Article | Received 21 February 2026; Revised 13 April 2026; Accepted 14 April 2026; Published 20 May 2026
https://doi.org/10.55092/aimat20260006

High-precision automated nanoparticle segmentation using a deep learning framework with boundary-aware and attention networks

Xiao Han[1,2], Linfeng Jin[3,*], Yijun Zhong[1,2,*] and Changfa Guo[1,2,*]

[1] Key Laboratory of the Ministry of Education for Advanced Catalysis Materials, College of Chemistry and Materials Science, Zhejiang Normal University, Jinhua 321004, China
[2] Zhejiang Key Laboratory of Advanced Catalysis and Adsorption Materials, College of Chemistry and Materials Science, Zhejiang Normal University, Jinhua 321004, China
[3] Department of Physics, College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China

* Correspondence authors; E-mails: jlf@zjnu.cn (L.J.); yjzhong@zjnu.edu.cn (Y.Z.); changfa.guo@zjnu.edu.cn (C.G.).

Highlights:

- A deep learning framework is developed for automated nanoparticle segmentation.
- Boundary-aware dynamic snake convolution captures complex nanoparticle boundaries.
- Bi-level routing attention enhances global context modeling and feature selection.
- A visualization system enables automated nanoparticle morphology analysis.

Abstract: Accurate characterization of nanoparticle geometry and morphology is essential for understanding their structure–property relationships. However, in most electron microscopy images, nanoparticles are densely distributed and are often affected by strong background noise and particle overlap, making conventional manual analysis time-consuming and inefficient. To address this issue, this study proposes Nanoparticle Segmentation You Only Look Once (NSYOLO), an enhanced deep learning–based instance segmentation framework for the automatic and high-precision recognition and segmentation of nanoparticles in electron microscopy images. The framework is trained on a multi-type dataset comprising nanocubes, nanospheres, and nanorods, and introduces a boundary-aware dynamic snake convolution (BADSConv) module to enhance boundary feature representation, along with a bi-level routing attention (BRA) mechanism to improve global feature modeling. Experimental results demonstrate that NSYOLO increases mean Average Precision (mAP)@0.5 from 0.906 to 0.957 and outperforms open-source automated tools, such as ImageJ and ImageDataExtractor, particularly in images with complex backgrounds and overlapping particles. In addition, the NSYOLO-based analysis system is developed to enable automated nanoparticle segmentation, size statistics, and the generation of editable Word reports without requiring any programming experience, thereby providing an efficient, reliable, and user-friendly solution for high-throughput nanoparticle morphology analysis.

Keywords: electron microscopy image; nanoparticle; instance segmentation; NSYOLO; deep learning

1. Introduction

Nanomaterials have attracted widespread attention in catalysis, energy, optics and biomedicine due to their unique physicochemical properties [1–5]. These properties are closely related to morphology, size, and surface characteristics; therefore, accurate morphological characterization of nanoparticles is essential for understanding and optimizing their performance. Transmission electron microscopy (TEM) and scanning electron microscopy (SEM) are widely used to observe and analyze nanomaterials because of their high-resolution imaging capabilities at the nanoscale [6–8]. However, conventional analysis approaches, such as ImageJ [9], often rely on manual measurement and annotation of individual nanoparticles. When dealing with images containing numerous particles, this procedure becomes labor-intensive and time-demanding, while also introducing variability associated with user experience and subjective judgment. As a result, the consistency and reproducibility of the measurements may be affected. These issues become more pronounced in high-throughput characterization tasks and large-scale statistical analyses, where efficient and reliable processing is essential.

With the rapid development of computer vision and deep learning technologies, automated analysis of microscopic images has attracted increasing attention in recent years. Earlier studies on microscopic image processing mainly employed conventional segmentation techniques, including Otsu thresholding [10], Canny edge detection [11], watershed segmentation [12], and k-means clustering [13]. These approaches generally distinguish foreground objects from the background according to grayscale intensity, edge information, or regional connectivity, offering advantages such as straightforward implementation and relatively low computational demand. Nevertheless, SEM/TEM images often contain nanoparticles with diverse morphologies, uneven intensity distributions, indistinct boundaries, and considerable background interference. Under such conditions, traditional methods frequently fail to extract complete particle contours and are strongly influenced by parameter selection, resulting in limited robustness and poor generalization across different imaging scenarios. By comparison, deep learning–based methods can learn hierarchical semantic and spatial representations directly from annotated datasets through end-to-end optimization [14], allowing more accurate segmentation in images with complicated backgrounds and weak contrast. Representative deep learning-based semantic segmentation methods include Fully Convolutional Network (FCN) [15], Unet and its variants [16−18], and the DeepLab series [19−21]. These methods achieve pixel-level classification; however, they are limited in distinguishing multiple instances of the same class. Instance segmentation methods, such as You Only Look At CoefficienTs (YOLACT) [22] and Mask Region-based Convolutional Neural Network (Mask R-CNN) [23], can classify individual particles, separate instances of the same class, and accurately delineate their boundaries. These methods are particularly suitable for segmenting densely distributed, similarly shaped, yet independent nanoparticles in SEM/TEM images.

In nanoparticle characterization, machine learning and deep learning have been widely applied to the automated analysis and feature extraction of microscopic image data. Existing studies have achieved substantial progress in algorithm performance and task adaptability. Sun *et al.* proposed a general framework based on the lightweight deep learning network Nanoparticle Segmentation Network (NSNet) for fast and accurate online statistical analysis of nanoparticle morphology in complex SEM/TEM images [24]. Kim *et al.* proposed a general method that combines computer vision and machine learning to automatically

extract nanoparticle size, size distribution, and morphology information from SEM images [25]. This methodology facilitates high-throughput analysis in complex scenarios characterized by overlapping particles, rod-shaped nanostructures, and core–shell architectures. Its efficacy has been substantiated through validation on SEM images that include nanoparticles with diverse geometries and size distributions. Glaubitz *et al.* [26] have further integrated machine learning models with cost-effective real-time characterization techniques, such as dynamic light scattering and ultraviolet–visible spectroscopy, to evaluate the size and morphology of gold nanoparticles. Despite these advancements, several practical challenges persist. Many existing tools necessitate intricate software installation and programming expertise, which can impede accessibility for researchers in materials science. Additionally, current methods typically extract only a limited set of morphological descriptors, thereby constraining the scope of characterization and their capacity to conduct comprehensive multi-parameter quantitative analysis of nanoparticles.

To tackle the identified challenges, this study presents an automated approach for nanoparticle segmentation and morphological analysis. The detailed workflow is shown in Figure 1. Initially, a high-quality dataset of electron microscopy images was assembled from previously published studies, featuring three key nanoparticle shapes: nanospheres, nanorods, and nanocubes (Figure 1a). The selected studies are predominantly well-cited and publicly available, ensuring the dataset's reliability and representativeness. Following this, a boundary-aware dynamic snake convolution (BADSConv) module was crafted, leveraging the You Only Look Once version 11 (YOLOv11) model [27]. This module, an advancement of the dynamic snake convolution (DSConv) [28], is adept at capturing intricate nanoparticle boundaries. Additionally, a bi-level routing attention (BRA) mechanism [29] was integrated, leading to the development of the Nanoparticle Segmentation You Only Look Once (NSYOLO) model for nanoparticle instance segmentation (Figure 1b). Finally, a system for analyzing nanoparticle geometric morphology was developed using the Python Flask framework (Figure 1c). Through a user-friendly web interface, users can conduct particle segmentation, detect scale bars, and perform quantitative analysis of morphological parameters, as well as create visualization charts and editable Word reports by uploading microscopy images. This provides a seamless automated analysis pipeline from image input to result output.

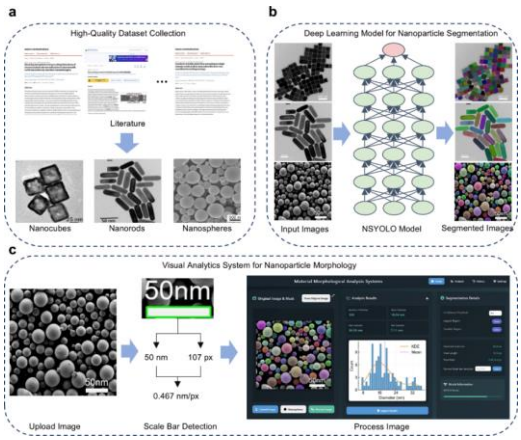

Figure 1. Overall workflow of the proposed automated nanoparticle segmentation and morphology analysis methods. **(a)** Construction of the electron microscopy image dataset of nanoparticles; **(b)** Deep learning–based Nanoparticle Segmentation You Only Look Once (NSYOLO) model for nanoparticle instance segmentation; **(c)** Workflow of the nanoparticle morphology analysis framework, including image upload, scale bar detection, and visualization.

