. Modern detection architectures generate large sets of spatially redundant candidate predictions that must be consolidated into a discrete set of final detections through a dedicated decoding stage. Existing decoding strategies are typically based on local suppression mechanisms that process candidate predictions independently or through pairwise interactions, limiting the explicit exploitation of the global spatial organization emerging from the candidate field. In this work, we investigate a swarm-based decoding strategy for pixel-based detection inspired by Glowworm Swarm Optimization (GSO). The proposed formulation models candidate predictions as interacting agents whose dynamics are guided by confidence-driven neighborhood interactions, progressively organizing into spatially coherent structures corresponding to object hypotheses. Final detections are extracted through density-based clustering over the converged swarm configuration. The proposed approach operates as a plug-in decoding module and does not require modifications to the underlying detector architecture. Experiments are conducted on a gravity-point-based detector for nuclei localization in cervical cytology imaging, a particularly suitable setting due to the presence of highly crowded and spatially correlated candidate predictions. The analysis highlights different operating behaviors between suppression-based and swarm-based decoding strategies, suggesting that collective swarm dynamics provide a viable alternative for aggregating dense candidate localization fields.

Keywords: Pixel-Based Detection · Swarm Intelligence · Swarm-Based Decoding · Glowworm Swarm Optimization

1 Introduction

Modern detection architectures generate large sets of spatially redundant candidate localizations that must be consolidated into a discrete set of final detections. This aggregation stage, commonly referred to as decoding, plays a central role in transforming dense prediction fields into interpretable object hypotheses. In pixel-based detection settings, candidate predictions often form highly correlated spatial structures around the same target, particularly in scenarios involving small, ambiguous, or densely packed objects [15]. Effectively exploiting these collective spatial patterns remains a challenging aspect of detection decoding. Most existing decoding strategies rely on local suppression mechanisms driven by confidence ranking and spatial proximity [10]. While highly effective in many practical settings, such approaches primarily process candidate predictions independently or through pairwise interactions, limiting the explicit exploitation of the global spatial organization emerging from the candidate field. In dense prediction regimes, however, groups of neighboring candidates may collectively encode structural information about the underlying target distribution [9].

Swarm intelligence provides an alternative perspective for addressing such aggregation problems through collective dynamics emerging from local agent interactions [11,17]. Rather than resolving redundancy through independent greedy decisions, swarm-based systems progressively organize candidate structures through iterative neighborhood interactions. Motivated by these considerations, in this work we investigate a swarm-based decoding strategy for pixel-based detection inspired by Glowworm Swarm Optimization (GSO) [13]. Rather than interpreting detector outputs as independent candidate predictions to be filtered through local suppression rules, the proposed formulation models the candidate field as a collective spatial system in which candidate localizations interact through confidence-driven neighborhood dynamics. Each candidate is represented as a glowworm agent whose luminosity is associated with the detector confidence score, while iterative local interactions progressively organize agents into spatially coherent structures corresponding to object hypotheses. Final detections are subsequently extracted through density-based clustering over the converged swarm configuration. The proposed approach operates as a plug-in decoding module that can be integrated into existing pixel-based detection pipelines without requiring modifications to the underlying detector architecture. The proposed decoder is evaluated on nuclei localization in cervical cytology imaging using a gravity-point-based detector [18], a particularly suitable setting due to the presence of highly crowded and spatially correlated candidate predictions. Although the experimental analysis is conducted in a cytological imaging scenario, the proposed decoding formulation remains independent of the application domain and can be integrated into any pixel-based detection framework producing scored candidate localization fields.

The remainder of this paper is organized as follows. Section 2 reviews related work on detection decoding and swarm intelligence in pattern recognition. Section 3 formalizes the decoding problem and presents the proposed swarm-based decoder. Section 4 describes the experimental protocol. Section 5 reports and discusses results. Section 6 concludes.

