YOLOv11-SAMNet: A Hybrid Detection and Segmentation Framework for Urine Sediment Analysis

Sania Akhtar[1(✉)], Muhammad Hanif[1], Hamdi Melih Saraoglu[2], Sham Lal[3], and Muhammad Waqas Arshad[4(✉)]

[1] Aerial Robotics and Vision Laboratory, GIK Institute of Engineering Sciences and Technology, Topi 23460, Pakistan
{sania.akhtar,Muhammad.hanif}@giki.edu.pk
[2] Department of Electrical and Electronics, Kutahya Dumlupinar University, Kutahya, Turkey
hmelih.saraoglu@dpu.edu.tr
[3] Faculty of Computing and AI, Air University, Islamabad, Pakistan
231183@students.au.edu.pk
[4] Department of Computer Science and Engineering, University of Bologna, 40126 Bologna, Italy
muhammadwaqas.arsha2@unibo.it

Abstract. Urinalysis plays a vital role in diagnosing kidney diseases and urinary tract infections, but conventional manual analysis is often subjective, time-intensive, and prone to variability. Existing detection and segmentation methods struggle with accurately identifying small and irregularly shaped urine sediment particles, leading to inconsistencies in results. To overcome these limitations, we introduce YOLOv11-SAMNet, a novel one-stage instance segmentation model that combines the YOLOv11 detector with the Segment Anything Model (SAM) for 14 classes. Our approach enhances both detection precision and segmentation accuracy by leveraging a feature extraction backbone for capturing key image details, a multi-scale fusion neck for integrating diverse feature representations, and a detection and segmentation head to refine instance segmentation, followed by a bounding box. The model performs exceptionally well on well-defined structures, such as Leukocytes (94.3% *mAP@50*) and Epithelial cells (92.5% *mAP@50*), but encounters challenges when segmenting small and irregular elements like bacteria (33.8% *mAP@50*) and mucus (27.2% *mAP@50*). To further enhance model performance, we developed an expert-assisted data preprocessing pipeline to improve the quality of training data. Our findings highlight the potential of YOLOv11-SAMNet in automating urinalysis, offering a more efficient and objective alternative to traditional manual analysis, ultimately contributing to improved clinical diagnostics.

Keywords: Yolov11 · Urine Sediment Analysis · SAM · Detection · Segmentation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15899, pp. 241–251, 2026.
https://doi.org/10.1007/978-3-031-97663-6_21

1 Introduction

Urinalysis plays a vital role in diagnosing kidney diseases and urinary tract infections (UTIs), helping detect conditions such as leukocyturia, proteinuria, and hematuria [1]. Microscopic examination of urinary sediment, comprising red blood cells (RBCs), white blood cells (WBCs), crystals, and other particles—provides essential insights into underlying medical conditions [2]. However, manual sediment analysis is often time-consuming, labor-intensive, and prone to observer variability, underscoring the need for reliable automated solutions. In recent years, deep learning has emerged as a powerful tool for automating urine sediment analysis, particularly through object detection and segmentation techniques. Nonetheless, most existing methods focus exclusively on either detecting particles or segmenting their boundaries, rarely addressing both simultaneously. This narrow focus limits the diagnostic value of such systems, as effective urinalysis requires not only accurate localization but also detailed morphological interpretation of sediment particles. For instance, previous studies using YOLOv5 variants [3] have achieved notable detection performance but did not incorporate instance-level segmentation, leaving a critical gap in comprehensive analysis.

To overcome this limitation, we propose YOLOv11-SAMNet, a novel one-stage instance segmentation model that seamlessly integrates the YOLOv11 detector with the Segment Anything Model (SAM). Unlike two-stage approaches such as Mask R-CNN, which achieve high accuracy at the cost of complex post-processing and longer inference times, our method offers a real-time, end-to-end solution that is well-suited for clinical workflows. One-stage instance segmentation models are gaining momentum in medical imaging for their ability to balance speed and precision, qualities essential for time-sensitive diagnostics. The proposed framework further incorporates expert-assisted data preprocessing strategies, including label verification, image quality enhancement, and data validation. These measures help ensure the robustness of the training dataset, thereby improving both detection and segmentation outcomes. YOLOv11-SAMNet features a three-part architecture: a backbone for hierarchical feature extraction, a neck for multi-scale feature fusion, and a head designed for simultaneous detection and segmentation. To the best of our knowledge, this is the first one-stage YOLO-based model tailored specifically for 14 urine sediment classes, achieving significant improvements in classification and localization accuracy. The key contributions of this paper are as follows:

- **Development of YOLOv11-SAMNet:** We present a novel one-stage instance segmentation framework that integrates the YOLOv11 object detector with the Segment Anything Model (SAM). This hybrid design is tailored to provide accurate and efficient segmentation of urine sediment particles in microscopic imagery.
- **Expert-Assisted Data Pre processing:** We introduce a systematic preprocessing pipeline involving domain expert validation, accurate labeling, quality enhancement, and data integrity checks to ensure high-quality inputs that enhance overall model performance.

– **Multi-Class Segmentation for Urine Sediment Analysis:** This work pioneers the application of a one-stage YOLO-based instance segmentation framework for 14 distinct urine sediment classes, significantly enhancing the accuracy of both detection and classification.

The remainder of this paper is organized as follows: Sect. 2 reviews the relevant literature. Section 3 outlines the proposed methodology and model architecture. Section 4 presents the output generation process. Section 5 discusses the experimental setup and results. Finally, Sect. 6 concludes the paper with key findings and future directions.

2 Related Work

One-stage models like YOLO perform detection and segmentation in a single pass, offering speed and efficiency. In contrast, two-stage models such as Mask R-CNN provide higher accuracy but with greater computational cost. Our proposed YOLOv11-SAMNet integrates the speed of one-stage detection with the segmentation precision of SAM, aiming for balanced, real-time performance in urinalysis. Table 1 and 2 One-stage models like YOLO perform detection and segmentation in a single pass, offering speed and efficiency.

Table 1. Summary of One-Stage Instance Segmentation Methods for Cell Segmentation

Reference	Problem Addressed	Method Used	Dataset/Domains	Performance
[1]	Slow, subjective manual urine sediment analysis; existing methods lack accuracy and speed for small, dense targets	YOLOv5n with FCN-enhanced head, multi-scale feature fusion	Custom urine sediment dataset (500 images)	mAP50 = 91.8%, FPS = 63.3
[5]	Speed and efficiency limitations in cell segmentation	Cell-DETR (Transformer + CNN)	Cell segmentation dataset	Jaccard 0.82 (A), 0.84 (B)
[6]	Heterogeneity in breast tumor cells	SOLOv2 for tumor cell instance segmentation	Breast tumor dataset	mAP_{50} = 68.00%

Table 2. Summary of Two-Stage Instance Segmentation Methods for Cell Segmentation

Reference	Problem Addressed	Method Used	Dataset/Domains	Performance
[7]	Weak boundaries, overlapping cells	Mask R-CNN with anchor box adjustment	Cancer cell dataset	IoU 87.9%
[8]	Variation in cell shape	Mask R-CNN with contour refinement	Live cell dataset	F1-score 89.5%
[9]	Clumped objects, high overlap	Mask R-CNN with attention mechanism	Histopathology images	IoU 92.3%
[10]	Small objects in microscopy images	Mask R-CNN with feature fusion	Fluorescence microscopy dataset	Mean IoU 84.6%, Error Rate 2.7%

3 Methodology

The proposed approach, illustrated in Fig. 1, integrates YOLOv11 with the Segment Anything Model (SAM) to enable accurate instance segmentation of urine sediment particles. It comprises three main stages: expert-assisted data preprocessing, feature extraction, and instance segmentation.