2. Materials and methods

2.1. Dataset construction

The quality and diversity of datasets are pivotal in influencing the training and optimization processes of deep learning models. In this study, we developed a dataset comprising 721 electron microscopy images (SEM and TEM) of nanoparticles, which includes 267 images of nanocubes, 260 images of nanorods, and 277 images of nanospheres. These images were sourced from publicly available literature and represent nanoparticle morphologies that are extensively examined in the field of materials science, with appropriate citations to the original sources included throughout the manuscript. To standardize the input for the model while preserving structural integrity, each image was resized to maintain its original aspect ratio and subsequently padded to a resolution of 640 ×640 pixels. This approach avoids geometric distortion while ensuring consistent input dimensions for the model and has been adopted in You Only Look Once (YOLO) based framework [30]. Pixel-level annotations were manually generated using the open-source tool LabelMe [31], enabling accurate delineation of nanoparticle boundaries across different morphology types. The dataset was randomly split into training, validation, and testing subsets at a ratio of 8:1:1. This splitting strategy follows common practices in deep learning ensuring sufficient samples for training while maintaining reliable validation and unbiased performance evaluation. To avoid data leakage, these subsets were strictly separated, and data augmentation was applied only to the training set. To improve model robustness and generalization under diverse imaging conditions, multiple data augmentation strategies, including random rotation, horizontal and vertical flipping, cropping, and scaling, were employed. After augmentation, the training set is expanded, resulting in a total of 4226 images in the dataset. Representative SEM and TEM images in Supplementary Figure S1 a–f illustrate typical samples, including nanocubes, nanorods, and nanospheres synthesized under different experimental conditions. These samples exhibit variations in particle size, shape regularity, and imaging contrast, highlighting the morphological diversity and imaging complexity of the dataset.

2.2. Model architecture

We adopted YOLOv11n (where “n” denotes the lightweight base version, hereafter referred to as YOLOv11) as the baseline segmentation framework. YOLOv11 represents a recent generation within the YOLO series of object detection and segmentation models. Its overall architecture mainly consists of two components: a Backbone network for feature extraction and a Neck–Head structure for feature fusion and prediction. The Backbone is responsible for extracting multi-scale semantic and texture information from the input images and is composed of convolutional layers, C3K2 modules, a Spatial Pyramid Pooling–Fast (SPPF) module, and a C2PSA module. Among them, the C3K2 module is a Cross Stage Partial-based structure containing multiple convolutional operations for efficient feature learning; the SPPF module enlarges the receptive field through accelerated multi-scale spatial pooling; and the C2PSA module integrates attention mechanisms to strengthen feature representation capability. In addition, the C3K2 module adopts different structural forms according to whether cross-channel feature recombination and bottleneck operations are introduced, allowing the network to adapt to feature extraction requirements at different stages.

To overcome the difficulties associated with nanoparticle SEM/TEM images, including irregular boundaries, substantial scale variations, and insufficient global contextual modeling, an enhanced instance segmentation framework termed NSYOLO is developed based on YOLOv11. The overall architecture of NSYOLO is illustrated in Figure 2. Instead of directly increasing network depth or stacking additional modules, the proposed framework introduces task-oriented architectural refinements at different stages of the network. These modifications jointly improve fine-grained boundary perception and high-level semantic feature representation, thereby enhancing segmentation performance for complex nanoparticle images.

First, to mitigate boundary information degradation during multi-scale feature fusion, the C3BADS module is introduced into the neck network, replacing the second C3K2 block in the original YOLOv11 architecture. Based on the proposed BADSConv, C3BADS employs a CSP-style dual-branch design, splitting input features into two parallel branches. One branch consists of BADSConv followed by cascaded Bottleneck boundary-aware dynamic snake module (BottleneckBADS) blocks, progressively enhance boundary-sensitive feature representations, while the other branch functions as a lightweight shortcut to preserve original features and maintain computational efficiency. The outputs of the two branches are concatenated along the channel dimension and further fused to produce the final output feature map. By introducing a boundary-guided deformation mechanism, the module adaptively adjusts convolutional sampling positions, guiding sampling points to align with irregular particle contours while effectively suppressing redundant background deformations. Consequently, fine boundary details are better preserved during feature aggregation, significantly enhancing the representation of elongated, curved, and low-contrast nanoparticle structures.

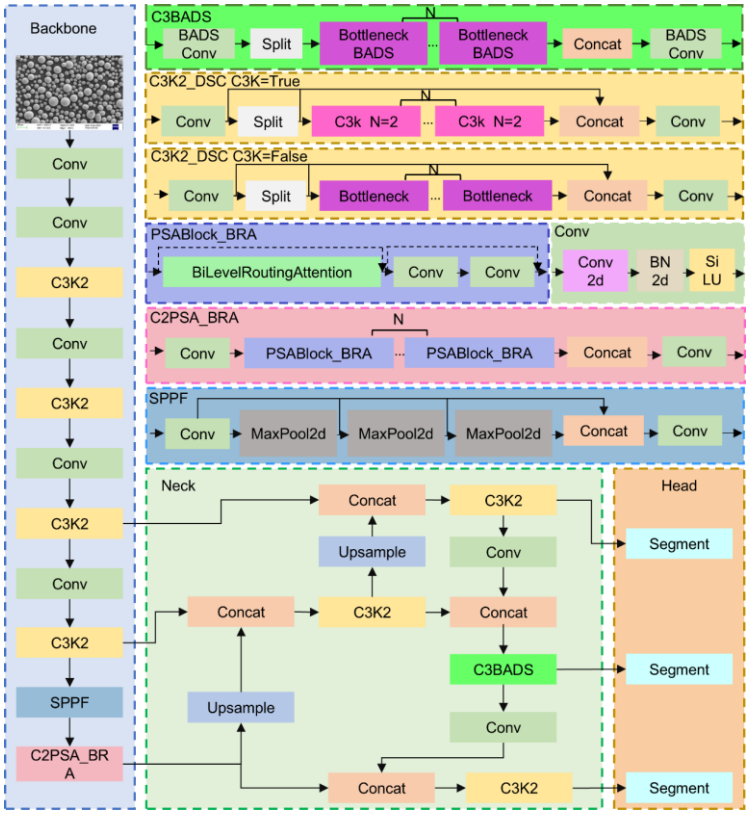

Figure 2. Overall architecture of the proposed NSYOLO model for nanoparticle segmentation in SEM/TEM images.

Second, to enhance global semantic modeling and mitigate background interference in high-level features, the C2PSA_BRA module is integrated into the Backbone, replacing the original C2PSA module that follows the SPPF block. By incorporating BRA, this module facilitates the joint modeling of global contextual dependencies and selective feature aggregation, thereby enabling the network to prioritize discriminative semantic features while suppressing irrelevant background responses.

The two modules operate at the multi-scale feature fusion stage and the high-level semantic modeling stage, forming a complementary optimization mechanism across network hierarchies. This enables progressive enhancement, from local boundary refinement to global semantic representation modeling.

2.2.1. BADSConv

DSConv is a structure-aware convolution operator derived from deformable convolution [32], initially proposed for tasks involving tubular structure segmentation. As depicted in Figure 3a, traditional convolution employs a fixed, regular grid for feature sampling [33], with consistent kernel shapes and sampling locations throughout the spatial domain. This approach imposes inherent limitations when modeling targets with complex morphologies or highly curved boundaries. The sampling pattern of deformable convolution is illustrated in Figure 3b, where learnable offsets are introduced at each sampling point. These offsets enable the convolution kernel to adaptively adjust its sampling positions, thereby enhancing its capacity to capture irregular structures. However, due to the high degree of freedom in offset learning, deformable convolution may generate redundant or meaningless deformations in non-structural regions, resulting in unstable feature representations. In contrast, DSConv imposes structural constraints on the direction and shape of the offsets, directing sampling points to follow the geometric trajectory of the target in a snake-like manner (Figure 3c). This design allows DSConv to better conform to elongated and curved structures while maintaining adaptability and effectively reducing invalid deformations. The computational formulation of DSConv is presented in Equation (1).

$$P_i = (x_i, y_i) = \left(x_c + s_i + \sum_{k=1}^{i} \Delta x_k, y_c + t_i + \sum_{k=1}^{i} \Delta y_k \right) \quad (1)$$

Here, (x_c, y_c) denotes the center position of the convolution kernel; s_i and t_i represent the sampling step sizes of regular convolution along the x and y directions, respectively; and Δx_k and Δy_k denote the accumulated learnable offsets along the corresponding directions. By constraining offset directions, DSConv enables continuous and adaptive sampling along object boundaries and dominant structural orientations.

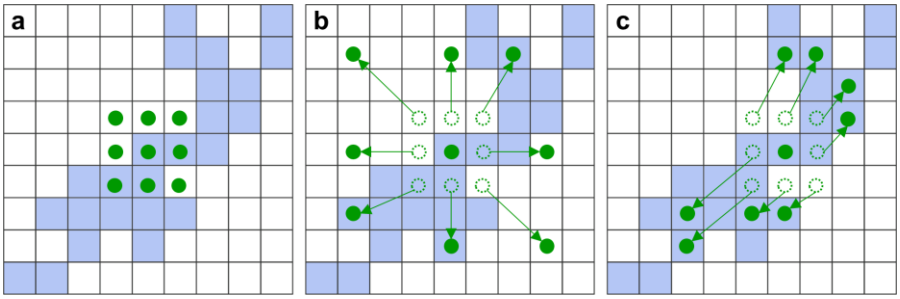

Figure 3. Sampling patterns of different convolution operators: **(a)** conventional convolution; **(b)** Deformable convolution; **(c)** DSConv.