2 Related Work

Decoding strategies represent a fundamental component of modern detection pipelines, particularly in dense prediction settings where large numbers of spatially redundant candidates must be consolidated into a discrete set of final detections. Non-maximum suppression (NMS) [14] remains the dominant decoding approach due to its widespread adoption and computational efficiency. Several extensions have been proposed to improve candidate aggregation in crowded scenarios. Soft-NMS [2] replaces hard suppression with score attenuation in order to reduce false negatives, while Weighted Box Fusion [20] aggregates overlapping predictions through weighted averaging. Learnable suppression approaches [9] further parameterizing the decoding stage through auxiliary trainable modules. Despite their differences, these methods primarily rely on local suppression mechanisms driven by confidence ranking and spatial proximity. Alternative decoding formulations based on collective aggregation have also been explored. Mean-shift clustering [4], for instance, iteratively aggregates candidates through kernel-based mode seeking and is conceptually closer to global candidate organization strategies. However, such approaches do not explicitly model candidate predictions as interacting entities evolving through distributed collective dynamics.

The emergence of coherent global behavior from simple local interactions is a central principle of swarm intelligence [11,17]. Swarm-based optimization methods such as Particle Swarm Optimization [12] and Ant Colony Optimization [5] have been successfully applied to several pattern recognition and computer vision problems, including clustering, feature selection, and image segmentation [16]. Among swarm-based methods, GSO [13] is particularly relevant in the context of detection decoding because it was specifically designed to support the simultaneous emergence of multiple local groups through adaptive luciferin-driven neighborhood interactions. Unlike single-optimum swarm methods, GSO naturally supports multi-modal convergence, making it particularly suitable for scenarios in which multiple spatial peaks corresponding to distinct object hypotheses must be localized concurrently. These properties become particularly relevant in modern detection frameworks, where object hypotheses often emerge from dense spatial localization fields rather than from sparse proposal generation stages [3]. Such formulations naturally produce highly redundant and spatially correlated candidate predictions distributed over the image domain, creating a suitable setting for collective decoding strategies. By operating directly over dense candidate localization fields, swarm-based decoding formulations can exploit collective spatial interactions that are not explicitly modeled by conventional suppression-based approaches [7]. Although the proposed strategy is evaluated on a gravity-point-based detector, the formulation itself remains independent of the underlying detector architecture and only requires a scored candidate localization field as input. To the best of our knowledge, this work represents the first application of Glowworm Swarm Optimization as a decoding strategy for pixel-based detection.

3 Methodology

The proposed framework formulates detection decoding as a collective spatial aggregation process operating over dense candidate localization fields. The decoding stage models candidate predictions as interacting swarm agents whose dynamics progressively organize the candidate field into coherent spatial structures corresponding to object hypotheses. The following sections describe the candidate representation and the proposed swarm-based decoding procedure.

3.1 Pixel-Based Candidate Representation

Pixel-based detectors generate large sets of scored candidate localizations distributed over the image domain. Rather than interpreting these predictions as independent hypotheses to be filtered through local suppression rules, the proposed formulation models the detector output as a spatial candidate field whose structure emerges from the collective organization of neighboring candidates.

Let f_θ denote a dense pixel-based detector parameterized by θ. Given an input image I, the detector produces a set of P candidate localizations:

$$f_\theta(I) \;=\; \big\{(\mathbf{p}_i, s_i)\big\}_{i=1}^{P}, \qquad \mathbf{p}_i \in \mathbb{R}^2, \;\; s_i \in [0,1] \tag{1}$$

where $\mathbf{p}_i$ denotes the spatial position of candidate i and s_i its associated confidence score. The number of generated candidates is typically large, depending on the detector resolution and prediction density.

The objective of the decoding stage is to transform the dense candidate field into a discrete set of final detections:

$$D\Big(\{(\mathbf{p}_i, s_i)\}_{i=1}^{P}\Big) \;=\; \big\{(\hat{\mathbf{c}}_k, \hat{s}_k)\big\}_{k=1}^{K}, \qquad K \ll P \tag{2}$$

where $\hat{\mathbf{c}}_k$ and $\hat{s}_k$ denote the position and confidence score of the k-th final detection, respectively. The resulting candidate field constitutes the input of the proposed decoding stage, which aims to aggregate spatially correlated predictions into a discrete set of final detections.