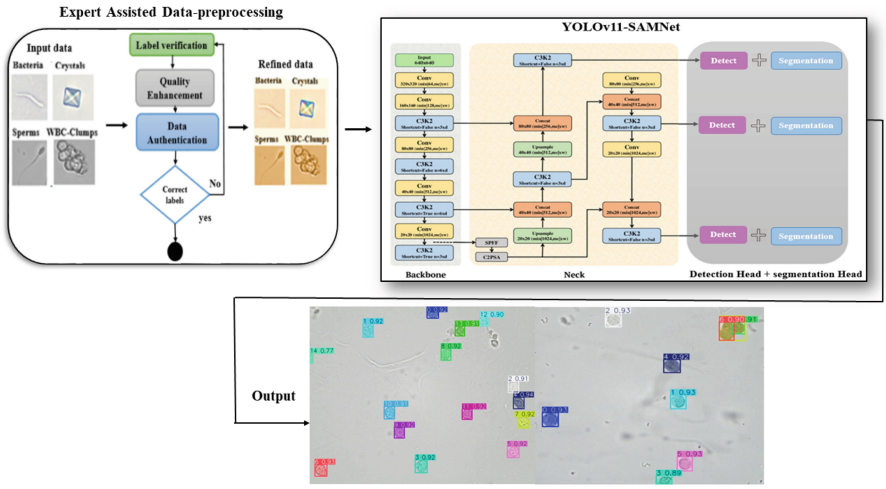

Fig. 1. Workflow of the proposed methodology

3.1 Expert-Assisted Data Preprocessing

The first stage ensures high-quality data labeling and refinement through verification by medical experts, allowing for the correction of inaccurate labels before advancing to the next phase.

3.2 YOLOv11-SAMNet Architecture

The YOLOv11-SAMNet architecture is composed of three key components: the backbone, neck, and head.

Backbone Network: The Conv Block processes input of size $C \times H \times W$ using a 2D convolution, Batch Normalization, and SiLU activation.

Neck Network: The neck network enhances feature fusion across multiple scales to improve detection and segmentation performance. It employs upsampling and down sampling operations to integrate shallow and deep features.

Head Network: The head network comprises two primary branches:

- Object Detection Head: Responsible for bounding box regression and object classification. It includes three detection heads operating at different spatial resolutions (32×32, 64×64, and 128×128), as shown in Fig. 1.
- Instance Segmentation Head: This branch integrates the Segment Anything Model (SAM) for precise instance-level segmentation. It processes features from the C3 layers of the neck using a lightweight fully convolutional network

(FCN) composed of three convolutional layers followed by an upsampling module. The FCN refines deep feature maps to generate high-resolution segmentation masks.

The overall loss function of YOLOv11-SAMNet consists of four components and is defined as follows:

$$L_{\text{total}} = L_{\text{box}} + L_{\text{cls}} + L_{\text{obj}} + L_{\text{seg}}. \tag{1}$$

where L_{box} is the bounding box regression loss, L_{cls} is the classification loss, L_{obj} is the object confidence loss, and L_{seg} is the segmentation loss.

3.3 Bounding Box Regression Loss

The bounding box regression in YOLOv11 utilizes the Complete Intersection over Union (CIoU) loss, which integrates additional components to account for aspect ratio consistency and center point distance, thereby improving localization accuracy:

$$L_{\text{bbox}} = 1 - \text{CIoU}(d, d_{gt}) + \frac{\psi^2(d, d_{gt})}{s^2} + \gamma u. \tag{2}$$

where d and d_{gt} represent the predicted and ground truth bounding box coordinates, respectively; ψ^2 denotes the squared Euclidean distance between the center points; s is the diagonal length of the smallest enclosing box; and u is a penalty term for aspect ratio discrepancy, modulated by a weighting factor γ.

3.4 Classification and Object Loss

In YOLOv11, both the classification lossL_{class} and $L_{\text{obj_conf}}$ are computed using Distribution Focal Loss (DFL). This loss function is specifically designed to enhance classification accuracy in dense object detection tasks by modeling the distribution over continuous regression targets. The DFL effectively balances confidence learning and class prediction, and is defined as:

$$L_{\text{class}} = -\frac{1}{M} \sum_{j=1}^{M} \Big[(1 - q_j)^{\delta} z_j \log(q_j) \Big]. \tag{3}$$

where M is the total number of predictions, q_j is the predicted probability of class j, z_j is the ground truth label, and δ is the focusing parameter that down-weights easy examples to focus on hard-to-classify instances.