In the process of segmenting nanoparticle SEM/TEM images, particles frequently display highly irregular boundaries, substantial scale variations, and pronounced background interferences. Under these circumstances, DSConv may still generate redundant deformations in areas with weak boundary cues, potentially compromising boundary localization accuracy. To address this limitation, we propose BADSConv, which integrates explicit boundary guidance into the offset learning process, allowing convolutional sampling to adaptively concentrate on nanoparticle boundaries while minimizing unnecessary background deformations. Figure 4 illustrates the overall framework of BADSConv.

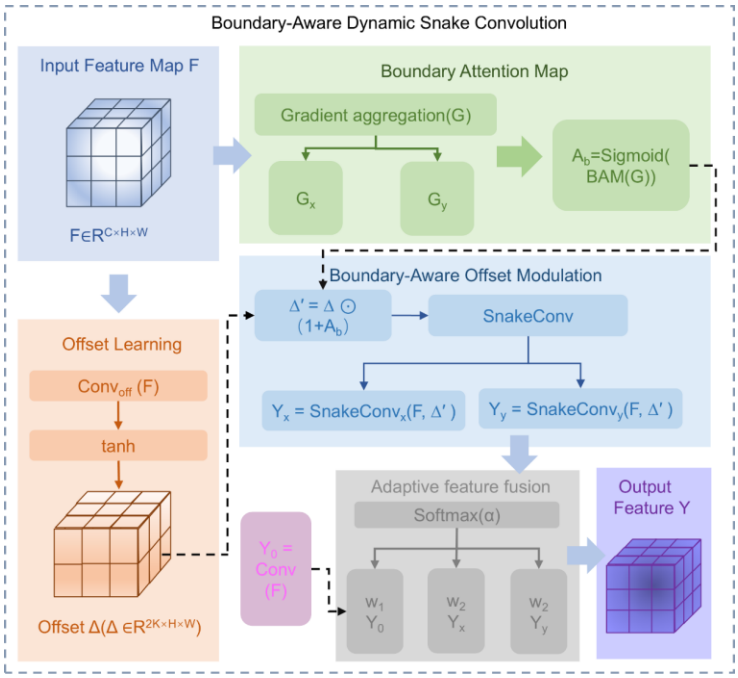

Figure 4. Architecture of the proposed BADSConv.

The pseudocode for the proposed BADSConv is provided in Algorithm 1.

Algorithm 1 Algorithmic workflow of the proposed BADSConv module.

Input: Input feature map F ∈ RC × H × W, kernel size **K**, learnable fusion parameter vector **α** = [α1, α2, α3] ∈ R3
Output: Output feature map Y ∈ RC′ × H′ × W′
// Step 1: Multi-scale gradient extraction
Gx ⇐ |F [:, :, 1 :, :] − F [:, :, : −1, :]|
Gy ⇐ |F [:, :, :, 1 :] − F [:, :, :, : −1]|
G ⇐ Meanc(Gx + Gy)
G ⇐ (AvgPool3 × 3(G) + AvgPool5 × 5(G) + AvgPool7 × 7(G)) / 3
// Step 2: Boundary-aware attention
Ab ⇐ Sigmoid(BAM(G))
// Step 3: Dynamic offset learning
Δ ⇐ tanh(Convoffset(F))
Δ′ ⇐ Δ ⊙ (1 + Ab)
// Step 4: DSConv
Y0 ⇐ Conv(F)
Yx ⇐ DSConvx(F, Δ′)
Yy ⇐ DSConvy(F, Δ′)
// Step 5: Adaptive feature fusion
[w1, w2, w3] ⇐ Softmax(α)
Y ⇐ Conv1 × 1(w1Y0 + w2Yx + w3Yy)
return Y

The computational process of BADSConv is formulated in Equations (2)–(13).

Given an input feature map:

$$F \in R^{C \times H \times W} \tag{2}$$

BADSConv first predicts deformation offsets using a lightweight convolutional module:

$$\Delta = \tanh\left(\mathrm{Conv}_{\mathrm{off}}(F)\right) \tag{3}$$

Here, $\Delta \in R^{2K \times H \times W}$ denotes the learned two-dimensional offsets of the K sampling points in snake convolution, each point corresponding to horizontal and vertical displacements. The hyperbolic tangent function constrains offset magnitude within a bounded range, enhancing training stability. This constraint prevents excessive or noisy offset responses in regions with weak or ambiguous boundary cues, common in nanoparticle SEM/TEM images, thereby reducing redundant deformations and improving deformable convolution robustness.

To explicitly encode the boundary information, spatial gradients are computed along horizontal and vertical directions:

$$G_x = |F_{:,:,1:,:} - F_{:,:,:-1,:}| \tag{4}$$

$$G_y = |F_{:,:,:,1:} - F_{:,:,:,:-1}| \tag{5}$$

The gradient responses are aggregated across the channels to obtain a single-channel boundary descriptor:

$$G = \frac{1}{C}\sum_{c=1}^{C}\left(G_x^c + G_y^c\right) \tag{6}$$

This highlights the regions with prominent structural transitions.

Based on the boundary descriptor, a boundary-aware attention map is generated:

$$A_b = \sigma\left(\mathrm{BAM}(G)\right) \tag{7}$$

where $\sigma()$ denotes the sigmoid function and $A_b \in [0,1]$ indicates boundary likelihood at each spatial location. Predicted offsets are adaptively modulated by the boundary attention:

$$\Delta' = \Delta \odot (1 + A_b) \tag{8}$$

where $\odot$ denotes element-wise multiplication. This amplifies deformations near object boundaries while minimizing changes in the non-boundary regions.

To enhance directional sensitivity and modeling flexibility, BADSConv employs two parallel DSConv branches operating along orthogonal directions:

$$Y_x = \mathrm{DSConv}_x\left(F, \Delta'\right) \tag{9}$$

$$Y_y = \mathrm{DSConv}_y\left(F, \Delta'\right) \tag{10}$$

where DSConv_x and DSConv_y denote the DSConv in the horizontal and vertical directions, respectively.

In parallel, a standard convolution branch is introduced to preserve the local appearance and semantic information:

$$Y_0 = \mathrm{Conv}(F) \tag{11}$$

Finally, outputs from the three branches are adaptively fused using the learnable fusion parameters α:

$$[w_1, w_2, w_3] = \mathrm{Softmax}(\alpha) \tag{12}$$

$$Y = \mathrm{Conv}_{1\times 1}\left(w_1 Y_0 + w_2 Y_x + w_3 Y_y\right) \tag{13}$$

where the learnable weights dynamically balance the contributions of the standard and directional DSConv. To evaluate the performance advantage of BADSConv over DSConv, we conducted ablation experiments. Supplementary Figure S2 presents the training loss curves of baseline YOLOv11 model, as well as YOLOv11 enhanced with DSConv and BADSConv. The introduction of BADSConv leads to faster and more stable model convergence.

Supplementary Table S1 summarizes the quantitative results of the ablation experiments, showing that BADSConv achieves the best performance across most evaluation metrics. However, in terms of parameter count and GFLOPS, BADSConv is slightly inferior to DSConv.

2.2.2. BRA

BRA is an efficient attention mechanism that incorporates dynamic sparsity and query awareness. BRA is designed to enable sparse and efficient feature interaction while preserving global dependency modeling through a two-layer mechanism: "region-level filtering" followed by "label-level aggregation". The overall structure of BRA is illustrated in Figure 5. The mathematical formulations in Equations (14)–(21) are based on the original BRA method proposed in Reference [29].

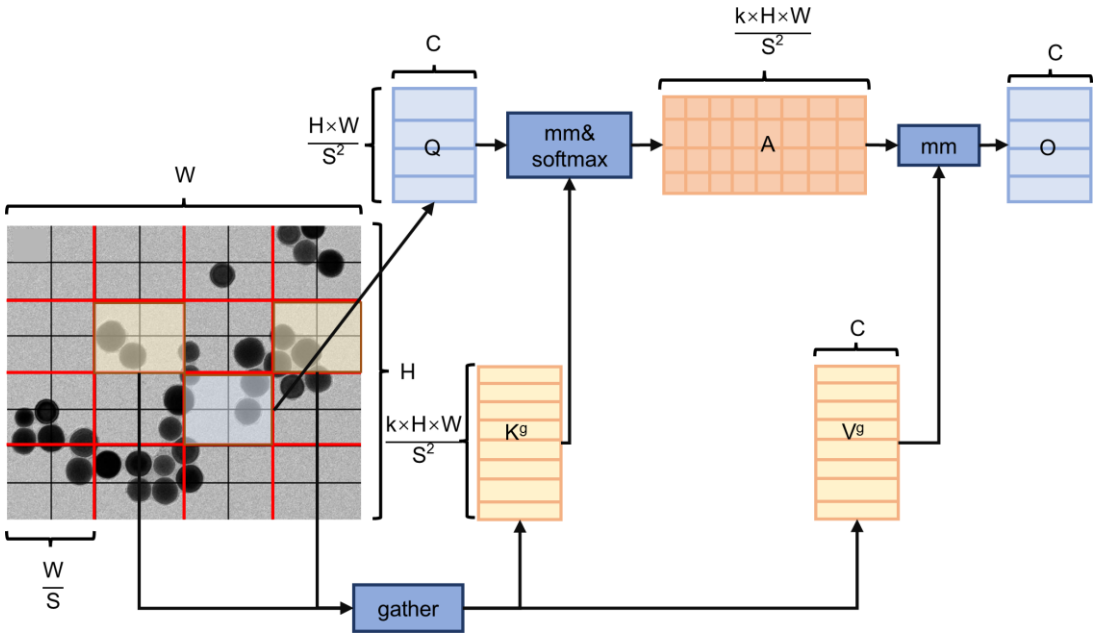

Figure 5. Architecture of BRA. Where, H and W denote the spatial height and width, respectively. C denotes the channel dimension. S is the number of spatial partitions per dimension. k denotes the number of top-ranked regions selected for each query region during region-level routing.