3.2 Swarm-Based Decoding

The proposed decoder models each candidate localization as a swarm agent evolving in the two-dimensional image plane. Candidate interactions are governed by confidence-driven local dynamics inspired by GSO [13], progressively organizing neighboring candidates into spatially coherent groups corresponding to object hypotheses. The complete decoder is summarized in Algorithm 1.

Initialization. Given the candidate field $\{(\mathbf{p}_i, s_i)\}_{i=1}^{P}$, each candidate is initialized as an agent positioned at $\mathbf{p}_i^{(0)} = \mathbf{p}_i$, with luciferin value $\ell_i^{(0)} = s_i$ and sensor range $r_i^{(0)} = r_s$, where r_s denotes the maximum sensor range.

Luciferin Update. At each iteration t, the luciferin of agent i is updated as:

$$\ell_i^{(t)} = (1-\rho)\,\ell_i^{(t-1)} + \gamma\, s_i \tag{3}$$

where $\rho \in (0,1)$ is a decay coefficient and $\gamma > 0$ controls the contribution of the detector confidence score. Because s_i is fixed by the detector output, the recurrence converges monotonically and preserves the relative luminosity ordering induced by the network.

Neighborhood and Movement. The neighborhood of agent i at iteration t is defined as the set of agents within its current sensor range that exhibit a strictly higher luciferin value:

$$\mathcal{N}_i^{(t)} = \Big\{ j \neq i \; : \; \big\|\mathbf{p}_j^{(t)} - \mathbf{p}_i^{(t)}\big\| < r_i^{(t)} \;\wedge\; \ell_j^{(t)} > \ell_i^{(t)} \Big\} \tag{4}$$

When $\mathcal{N}_i^{(t)} \neq \emptyset$, agent i moves toward the luminosity-weighted barycenter of its brighter neighbors:

$$\bar{\mathbf{p}}_i^{(t)} \;=\; \sum_{j \in \mathcal{N}_i^{(t)}} w_{ij}^{(t)}\, \mathbf{p}_j^{(t)}, \qquad w_{ij}^{(t)} \;=\; \frac{\ell_j^{(t)} - \ell_i^{(t)}}{\sum_{k \in \mathcal{N}_i^{(t)}} \big(\ell_k^{(t)} - \ell_i^{(t)}\big)} \tag{5}$$

$$\mathbf{p}_i^{(t+1)} = \mathbf{p}_i^{(t)} + \delta \cdot \frac{\bar{\mathbf{p}}_i^{(t)} - \mathbf{p}_i^{(t)}}{\|\bar{\mathbf{p}}_i^{(t)} - \mathbf{p}_i^{(t)}\|} \tag{6}$$

where δ denotes the movement step size. Agents without brighter neighbors remain stationary at the current iteration.

Adaptive Sensor Range. The sensor range of each agent is updated at every iteration to regulate the local neighborhood density:

$$r_i^{(t+1)} = \min\Big(r_s,\; \max\big(0,\; r_i^{(t)} + \beta\,(n_t - |\mathcal{N}_i^{(t)}|)\big)\Big) \tag{7}$$

where n_t is the target neighborhood size and β controls the adaptation rate. This mechanism allows agents in sparsely populated regions to extend their sensor range in search of brighter neighbors, while agents in densely populated regions contract it to focus on the local cluster, supporting the multi-modal convergence property of GSO.