Similarly, the object confidence loss is formulated as:

$$L_{\text{obj_conf}} = -\frac{1}{M} \sum_{j=1}^{M} \Big[z_j \log(\sigma(r_j)) + (1 - z_j) \log(1 - \sigma(r_j)) \Big]. \tag{4}$$

where r_j is the objectness score for the predicted bounding box, and $\sigma(.)$ is the sigmoid function.

3.5 Segmentation Loss

The segmentation loss incorporates pixel-wise loss with instance-based refinement from SAM-generated masks:

$$L_{\text{seg}} = -\frac{1}{M}\sum_{k=1}^{M}\frac{1}{B_k}\sum_{l=1}^{D_k}\left[\beta g_k^r \log(g_k^q) + (1-\beta)(1-g_k^r)\log(1-g_k^q)\right] \times v_l \times \text{Trim}(c_l). \tag{5}$$

where B_k is the area of the predicted mask, g_k^r and g_k^q represent the ground truth and predicted masks, β is a balancing factor, v_l is a weight term emphasizing boundary regions, and $\text{Trim}(c_l)$ ensures loss computation only within the bounding box:

$$\text{Trim}(c_l) = \begin{cases} 1, & \text{if } c_l \in S_k \text{ and } g_k^q > \delta \\ 0, & \text{otherwise} \end{cases}. \tag{6}$$

where S_k is the predicted bounding box, and δ is a mask confidence threshold to filter low-confidence regions.

4 Output Generation

The final output of YOLOv11-SAMNet, illustrated in Fig. 4, includes:

- Bounding boxes annotated with classification scores for detected urine sediment particles.
- High-resolution instance segmentation masks for each identified particle, enabling precise localization and shape delineation.

This integrated methodology delivers a robust and real-time solution for automated urine sediment analysis, effectively balancing detection precision and segmentation accuracy for clinical use.

5 Experimental Analysis

5.1 Dataset

The dataset used in this study, as described in [2], was carefully curated with meticulous annotation and validation by medical professionals to ensure accuracy. Labeling was performed using the Roboflow platform, which enabled the creation of high-quality annotations essential for reliable model training and evaluation. Over a period of five months, 7,201 microscopic images were annotated, representing 14 distinct classes of urine sediment. These classes include Miscellaneous, Bacteria, Mucus, Budding Yeast, Sperm, Yeast, RBC-Clumps, WBC-Clumps, Cast, Crystals, Leukocytes, Epith, Erythrocytes, Epithin, and Mycete. Figure 2 illustrates representative labeled samples from the dataset. To facilitate robust training and validation, the dataset was divided into 70% for training, 20% for validation, and 10% for testing. This stratified split ensures effective performance evaluation across all categories.

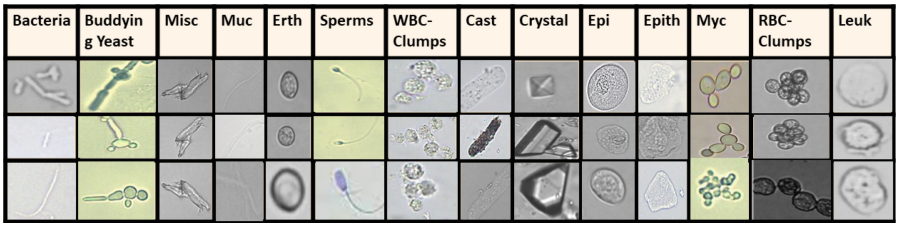

Fig. 2. Sample images labeled using the Roboflow

5.2 Performance Metrics and Implementation Details

In this study, we evaluate the performance of the YOLOv11-SAMNet model using several metrics: precision (P), recall (R), F1-score, and mean Average Precision at 0.5 IoU (*mAP@50*). These metrics are divided into two main categories: object detection quality and segmentation quality. Object detection quality is measured using precision (P(box)), recall (R(box)), and F1-score (F1(box)), while segmentation quality is assessed using precision (P(seg)), recall (R(seg)), and F1-score (F1(seg)). In contrast to models like Mask R-CNN, where detection quality is typically measured with P(box), R(box), and F1(box) while segmentation performance is evaluated using *mAP@50*, YOLOv11-SAMNet separately evaluates both tasks. The object detection performance is assessed using P(box), R(box), and F1(box), whereas segmentation performance is measured with P(seg), R(seg), and F1(seg). The YOLOv11-SAMNet model is trained using the AdamW optimizer, with a learning rate of 0.000556 and momentum set to 0.9. Weight decay parameters include 81 weight parameters with a decay of 0.0, 88 weight parameters with a decay of 0.0005, and 87 bias parameters with a decay of 0.0.