First, the input feature map F is defined as Equation (14).

$$F \in R^{H\times W\times C} \tag{14}$$

The feature map is partitioned into $S \times S$ non-overlapping spatial regions, resulting in S^2 total regions, each containing $(H \times W) / S^2$ feature tokens. After rearranging these subregions, a regional feature tensor is obtained, as expressed in Equation (15).

$$F \in R^{S^2 \times \frac{H \times W}{S^2} \times C} \tag{15}$$

The query (Q), key (K), and value (V) matrices are obtained via linear projections, as defined in Equation (16).

$$Q = FWq,\ K = FWk,\ V = FWv \tag{16}$$

where Wq, Wk, and Wv denote the learnable weights for the linear projections.

A directed graph is constructed to dynamically determine the regions attended by each query region. Specifically, average pooling is applied to the Q and K vectors within each region to obtain region-level representations Q^r and K^r. Then, the adjacency matrix A^r of the regional affinity graph is computed by multiplying Q^r with the transpose of K^r. Each element in this matrix directly represents the semantic correlation strength between the corresponding pair of regions, as defined in Equation (17).

$$A^r = Q^r (K^r)^T \tag{17}$$

Next, to prune the affinity graph, only the top-k most relevant connections per region are retained. This is implemented via row-wise top-k producing the routing index matrix I^r, as defined in Equation (18).

$$I^r = \mathrm{topkIndex}(A^r) \tag{18}$$

Therefore, the i-th row of the matrix, and I^r represents the indices of the top-k regions that exhibit the highest relevance to the i-th region.

Finally, fine-grained attention computation is performed based on the routing index matrix I^r. For each query region, the top-k relevant regions are selected according to Ir, and the corresponding tokens from these regions are aggregated, resulting in $k \times (H \times W / S^2)$ key–value pairs for attention computation, where each region contributes $(H \times W / S^2)$ tokens. For each query token in the i-th region, the model attends to all key–value pairs within the concatenated set of the k routed regions specified by $I^r_{(i,1)}, I^r_{(i,2)}, \ldots, I^r_{(i,k)}$. By aggregating the corresponding key and value tensors, the focused representations K_g and V_g are obtained, after which an attention operation is applied. The LCE function is implemented via depthwise convolution with a 5×5 kernel, as defined in Equations (19)–(21).

$$K^g = \mathrm{gather}(K, I^r) \tag{19}$$

$$V^g = \mathrm{gather}(V, I^r) \tag{20}$$

$$O = \mathrm{Attention}(Q, K^g, V^g) + \mathrm{LCE}(V) \tag{21}$$

BRA is integrated into the C2PSA structure of YOLOv11, replacing the original attention mechanism in the Position-Sensitive Attention Block (PSABlock) to enhance high-level semantic feature modeling.

As the core component of the C2PSA module, the PSABlock is designed to recalibrate feature responses along both spatial and channel dimensions in high-level feature maps. By incorporating an

attention mechanism, the PASBlock highlights semantically relevant regions while reducing the influence of redundant background information, thereby improving the network's ability to represent global object features. In conventional object detection and segmentation tasks, this structure has demonstrated effectiveness in enhancing feature discrimination and semantic awareness. Nevertheless, the original attention mechanism employed in the PASBlock mainly depends on dense local feature interactions, which limits its capability to capture long-range dependencies and semantic correlations across distant regions. This issue becomes more evident in nanoparticle SEM/TEM images, where nanoparticles often exhibit diverse morphologies, significant scale differences, and strong spatial correlations among adjacent particles.

Benefiting from its dynamic routing mechanism, BRA allows the network to focus selectively on regions with strong semantic relevance, thereby improving global contextual perception and enhancing the coherence of high-level semantic representations. By integrating BRA into the PSABlock within the C2PSA module, the proposed framework alleviates the insufficient global modeling capability of the original attention mechanism and strengthens semantic relationship modeling among nanoparticles. As a result, the network can generate more informative and discriminative high-level features, providing a more reliable semantic basis for subsequent instance segmentation with refined boundary prediction.

2.3. Loss function

In instance segmentation of nanoparticles from electron microscopy images, the design of the loss function is essential for achieving accurate object localization and reliable mask prediction. Due to the complex morphologies and significant scale variations of nanoparticles, the YOLOv11 framework was optimized in an end-to-end manner using a multitask joint loss strategy. This optimization scheme simultaneously improves object localization, category classification, and instance mask generation. The overall loss function mainly consists of two parts: the detection loss and the segmentation loss, as defined in Equation (22).

$$L = L_{det} + L_{seg} \tag{22}$$

The detection loss (L_{det}) comprises Classification Loss (L_{cls}), Box Regression Loss (L_{box}), and Distribution Focal Loss (L_{dfl}), which collectively optimize the object detection performance. The Segmentation Loss (L_{seg}) constrains pixel-level region predictions, thereby enabling instance segmentation. The overall loss formulation is expressed as:

$$L = L_{cls} + L_{box} + L_{dfl} + L_{seg} \tag{23}$$

(1) Lcls

YOLOv11 employs the Binary Cross-Entropy (BCE) loss function [34] in the classification branch to quantify the discrepancy between predicted class probabilities and ground-truth labels, thereby enhancing the model's class discrimination capability. The corresponding calculation is expressed as:

$$L_{cls} = -\sum_{i=1}^{C} [\, y_i \log(p_i) + (1-y_i) \log(1-p_i)] \tag{24}$$

Here, C denotes the total number of categories, y_i is the ground-truth label for the i-th category, and p_i is the predicted probability for that category.

(2) Lbox

For target localization, YOLOv11 employs the Complete Intersection over Union (CIOU) loss function [35] to enhance both geometric overlap and shape consistency between predicted bounding boxes and ground truth boxes. Unlike traditional Smooth L1 [36] or IOU loss, CIOU loss considers the center-point distance, overlap area, and aspect ratio difference between boxes, thereby more effectively optimizing target localization accuracy. The calculation is expressed as:

$$L_{box} = 1\text{-}IOU + \frac{\rho^2\left(b,b^{gt}\right)}{c^2} + \alpha v \tag{25}$$

Here, IOU denotes the intersection-union ratio between the predicted and ground-truth boxes; $\rho(b,b_{gt})$ represents the Euclidean distance between their center points; c is the diagonal length of the smallest enclosing rectangle covering both boxes; v quantifies aspect ratio consistency; and α is a weighting parameter controlling the relative contribution of each term.

(3) Ldfl

In L_{box}, YOLOv11 employs the Distribution Focal Loss to enhance the precision of predicted box coordinates. Unlike directly predicting continuous coordinate values, Distribution Focal Loss treats each coordinate as a probability distribution over discrete interval [37], achieving more accurate regression by learning the distribution's centroid. The calculation is expressed as:

$$DFL(S_i,S_{i+1}) = -\left(\left(y_{i+1}-y\right)\log(S_i) + \left(y-y_i\right)\log(S_{i+1})\right) \tag{26}$$

Here, S_i and S_{i+1} denote the predicted probabilities corresponding to discrete coordinate intervals y_i and y_{i+1}, respectively and y is the true continuous coordinate value.

(4) Lseg

In mask prediction, the segmentation branch of YOLOv11 also employs BCE loss to quantify the discrepancy between predicted and ground-truth masks, thereby optimizing pixel-level region classification accuracy. This loss is consistent with the BCE formulation used in the classification branch and effectively enhances segmentation performance for nanoparticle boundaries and morphological details.

2.4. Model training

Model training and validation were conducted on a computing platform equipped with an Intel® Xeon® Gold 6430 CPU, an NVIDIA GeForce RTX 4090 GPU, and 30 GB of memory. The system operated on Ubuntu 22.04, with the development environment comprising Python 3.10 and the PyTorch framework. Detailed specifications of the experimental environment are summarized in Supplementary Table S2. The hyperparameters used for model training are listed in Supplementary Table S3.