Cluster Extraction. After T iterations, DBSCAN [6] is applied to the converged swarm configuration $\{\mathbf{p}_i^{(T)}\}_{i=1}^{P}$ with neighborhood radius ε and minimum cluster size $m_{\min}$. For each cluster C_k, the final detection center is assigned to the position of the highest-luciferin agent in the cluster:

$$\hat{\mathbf{c}}_k = \mathbf{p}_{i^*}^{(T)}, \qquad i^* = \arg\max_{i \in C_k} \ell_i^{(T)} \tag{8}$$

while the cluster confidence is computed as:

$$\hat{s}_k = \sum_{i \in C_k} \big(\ell_i^{(T)}\big)^2 \tag{9}$$

The quadratic aggregation in Eq. (9) rewards dense clusters of high-luciferin agents, amplifying the collective evidence of well-supported detections relative to isolated or weakly-supported activations. The resulting detections therefore encode both the spatial convergence of the swarm dynamics and the collective confidence emerging from neighboring candidate interactions.

Algorithm 1 Swarm-Based Decoding for Pixel-Based Detection

Require: Candidate field $\{(\mathbf{p}_i, s_i)\}_{i=1}^{P}$; parameters $\rho, \gamma, \delta, \beta, n_t, r_s, T, \varepsilon, m_{\min}$

Initialize $\ell_i^{(0)} \leftarrow s_i,\ r_i^{(0)} \leftarrow r_s$ for all i
for $t = 1, \ldots, T$ **do**
  Update luciferin $\ell_i^{(t)}$ via Eq. (3) for all i
  Compute neighborhoods $\mathcal{N}_i^{(t)}$ via Eq. (4)
  Update positions $\mathbf{p}_i^{(t+1)}$ via Eqs. (5)–(6)
  Update sensor ranges $r_i^{(t+1)}$ via Eq. (7)
end for
Apply DBSCAN$(\varepsilon, m_{\min})$ to $\{\mathbf{p}_i^{(T)}\}_{i=1}^{P}$

Ensure: Detections $\{(\hat{\mathbf{c}}_k, \hat{s}_k)\}_{k=1}^{K}$ via Eqs. (8)–(9)

4 Experimental Setup

4.1 Dataset

The experimental evaluation is conducted on the publicly available Cervix93 dataset [1], a cervical cytology benchmark for nuclei localization. The dataset consists of 93 image stacks acquired at Moffitt Cancer Center (Tampa, FL) using an integrated microscope system (Stereologer, SRC Biosciences) at 40× magnification. Each stack contains between 10 and 20 images acquired at equally spaced focal planes, with a resolution of 1280×960 pixels. The dataset contains a total of 2,705 manually annotated nuclei, each represented by the spatial coordinates of its center.

This dataset provides a particularly suitable setting for evaluating decoding strategies in dense detection regimes due to the presence of small and densely distributed nuclei, which naturally generate highly redundant candidate localization fields during inference.

4.2 Detector Configuration

The experimental evaluation is conducted using GravityNet [18], a gravity-point-based detector designed for nuclei localization in cytological imaging. The detector employs a ResNet-152 backbone [8] with the gravity-point configuration defined by step 15, as adopted in [19] for the same dataset.

The network is trained for 50 epochs using the Adam optimizer with learning rate 10^{-4} and batch size 8. Training is performed on full-resolution images with min-max intensity normalization using the default split provided by the author of the dataset [1]. All experiments are conducted on an NVIDIA V100 GPU.

The underlying detection network remains identical across all decoding conditions in order to isolate the effect of the decoding strategy from the detector architecture itself.

4.3 Swarm Decoder Configuration

Three decoding configurations are considered in the experimental evaluation.

The first configuration does not apply any post-processing strategy, directly emitting candidate predictions above a confidence threshold as final detections. This setting provides a reference condition for analyzing the redundancy of the raw candidate localization field.

The second configuration employs Non-Maximum Suppression (NMS), performing greedy candidate suppression based on confidence ranking and spatial proximity using suppression radius r_{NMS}.

The third configuration corresponds to the proposed swarm-based decoder. The decoder uses luciferin dynamics parameters $\rho = 0.4$ and $\gamma = 0.6$, neighborhood adaptation parameters $\beta = 0.08$, $n_t = 5$, and $r_s = r_{\mathrm{NMS}}$, movement step size $\delta = r_s/10$, number of swarm iterations $T = 25$, and DBSCAN clustering parameters $\varepsilon = r_s/3$ and $m_{\min} = 2$.