5.3 Experimental Results of YOLOv11-SAMNet

This section presents the performance results of the YOLOv11-SAMNet model across 14 classes of urine sediment, as summarized in Table 3. The model demonstrates its highest performance for Leukocytes (Leuko), achieving a recall of 94.6% and an *mAP@50* of 94.3%. It is closely followed by Epithelial cells (Epith), which reach an *mAP@50* of 92.5%. These results highlight the model's strong capability to accurately detect well-defined and common cell types. In contrast, bacteria and mucus exhibit the lowest performance, with *mAP@50* values of 38.0% and 27.2%, respectively. These lower results suggest difficulties in detecting irregularly shaped and smaller particles, likely caused by their overlapping structures and variability in appearance. Figure 3 illustrates the confusion matrix, providing deeper insights into the model's performance. It shows a high true positive rate for well-classified particles such as leukocytes (0.94) and epithelial cells (0.92). However, there is noticeable misclassification of smaller and less distinct particles. Specifically, bacteria (0.56) and mucus (0.17) exhibit

Table 3. Performance metrics per class

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95
all	1440	11082	0.593	0.576	0.589	0.295
Bacteria	42	1578	0.480	0.318	0.338	0.0977
Budding yeast	4	29	1.000	0.0478	0.384	0.107
Miscellaneous	66	95	0.466	0.537	0.484	0.204
Mucus	120	532	0.482	0.205	0.272	0.092
RBC-Clumps	16	116	0.526	0.534	0.494	0.177
Sperms	9	91	0.342	0.514	0.514	0.143
WBC-Clumps	118	298	0.540	0.595	0.566	0.216
Yeast	10	50	0.281	0.350	0.400	0.210
Cast	282	726	0.557	0.815	0.754	0.456
Cryst	172	342	0.760	0.854	0.858	0.545
Epith	293	1269	0.721	0.925	0.911	0.598
Epithn	85	124	0.746	0.855	0.871	0.390
Eryth	202	4695	0.904	0.822	0.860	0.507
Leuko	264	1185	0.782	0.946	0.943	0.590

significant confusion with the background and other sediment classes, indicating challenges in differentiating them from artifacts or noise present in the dataset. Furthermore, classes such as budding yeast (0.27) and miscellaneous particles (0.54) show moderate levels of confusion, suggesting that these categories could benefit from further refinement in future iterations.

5.4 Comparison With SOTA Models

In the comparison of state-of-the-art models for object instance segmentation, as shown in Table 4, our YOLOv11-SAMNet model demonstrates competitive performance on a specialized urine sediment dataset containing 14 distinct classes. While YOLOv5n achieves the highest *mAP@50* score of 91.8% on a 3-class dataset, our model attains a *mAP@50* of 60%, but shows strong precision metrics, with 93% bounding box precision and 85% segmentation precision. These results highlight the model's high confidence in positive detections, even with the more complex multi-class urine dataset. In comparison, other models such as Mask R-CNN and MaskDis R-CNN, both trained on the extensive 91-class COCO dataset, achieve mAP scores of 87.4% and 88.8%, respectively. Despite the difference in mAP performance, YOLOv11-SAMNet demonstrates its potential for specialized medical imaging applications, with precision and recall F1 scores of 82% for bounding boxes and 81% for segmentation tasks. To the best of our knowledge, this is the first work to apply instance segmentation on 14 distinct classes using the Urinalysis dataset. Moreover, there is limited research on