2.5. Model evaluation indicators

To systematically evaluate the performance of the proposed model for instance segmentation of electron microscopy images, multiple evaluation metrics were computed on the test dataset. The evaluation metrics comprised Precision, Recall, F1 Score, Intersection over Union (IoU), and mean Average Precision (mAP@0.5 and mAP@0.5–0.95). Specifically, mAP@0.5 uses an IoU threshold of 0.5 to determine prediction correctness, primarily reflecting the model's fundamental detection capability. In contrast, mAP@0.5–0.95 computes the average precision over multiple IoU thresholds from 0.5 to 0.95

in increments of 0.05, providing a more comprehensive and stringent assessment of the model's accuracy in both bounding box and mask predictions. The computation formulas for these metrics are presented in Equations (27)–(32), where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives respectively. Here, Prediction refers to the predicted region, and GroundTruth denotes the annotated reference region. Equations (27)–(30) follow the standard evaluation metrics commonly used in object detection and segmentation tasks [38]. Equations (31)–(32) are based on the evaluation protocol proposed in the COCO benchmark [39].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{27}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{28}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{29}$$

$$\text{IoU} = \frac{|\text{Prediction} \cap \text{GroundTruth}|}{|\text{Prediction} \cup \text{GroundTruth}|} \tag{30}$$

$$\text{AP} = \int_0^1 p(r)dr \tag{31}$$

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{A P}_i \tag{32}$$

3. Results and discussion

3.1. Loss analysis

Loss functions are critical metrics for assessing the discrepancies between model predictions and ground-truth labels. To further examine the differences in training convergence between NSYOLO and the baseline model, Figure 6 illustrates the variations of the individual loss functions throughout the training process. Specifically, NSYOLO demonstrates faster convergence and a smoother trajectory, indicating enhanced accuracy and stability in bounding box regression (Figure 6a). The rapid decline of the loss function underscores the effectiveness of the C3BADS module in modeling target boundary features (Figure 6b). Additionally, the superior convergence behavior of the loss function reflects the efficacy of the C2PSA_BRA module in enhancing feature discriminability (Figure 6c). The continuous decline of the loss function indicates that the model effectively captures the distribution characteristics of small-scale particles and targets under complex background conditions (Figure 6d).

Overall, NSYOLO attains lower loss values across all loss components and exhibits faster and smoother convergence compared with baseline model, thereby fully demonstrating the superior optimization capability of the proposed network architecture in feature extraction and bounding box regression tasks. This stable and efficient convergence behavior is particularly critical for nanoparticle segmentation, given that nanoparticles frequently display significant scale variations, irregular boundary morphologies, and complex background interference. Under these conditions, the rapid and stable convergence of the L_{box} mitigates localization jitter for small or densely packed particles, whereas the

efficient reduction of segmentation and classification losses further enhances boundary delineation accuracy and ensures more reliable instance discrimination throughout training.

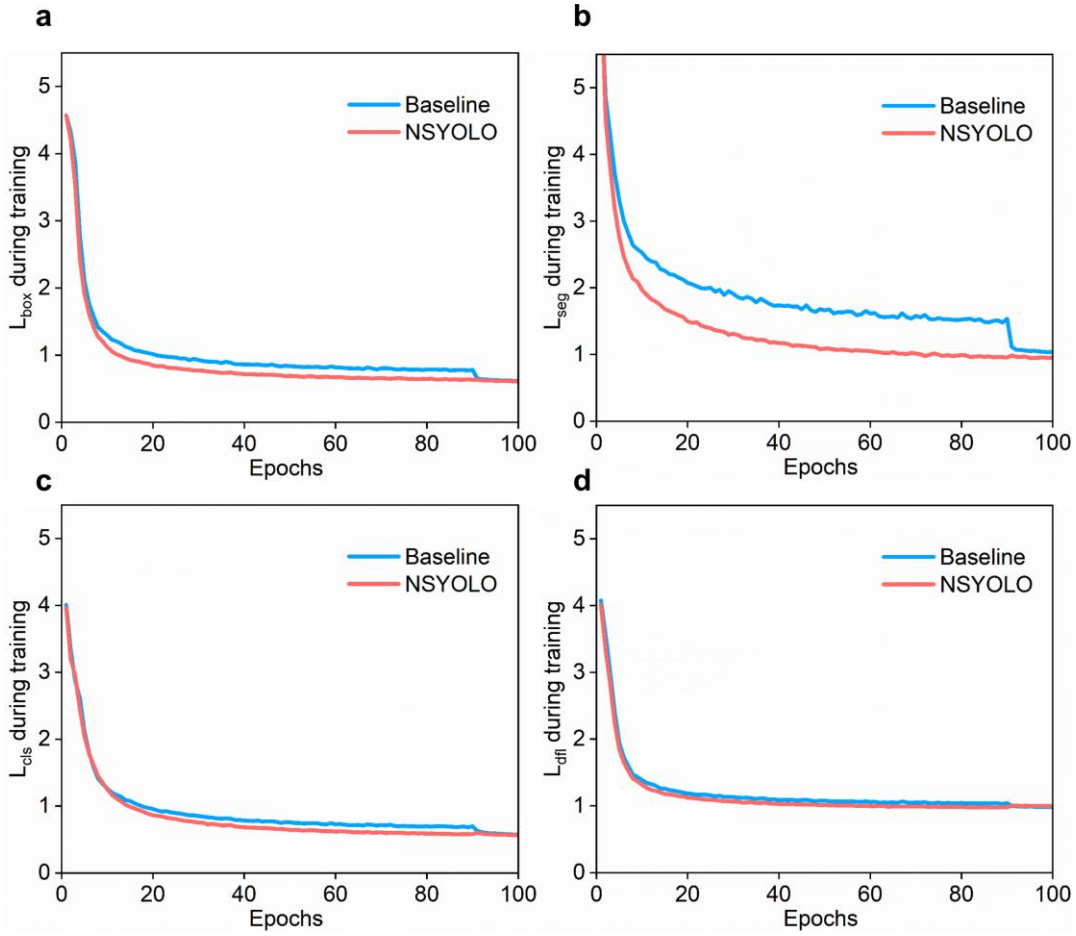

Figure 6. Training loss curves of NSYOLO and the baseline. **(a)** L_{box}, **(b)** L_{seg}, **(c)** L_{cls}, and **(d)** L_{dfl}.

3.2. Visualization and comparison experiment of segmentation results

We conducted a comparative evaluation of Unet, Mask R-CNN, DeepLabv3+, YOLOv11, and the proposed NSYOLO model across multiple performance metrics, as depicted in Figure 7. Overall, NSYOLO demonstrates superior performance across all evaluated metrics.

As illustrated in Figure 7a, NSYOLO achieves an approximately 6-percentage-point improvement in precision compared to the baseline YOLOv11 model, indicating a significant reduction in false-positive detections. This suggests that background regions or noise are less likely to be misclassified as nanoparticles, which is crucial for reliable particle counting and statistical analysis. As shown in Figure 7b, NSYOLO also attains the highest Recall, reflecting its robust capability to detect a larger proportion of true nanoparticles. This improvement is particularly critical for densely packed or small-scale nanoparticles, where missed detections can substantially bias size distribution and morphology statistics.

Furthermore, as shown in Figure 7c,d, NSYOLO exhibits notable improvements in the F1 Score and IoU relative to traditional segmentation models such as Unet and Mask R-CNN. These improvements indicate an enhanced balance between Precision and Recall, as well as more accurate boundary alignment and instance integrity, thereby fully demonstrating the superior segmentation accuracy and robustness of NSYOLO in complex nanoparticle imaging scenarios.