The spatial parameters governing the swarm geometry are coupled to the suppression radius r_{NMS} of the baseline ($r_s = r_{\mathrm{NMS}}$, $\delta = r_s/10$, $\varepsilon = r_s/3$), so that no spatial scale is introduced beyond the one already shared with the NMS configuration. The remaining luciferin and adaptation parameters (ρ, γ, β, n_t, $m_{\min}$) were fixed a priori to values yielding stable local swarm interactions, and the number of iterations T was set to operate in the converged regime of the dynamics (Section 5.1). No parameter was tuned on the test set, and the same configuration is used throughout all experiments.

The same detector predictions are used across all decoder configurations in order to ensure that performance differences are exclusively attributable to the decoding stage.

4.4 Evaluation Protocol

Performance is evaluated using precision, recall, and F1-score computed at different False Positives per Image (FPpI) operating points. A predicted localization is considered a true positive if its Euclidean distance from the nearest unmatched ground-truth nucleus center is at most 10 pixels, corresponding approximately to the average nucleus radius in the dataset.

The analysis is conducted over FPpI operating points ranging from 1 to 10, allowing the evaluation of the different decoding strategies under progressively less restrictive false positive regimes. Precision, recall, and F1-score are computed at each operating point in order to analyze the trade-off between candidate aggregation and false positive control.

The evaluation protocol is applied identically across all decoding configurations to ensure a fair comparison between suppression-based and swarm-based decoding strategies.

5 Results and Discussion

Table 1 reports quantitative results at selected FPpI operating points, while Fig. 1 shows the complete Precision, Recall, and F1-score curves across FPpI values from 1 to 10.

Table 1. Precision, Recall, and F1-score at FPpI operating points from 1 to 10. Best result per column in bold.

Decoder	@1	@2	@3	@4	@5	@6	@7	@8	@9	@10
					F1-score					
No post-processing	0.892	0.908	0.909	0.903	0.893	0.882	0.873	0.870	0.853	0.853
NMS	**0.923**	**0.936**	**0.930**	0.920	0.907	0.882	0.882	0.860	0.850	0.820
Swarm (GSO)	0.911	0.928	0.929	**0.925**	**0.913**	**0.896**	**0.886**	**0.882**	**0.869**	**0.853**
					Precision					
No post-processing	0.963	0.932	0.904	0.882	0.852	0.828	0.805	0.797	0.765	**0.765**
NMS	**0.964**	**0.933**	0.901	0.879	0.851	0.804	0.804	0.769	0.752	0.705
Swarm (GSO)	0.963	**0.933**	**0.905**	**0.889**	**0.856**	**0.826**	**0.806**	**0.800**	**0.778**	0.751
					Recall					
No post-processing	0.831	0.886	0.914	0.924	0.939	0.945	0.954	0.957	0.963	0.963
NMS	**0.885**	**0.940**	**0.961**	**0.965**	0.972	0.975	0.975	0.976	0.978	0.979
Swarm (GSO)	0.865	0.923	0.954	0.964	**0.977**	**0.980**	**0.983**	**0.984**	**0.985**	**0.987**

The experimental analysis reveals a consistent operating-point trade-off between suppression-based and swarm-based decoding strategies. In the low-FPpI regime, NMS achieves the highest F1-score, reaching 0.923 at FPpI= 1. This behavior is consistent with the greedy confidence-ranking mechanism of NMS, which favors highly confident and spatially isolated candidates under restrictive false positive budgets.

As the FPpI operating point increases, the proposed swarm-based decoder progressively surpasses NMS in both precision and recall, resulting in improved F1-score at medium and high FPpI regimes. At FPpI= 10, the proposed decoder achieves F1-score = 0.853, precision = 0.751, and recall = 0.987, outperforming NMS across all three metrics. The simultaneous improvement in precision and recall suggests that the swarm dynamics effectively exploit collective spatial evidence emerging from neighboring candidate predictions rather than relying exclusively on individual candidate confidence scores.