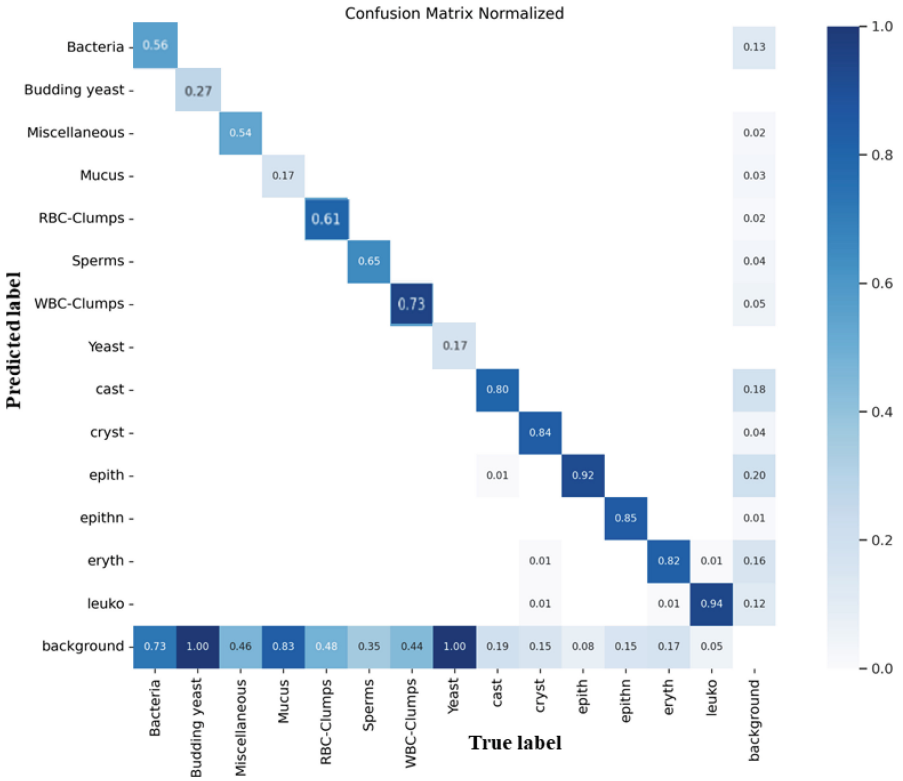

Fig. 3. Confusion matrix of the YOLOv11-SAMNet

instance segmentation in this domain, primarily due to the challenges associated with pixel-level labeling. Urine particles are typically small and share similar morphological characteristics, making precise annotation particularly challenging.

Figure 4 illustrates the prediction results of YOLOv11-SAMNet for urine sediment analysis. Each detected particle is enclosed within a bounding box, with distinct colors representing different sediment classes. The numerical values beside the bounding boxes correspond to the model's confidence scores for each detection. In the first image, the model accurately detects 65 Red Blood Cells (RBCs), highlighting its ability to localize small, densely packed particles. In the second image, the model identifies 21 epithelial cells, which are comparatively larger and more distinct. In the third image, the model successfully localizes 8 casts, demonstrating its proficiency in detecting elongated structures within urine samples.

It is important to note that particles of the same class do not share a uniform color. Instead, the visualization adopts an instance segmentation approach, where distinct colors are assigned to each individual particle based on its loca-

Table 4. Comparison of mAP and Inference Time Across Existing Object Instance Segmentation Methods

Method/Technique	Dataset	*mAP@50*	P(box)	R(box)	F1(box)	P(seg)	R(seg)	F1(seg)
YOLOV5n [1]	3 classes	91.8	84.5	89.5	86.8	86.5	91.5	88.8
Mask R-CNN [1]	91 classes (COCO dataset)	87.4	88.2	84.82	84.6	–	–	–
MaskDis R-CNN [1]	91 classes (COCO dataset)	88.8	91.5	85.9	88.4	–	–	–
YOLOv11-SAMNet	14 classes	**60**	**93**	**82**	51	**85**	**75**	**81**