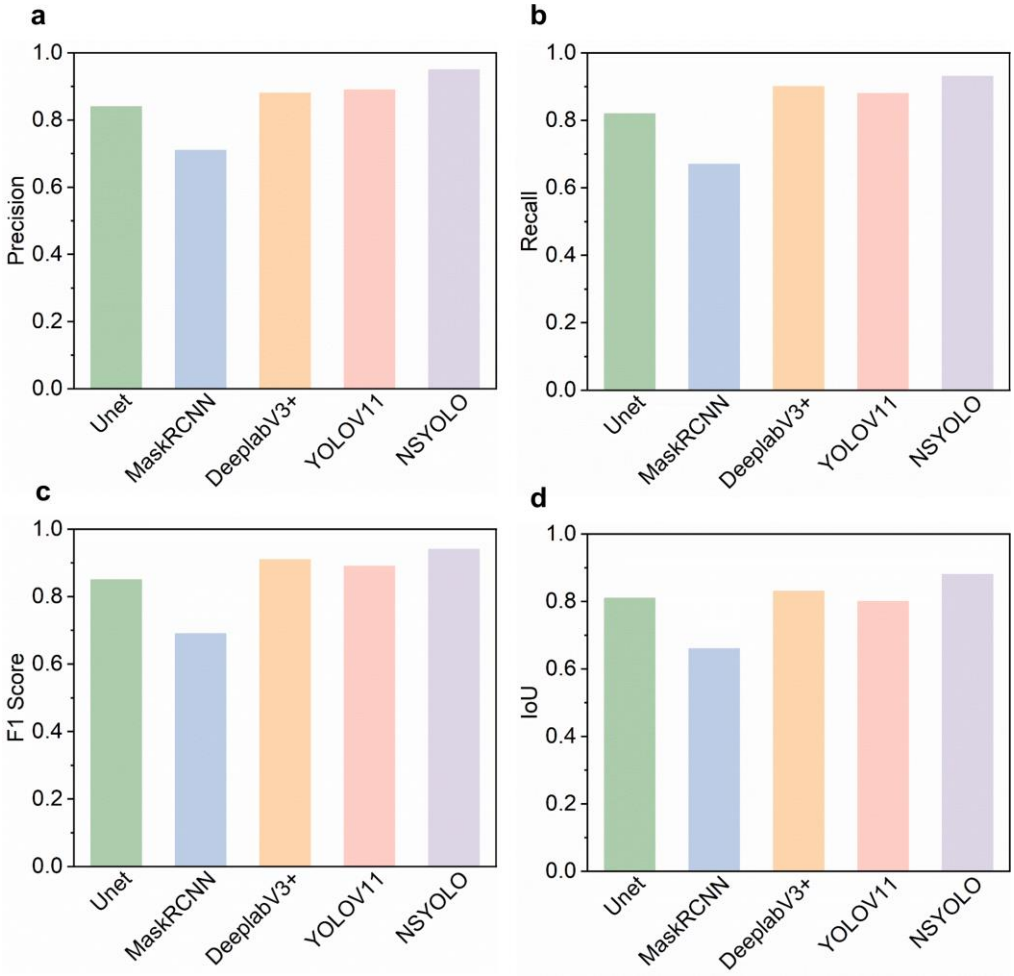

Figure 7. Quantitative performance comparison of different segmentation models evaluated on the test dataset: **(a)** Precision, **(b)** Recall, **(c)** F1 Score, and **(d)** IoU.

Figure 8 illustrates the qualitative segmentation outcomes of various models applied to SEM images of carbon nanospheres. Specifically, Figures 8a,b display the raw input image and the corresponding ground-truth annotation, respectively. The Unet and DeepLabv3+ models exhibit significant particle adhesion during segmentation (Figures 8c,d), complicating the differentiation of adjacent or contacting nanoparticles. The YOLOv11 model demonstrates limited efficacy in managing overlapping particles, failing to accurately separate them and resulting in missed detections in certain regions (Figure 8e). In contrast, NSYOLO demonstrates superior performance in boundary delineation and detail preservation, facilitating more accurate and reliable segmentation of nanoparticles and fully highlighting its advantages in complex scenarios (Figure 8f). This enhanced boundary delineation is attributed to the boundary-aware design of the BADSConv module, which explicitly guides feature sampling along object contours while suppressing redundant or noisy deformations in regions with weak boundary cues. By incorporating explicit boundary constraints into the deformable convolution process, BADSConv enables the network to effectively focus on fine-grained edge structures, thereby preserving subtle boundary details and improving instance separation in densely packed or overlapping nanoparticle scenarios. Consequently, NSYOLO achieves more accurate contour localization and more consistent segmentation performance under complex imaging conditions.

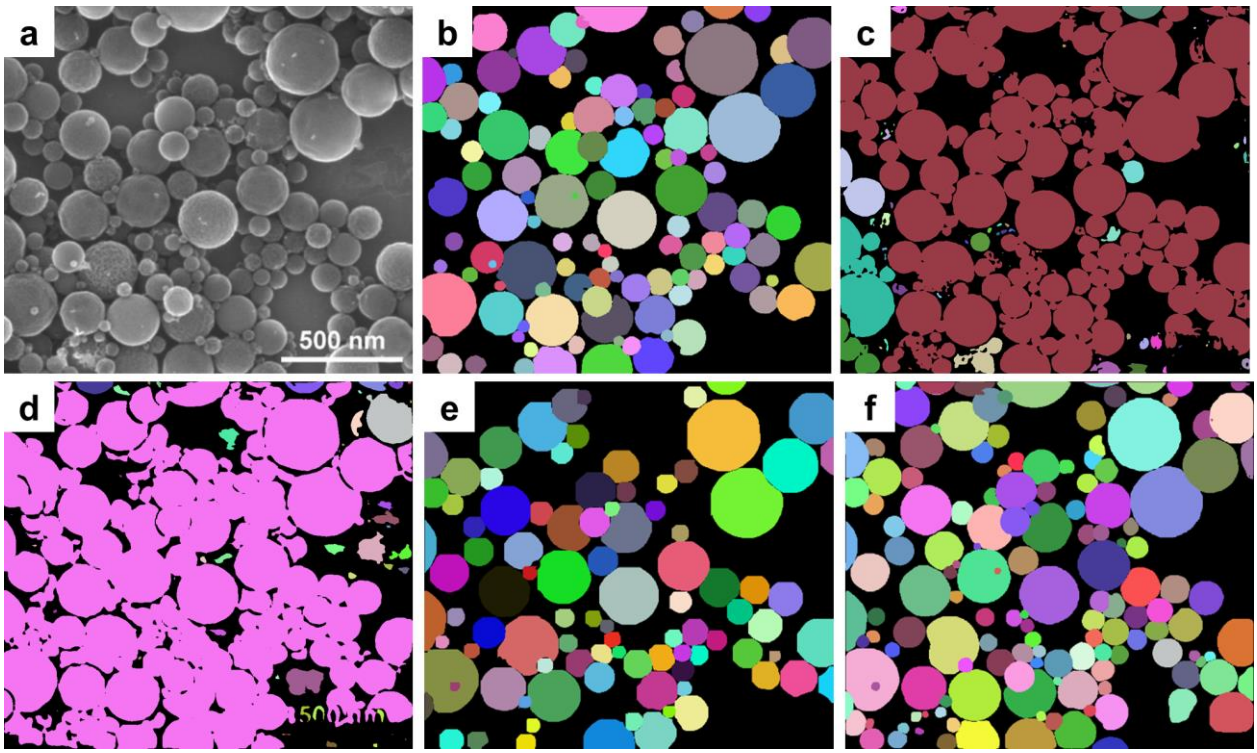

Figure 8. Visual comparison of segmentation results produced by different models: **(a)** SEM image of carbon nanospheres synthesized from $Mn(NO_3)_2$ and sugar precursors after HCl treatment [40], Reprinted with permission. Copyright 2013 American Chemical Society; **(b)** Ground-truth annotation; **(c–f)** Segmentation results generated by Unet, DeepLabV3+, YOLOv11, and NSYOLO, respectively.

Based on the segmentation results of the nanocubes and nanorods shown in Supplementary Figures S3 and S4, NSYOLO demonstrated clearer boundary delineation and superior capability in separating densely packed nanostructures compared with the other models.

We conducted a comparative analysis of the proposed method against existing open-source segmentation techniques for nanoparticle microscopy images, focusing on three representative nanostructures (Figure 9a,e,i). The comparison encompassed a traditional threshold-based segmentation algorithm implemented in ImageJ and the ImageDataExtractor approach. Visual examination of the segmentation outcomes reveals that threshold-based methods frequently exhibit excessive boundary thinning during segmentation (Figure 9b,f,j). This phenomenon arises primarily because these methods classify pixels solely based on image intensity, resulting in limited robustness when applied to microscopy images with low contrast or complex backgrounds. As illustrated in Figure 9c,g,k, although ImageDataExtractor demonstrates relatively satisfactory overall segmentation performance, issues of particle adhesion persist in certain instances, and its segmentation results are often accompanied by boundary expansion, leading to discrepancies between the segmented contours and the true particle shapes in the original images. In contrast, NSYOLO demonstrates clear advantages in boundary delineation accuracy, separation of overlapping particles, and overall segmentation consistency (Figure 9d,h,l). It is noteworthy that, in rare cases, extremely small or low-contrast particles may still be overlooked; however, such occurrences are infrequent and exert a limited impact on the overall segmentation performance.