The no post-processing baseline achieves consistently high recall but substantially lower precision, reflecting the strong redundancy of the raw candidate localization field. Both NMS and the proposed decoder significantly improve precision through spatial aggregation, confirming the importance of the decoding stage in dense localization settings. However, the observed differences between NMS and the proposed approach indicate that the aggregation mechanism itself plays a central role. While NMS performs greedy pairwise suppression, the swarm-based decoder progressively reorganizes neighboring candidates through iterative local interactions, enabling the emergence of coherent spatial structures corresponding to object hypotheses.

These observations support the interpretation of dense candidate localization fields as structured spatial systems rather than collections of independent

Table 2. Relative F1 variation (ΔF1) with respect to $T = 0$ (clustering-only reference, no movement) at selected FPpI operating points.

T	**Δ@1**	**Δ@3**	**Δ@5**	**Δ@10**	**mean (1–10)**
0 (ref.)	—	—	—	—	—
5	+0.000	+0.004	−0.001	+0.034	+0.005
10	+0.005	+0.006	−0.003	+0.030	+0.004
15	+0.008	+0.009	−0.002	+0.032	+0.006
20	+0.008	+0.009	−0.002	+0.033	+0.007
25	+0.009	+0.010	−0.001	+0.033	+0.007

predictions. In particular, the improved behavior of the swarm-based decoder at higher FPpI regimes suggests that collective aggregation mechanisms can preserve informative candidate structures that may be partially discarded by purely local suppression strategies. This behavior is especially relevant in dense detection scenarios characterized by spatially correlated candidate distributions, where neighboring predictions may jointly encode evidence about the underlying target configuration.

5.1 Effect of the Swarm Movement Dynamics

The proposed decoder combines two distinct stages: the iterative agent movement and the final density-based clustering. To isolate the contribution of the movement dynamics from the clustering stage, we conduct a controlled analysis varying the number of swarm iterations T, keeping all remaining components and the underlying candidate predictions fixed. The setting $T = 0$ corresponds to applying DBSCAN directly to the initial candidate positions, without any movement dynamics, and therefore serves as a clustering-only reference. Table 2 reports the relative F1 variation (ΔF1) with respect to this reference at representative operating points and averaged over the FPpI range 1–10.

Two observations emerge. First, the movement dynamics yield a consistent but moderate improvement over the clustering-only reference, with an average ΔF1 of +0.007 at $T = 25$. The improvement is not uniform across operating points: it is negligible in the low- and mid-FPpI regimes and concentrated at high FPpI, reaching +0.033 at FPpI= 10. This is consistent with the interpretation that the iterative reorganization mainly recovers informative candidate structures in the less restrictive false-positive regimes, while the clustering stage alone already accounts for most of the performance under restrictive budgets. Second, the configuration stabilizes well before the adopted horizon: the mean absolute F1 change between consecutive iteration counts drops below 0.001 beyond $T = 15$, indicating that the swarm reaches a stable configuration and that the choice $T = 25$ operates in the converged regime rather than at an arbitrary cut-off.

5.2 Limitations

The current study is limited to a single application domain and detector configuration, and additional evaluations across different detection architectures and candidate field distributions will be necessary to further assess the generality of the proposed formulation.

Furthermore, the comparison is restricted to suppression-based decoding and does not include alternative collective aggregation strategies, such as mean-shift-based approaches, which could help isolate the contribution of swarm dynamics from clustering-based aggregation mechanisms more broadly. Finally, the current implementation still relies on manually defined hyperparameters regulating neighborhood interactions and swarm evolution dynamics.

From a computational standpoint, the decoder is iterative: each of the T iterations requires, in the worst case, the evaluation of pairwise neighborhood relations among the P candidates, leading to a worst-case complexity of $\mathcal{O}(T\,P^2)$, followed by a final DBSCAN pass. This contrasts with the single-pass nature of greedy suppression and introduces an overhead that scales with the number of iterations and the density of the candidate field. While the analysis in Section 5.1 shows that the configuration converges well before the adopted horizon, so that a smaller T can be used without substantial loss, the accuracy gains reported here must be weighed against this additional cost; a detailed empirical characterization of decoding time across decoding strategies and hardware is left for future work.