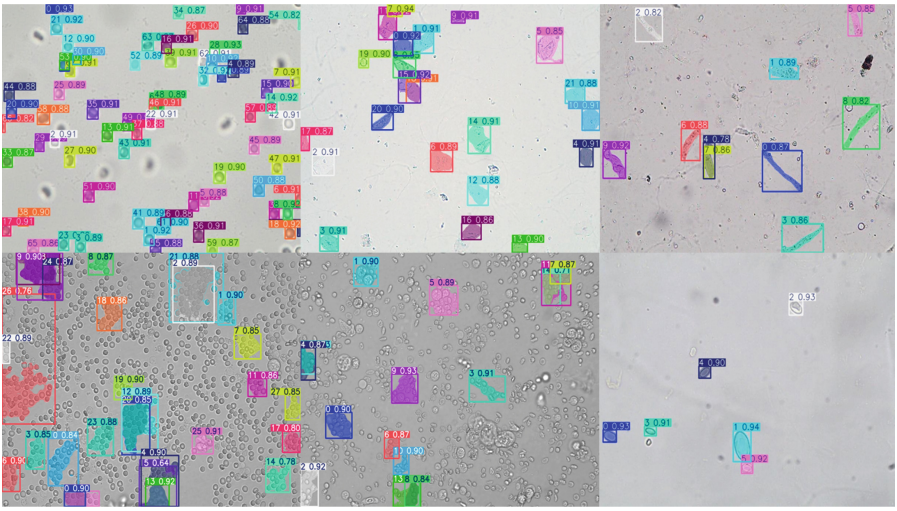

Fig. 4. Prediction results of YOLOv11-SAMNety

tion, not its class. This strategy facilitates the clear differentiation of overlapping or closely positioned objects, enhancing the interpretability of the detection results. These findings emphasize the effectiveness of YOLOv11-SAMNet in detecting and classifying various sediment types in microscopic images. However, challenges such as variations in image contrast and overlapping structures may still affect detection accuracy for certain classes.

6 Conclusion

In this study, we introduce YOLOv11-SAMNet, a novel one-stage instance segmentation model tailored for automated urine sediment analysis. By combining YOLOv11 with the Segment Anything Model (SAM), our approach significantly improves detection precision and segmentation accuracy across 14 distinct urine sediment classes. The model architecture includes a feature extraction backbone, a multi-scale fusion neck, and a specialized detection and segmentation head, optimizing instance segmentation performance. Experimental results show strong performance in detecting well-defined structures such as Leukocytes

(94.3% *mAP@50*) and Epithelial cells (92.5% *mAP@50*). However, challenges persist in segmenting smaller and irregularly shaped particles, such as bacteria (33.8% *mAP@50*) and mucus (27.2% *mAP@50*). To address these challenges, we developed an expert-assisted data preprocessing pipeline to refine the quality of training data, thus enhancing overall model performance. The encouraging results of YOLOv11-SAMNet demonstrate its potential to automate urinalysis, offering a more efficient and objective alternative to traditional manual analysis, ultimately supporting improved clinical diagnostics.

References

1. Tu, S., et al.: The urine formed element instance segmentation based on YOLOv5n. Sci. Rep. **14**(1), 28658 (2024)
2. Akhtar, S., et al.: An optimized data and model centric approach for multi-class automated urine sediment classification. IEEE Access (2024)
3. Suhail, K., et al.: Microscopic urinary particle detection by different YOLOv5 models with evolutionary genetic algorithm based hyperparameter optimization. Comput. Biol. Med. **169**, 107895 (2024)
4. Bai, B., et al.: YUSEG: YOLO and UNet is all you need for cell instance segmentation. Competitions in Neural Information Processing Systems. PMLR (2023)
5. Prangemeier, T., Reich, C., Koeppl, H.: Attention-based transformers for instance segmentation of cells in microstructures. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE (2020)
6. Priego-Torres, B.M., et al.: Deep learning-based instance segmentation for the precise automated quantification of digital breast cancer immunohistochemistry images. Expert Syst. Appl. **193**, 116471 (2022)
7. Ren, X., et al.: Mask-RCNN for cell instance segmentation. IEEE Trans. Med. Imaging (2020)
8. Loh, D.R., et al.: A deep learning approach to the screening of malaria infection: automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN. Comput. Med. Imaging Graph. **88**, 101845 (2021)
9. Qiu, X., et al.: Segmentation of multiple myeloma cells using feature selection pyramid network and semantic cascade mask R-CNN. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE (2022)
10. Mitate, E., et al.: Application of the sliding window method and Mask-RCNN method to nuclear recognition in oral cytology. Diagn. Pathol. **17**(1), 62 (2022)