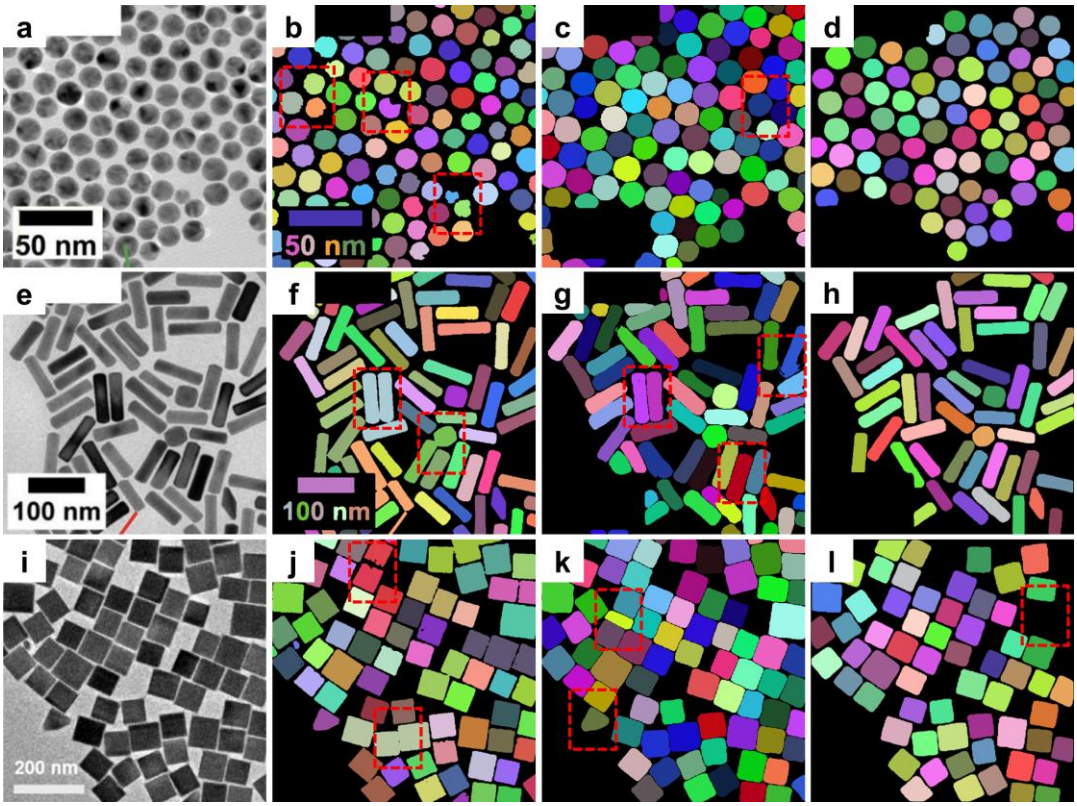

Figure 9. Comparison of nanoparticle segmentation results obtained using different methods for various nanoparticles. **(a)** Original TEM image of gold nanosheets (GNSs) [41]. Reprinted with permission. Copyright 2018 American Chemical Society; Segmentation results of the GNSs shown in **(a)** obtained using the **(b)** Otsu thresholding implemented in ImageJ; **(c)** ImageDataExtractor method, and **(d)** NSYOLO method; **(e)** Original TEM image of PEGylated gold nanorods [41]. Reprinted with permission. Copyright 2018 American Chemical Society; segmentation results of the PEGylated GNRs shown in **(e)** obtained using the **(f)** Otsu thresholding; **(g)** ImageDataExtractor method, and **(h)** NSYOLO method; **(i)** Original TEM image of the Ag nanocubes [42]. Reprinted with permission. Copyright 2023 American Chemical Society; Segmentation results of the Ag nanocubes shown in **(i)** obtained using **(j)** Otsu thresholding; **(k)** ImageDataExtractor method, and **(l)** NSYOLO method. The red dashed boxes indicate the particles with abnormal segmentation.

3.3. Ablation studies

Ablation studies were performed to assess the contribution of individual modules to the overall model performance. By selectively removing or replacing specific components while keeping all other conditions constant, performance variations were analyzed to evaluate the impact of each module on the model's overall performance.

The effects of the BADSConv and BRA modules on the model performance were systematically examined. Four experimental configurations were designed as follows: Scheme 1: Baseline YOLOv11 model; Scheme 2: The baseline model with the original C3K2 module at the P4 layer of the Neck network replaced by the C3BADS module; Scheme 3: Baseline model with the original C2PSA module replaced by the C2PSA_BRA module; Scheme 4: NSYOLO, the proposed model integrating both the

C3BADS and C2PSA_BRA modules. The four configurations were evaluated on the test set, and the corresponding experimental results are summarized in Table 1.

Table 1. Comparison of ablation experiment results evaluating the impact of the proposed C3BADS and C2PSA_BRA modules on the performance of YOLOv11-based nanoparticle segmentation model.

Scheme	C3BADS	C2PSA_BRA	Precision	Recall	F1 Score	mAP@0.5
Scheme 1	×	×	0.885	0.882	0.881	0.906
Scheme 2	√	×	0.903	0.890	0.897	0.920
Scheme 3	×	√	0.894	0.891	0.893	0.923
Scheme 4	√	√	0.946	0.927	0.935	0.957

The experimental results demonstrate that incorporating different modules leads to varying degrees of performance improvement. Compared with Scheme 1, Scheme 2 achieved higher precision and mAP@0.5 values of 0.903 and 0.920, respectively, following the incorporation of the C3BADS module. These results indicate that the BADSConv module effectively enhances the model's discriminative capability during feature extraction.

Following the introducing the C2PSA_BRA module, Scheme 3 showed notable improvements in Recall and F1 Score, increasing to 0.891 and 0.893, respectively, with mAP@0.5 reaching 0.923. These findings suggest that the BRA module allows the model to focus on more informative feature regions, thereby enhancing overall detection performance.

Scheme 4, which integrates both the C3BADS and C2PSA_BRA modules, demonstrates superior performance across all evaluation metrics, achieving Precision, Recall, F1 Score, and mAP@0.5 values of 0.946, 0.927, 0.935, and 0.957, respectively. This indicates that the two modules exhibit significant complementarity, collectively enhancing the model's detection capabilities. These performance improvements further substantiate the strong complementarity between the modules. Specifically, BADSConv enhances local boundary representation and geometric continuity at the instance level by explicitly modeling fine-grained contour features, while the BRA module captures long-range dependencies and global contextual relationships among nanoparticles. The integration of these modules enables the model to jointly optimize fine-grained boundary accuracy and global semantic consistency, which is particularly beneficial in scenarios involving densely packed or overlapping nanoparticles. Consequently, NSYOLO achieves superior overall segmentation performance compared to configurations employing either module individually.

Furthermore, confusion matrices were constructed for the four experimental schemes to facilitate a more intuitive analysis of classification performance. These matrices illustrate the correspondence between ground-truth labels and predicted categories, enabling a detailed assessment of classification accuracy and inter-class confusion. As depicted in Figure 10a–d, all four schemes demonstrated high prediction accuracy for the nanocube category. However, the baseline model exhibits significantly lower prediction probabilities for the remaining two categories compared to Scheme 4, suggesting that the concurrent integration of the C3BADS and C2PSA_BRA modules substantially enhances the predictive performance of the baseline model.

As shown in Figure 11, a systematic heatmap-based comparative analysis was conducted on the baseline, YOLOv11 augmented with BADSConv, YOLOv11 incorporating BRA, and YOLOv11

integrating both BADSConv and BRA. Three types of heatmaps—confidence heatmaps, mask probability maps, and boundary heatmaps were employed to intuitively evaluate the models' performance in terms of target recognition reliability, boundary representation accuracy, and geometric consistency. The results indicate that introducing BADSConv alone produces substantial improvements in detection confidence, continuity of mask probability distributions, and boundary quality, particularly in regions with complex contours where polygonal artifacts are effectively suppressed and smoother, more coherent segmentation boundaries are achieved. In comparison, incorporating BRA leads to notable gains in detection confidence and mask prediction accuracy, although its contribution to fine-grained boundary modeling is less pronounced than that of BADSConv. Moreover, the synergistic integration of BADSConv and BRA markedly enhances global semantic modeling capability, enabling the model to attain the best overall performance with respect to both detection confidence and boundary smoothness. Due to space limitations, the TEM image of nanocubes [43] are not shown.

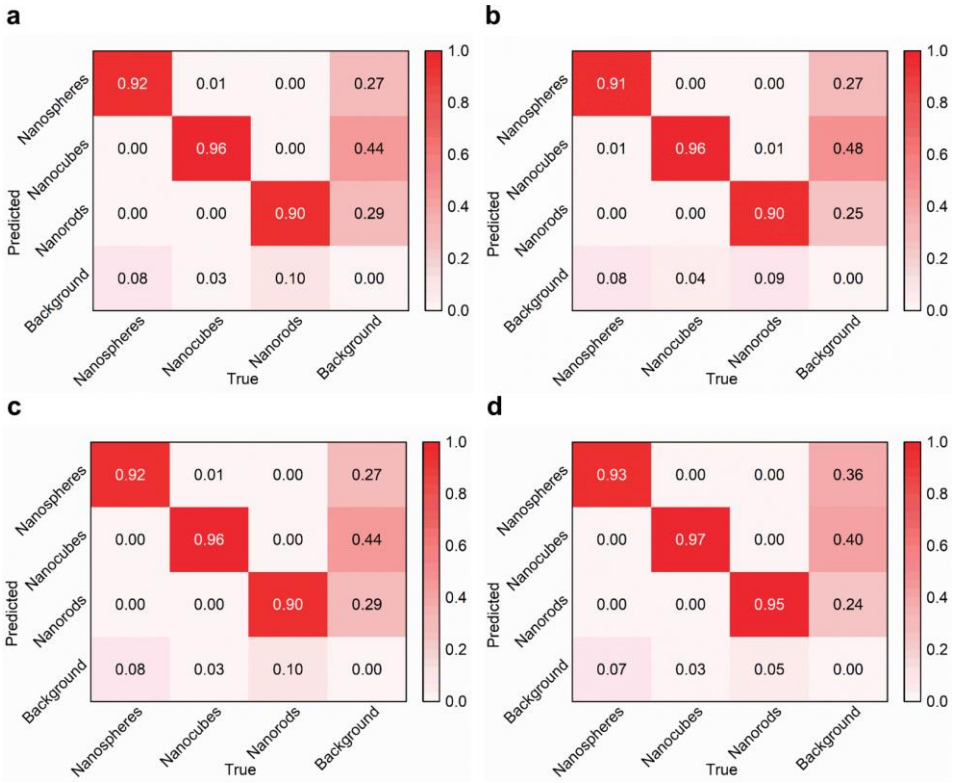

Figure 10. Confusion matrices for the four ablation experiment schemes. **(a)** Baseline YOLOv11 model; **(b)** model with the C3BADS module; **(c)** model with the C2PSA_BRA module; **(d)** NSYOLO model integrating both the C3BADS and C2PSA_BRA modules. The matrices illustrate the correspondence between the ground-truth labels and predicted categories, allowing for an intuitive evaluation of the classification accuracy and inter-class confusion across different model configurations.