6 Conclusions

This work introduced a swarm-based decoding strategy inspired by GSO for dense pixel-based detection. By modeling candidate predictions as interacting agents, the proposed decoder reformulates detection decoding as a collective spatial aggregation problem operating over dense candidate localization fields.

Experimental results on cervical cytology nuclei localization demonstrated a consistent operating-point trade-off between suppression-based and swarm-based decoding strategies. While NMS achieved superior performance in low-FPpI regimes, the proposed decoder obtained improved precision-recall behavior at medium and high FPpI operating points, suggesting that collective swarm dynamics can effectively exploit spatial evidence emerging from correlated candidate predictions.

Future work will investigate alternative spatial interaction models, adaptive neighborhood formulations, and structured interaction topologies aimed at reducing hyperparameter sensitivity and computational complexity in dense candidate aggregation settings.

7 Acknowledgment

This research was funded by the PNRR Mission 4, Component 2, Investment 1.2 Research Project MINERVA (CUP H33C25000280001).

References

1. Ahmady Phoulady, H., Mouton, P.R.: A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection. arXiv e-prints (2018). https://doi.org/10.48550/arXiv.1811.09651
2. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-nms — improving object detection with one line of code. In: 2017 IEEE International Conference on Computer Vision (ICCV). p. 5562–5570. IEEE (2017). https://doi.org/10.1109/ICCV.2017.593
3. Casas-Ordaz, A., Haro, E.H., Beltran, L.A., Alvarez, O., Mousavirad, S.J., Pérez-Cisneros, M., Oliva, D.: Particle swarm optimization: A survey of innovations over the last 10 years. Computer Science Review **60**, 100910 (2026). https://doi.org/10.1016/j.cosrev.2026.100910
4. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(5), 603–619 (2002). https://doi.org/10.1109/34.1000236
5. Dorigo, M., Gambardella, L.: Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation **1**(1), 53–66 (1997). https://doi.org/10.1109/4235.585892
6. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. p. 226–231 (1996)
7. Gad, A.G.: Particle swarm optimization algorithm and its applications: A systematic review. Archives of Computational Methods in Engineering **29**(5), 2531–2561 (2022). https://doi.org/10.1007/s11831-021-09694-4
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
9. Hosang, J., Benenson, R., Schiele, B.: Learning non-maximum suppression. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). p. 6469–6477. IEEE (2017). https://doi.org/10.1109/CVPR.2017.685
10. Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R.: A survey of deep learning-based object detection. IEEE Access **7**, 128837–128868 (2019). https://doi.org/10.1109/ACCESS.2019.2939201
11. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks. vol. 4, p. 1942–1948 vol.4 (1995). https://doi.org/10.1109/ICNN.1995.488968
12. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks. vol. 4, p. 1942–1948 vol.4 (1995). https://doi.org/10.1109/ICNN.1995.488968
13. Krishnanand, K.N., Ghose, D.: Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions. Swarm Intelligence **3**(2), 87–124 (2009). https://doi.org/10.1007/s11721-008-0021-5
14. Neubeck, A., Van Gool, L.: Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR'06). vol. 3, p. 850–855 (2006). https://doi.org/10.1109/ICPR.2006.479
15. Nikouei, M., Baroutian, B., Nabavi, S., Taraghi, F., Aghaei, A., Sajedi, A., Moghaddam, M.E.: Small object detection: A comprehensive survey on challenges, tech-

niques and real-world applications. Intelligent Systems with Applications **27**, 200561 (2025). https://doi.org/10.1016/j.iswa.2025.200561
16. Omran, M.G.H., Engelbrecht, A.P., Salman, A.: Particle Swarm Optimization for Pattern Recognition and Image Processing, p. 125–151. Springer (2006). https://doi.org/10.1007/978-3-540-34956-3_6
17. Reynolds, C.W.: Flocks, herds and schools: A distributed behavioral model. SIGGRAPH Comput. Graph. **21**(4), 25–34 (1987). https://doi.org/10.1145/37402.37406
18. Russo, C., Bria, A., Marrocco, C.: Gravitynet for end-to-end small lesion detection. Artificial Intelligence in Medicine **150**, 102842 (2024). https://doi.org/10.1016/j.artmed.2024.102842
19. Russo, C., Tanriverdi, Y.B., Bria, A., Marrocco, C.: A pixel-based anchor approach for nuclei detection in cervical cytology imaging. In: Pattern Recognition. ICPR 2024 International Workshops and Challenges. p. 268–278. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-031-87660-8_20
20. Solovyev, R., Wang, W., Gabruseva, T.: Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing **107**, 104117 (2021). https://doi.org/10.1016/j.imavis.2021.104117

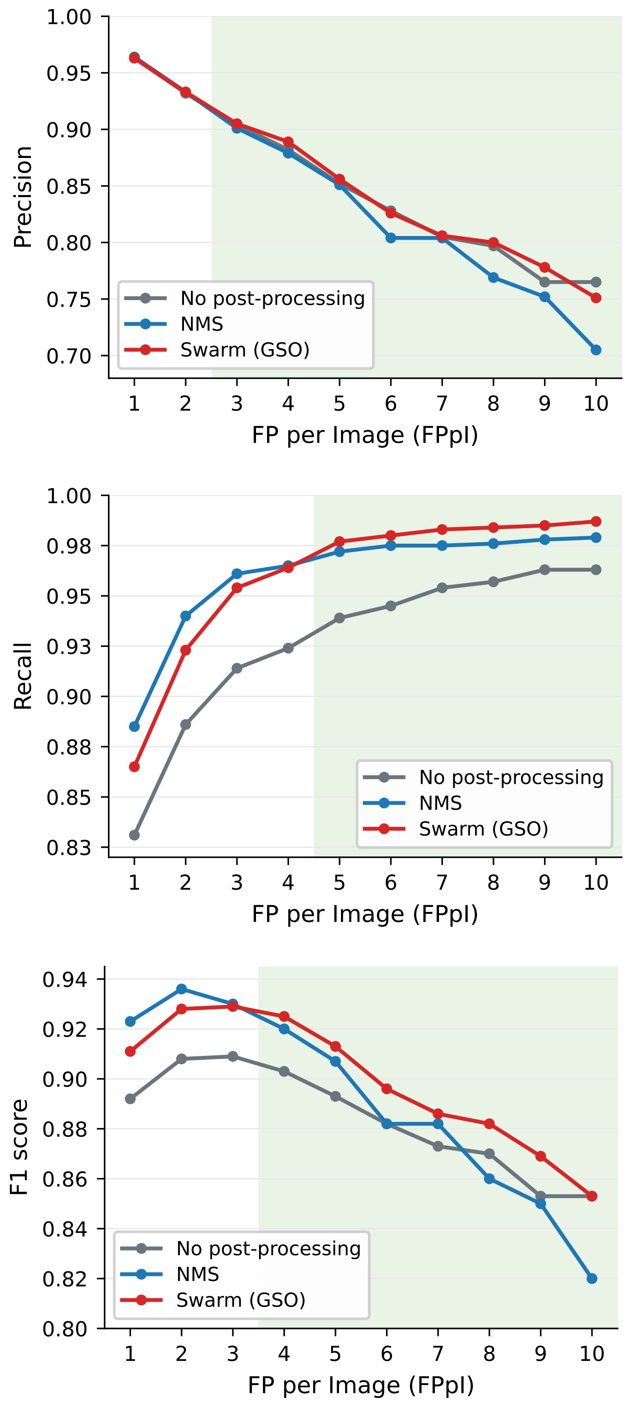

Fig. 1. Precision, Recall, and F1-score curves as a function of FPpI for the evaluated decoding strategies.