To further verify the effectiveness of the proposed method in instance boundary segmentation, a visual comparison of boundary prediction results obtained by different models is conducted. The original input image, which is a TEM image of nanorods with a high degree of spatial overlap between instances, poses significant challenges for accurate instance boundary delineation (Figure 12a). The boundary segmentation results produced by the YOLOv11 baseline model and the models enhanced with DSConv, BADSConv, BRA, and the combination of BADSConv and BRA, respectively (Figure 12b–f). In

regions with severe nanorod overlap, the baseline model and the model incorporating only DSConv still exhibit boundary discontinuities and blurred contours, making it difficult to reliably separate adjacent or overlapping instances (Figure 12b,c). In contrast, after introducing BADSConv, the boundary responses become more concentrated around object edges, and the contours of overlapping nanorods are more clearly delineated (Figure 12d). The predicted boundaries show improved alignment with the true nanorod shapes, indicating that BADSConv effectively enhances the modeling of local boundary features. Further comparison between Figure 12b and e shows that the introduction of the BRA module expands the overall response across instance regions. Boundary responses in several overlapping nanorod areas-weakly activated in the baseline model-being significantly strengthened, suggesting that BRA plays a positive role in integrating contextual information and improving the stability of responses in overlapping instances. When BADSConv and BRA are jointly incorporated (Figure 12f), the model achieves a more balanced performance in highly overlapping nanorod scenarios, effectively improving boundary continuity and instance completeness while maintaining accurate boundary localization. Compared with configurations using a single module, this combined design yields superior overall performance in complex overlapping instance segmentation tasks. These visualization results further demonstrate the complementary effectiveness of BADSConv and BRA in high-overlap nanorod instance boundary segmentation. Additional visualizations for spherical and cubic nanoparticles are provided in Supplementary Figures S5 and S6, further supporting the generality of the proposed approach.

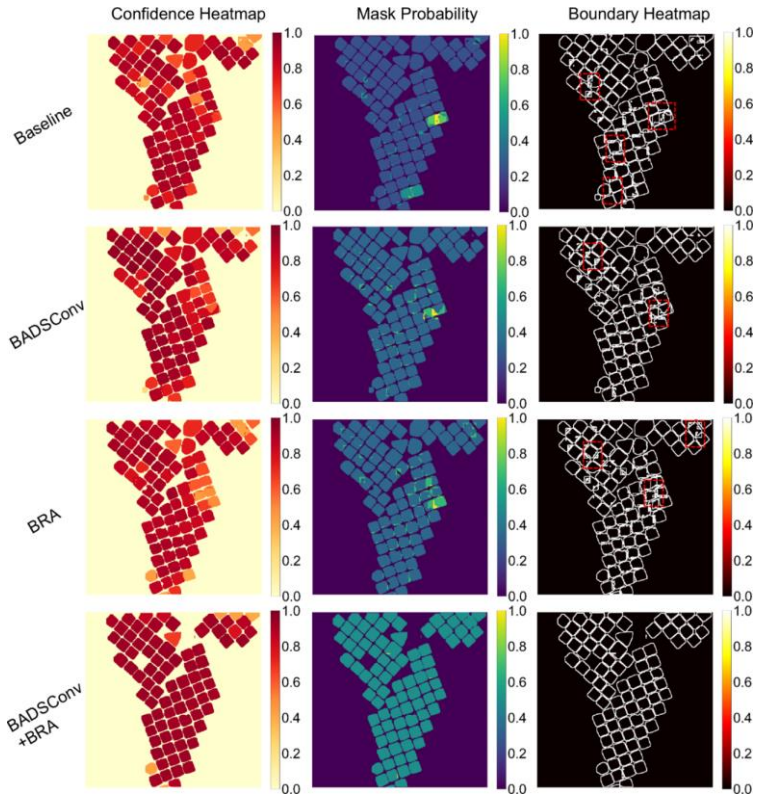

Figure 11. Comparative heatmap analysis of segmentation performance across four methods. The rows correspond to the baseline YOLOv11, YOLOv11 with BADSConv, YOLOv11 with BRA, and YOLOv11 integrating both BADSConv and BRA(NSYOLO). The columns show the confidence heatmap, the mask probability map, and the boundary heatmap, respectively.

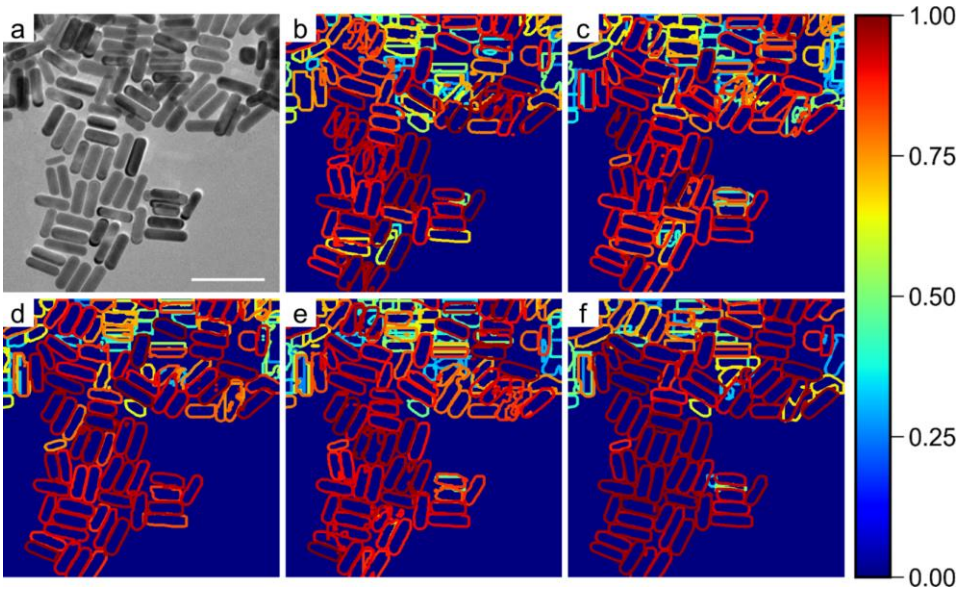

Figure 12 Visualization of instance boundary segmentation results produced by different models. **(a)** TEM image of Au NRs [44]; **(b)** YOLOv11; **(c)** YOLOv11 with DSConv; **(d)** YOLOv11 with BADSConv; **(e)** YOLOv11 with BRA; **(f)** YOLOv11 integrating both BADSConv and BRA (NSYOLO). The color intensity indicates the boundary response strength of the predicted instances.

3.4. Applicability and limitations of NSYOLO model

To further assess the generalization capability and practical applicability of the proposed model, additional analyses were performed. These analyses considered non-ideal imaging conditions, various imaging modalities, nanoparticle morphologies beyond the training dataset, and comparisons with alternative workflows based on pre-trained vision foundation models.

Initially, the model's robustness under suboptimal imaging conditions was examined. As illustrated in Supplementary Figure S7a,b, the segmentation performance significantly deteriorates for nanoparticles with blurred or indistinct boundaries, particularly when adjacent particles are closely connected or overlapping, due to the absence of clear boundary information. Conversely, as depicted in Supplementary Figure S7c,d, the proposed model achieves satisfactory segmentation performance for nanoparticles with relatively clear boundaries but irregular shapes. This behavior suggests that the model predominantly relies on boundary clarity and local intensity contrast for accurate segmentation. Given that images with defocus and charging artifacts are infrequently reported in the literature, where higher-quality images are typically presented, such cases were not systematically included in this study. Nevertheless, based on their known imaging characteristics, it is anticipated that the model would perform less effectively on such images. Specifically, defocus tends to smooth particle edges and suppress high-frequency details, while charging artifacts introduce local intensity distortions (e.g., bright spots or streaking), both of which can impede reliable boundary extraction. These observations indicate that the model performs robustly as long as particle boundaries remain distinguishable, but its performance diminishes when boundary information is significantly compromised.

Second, the model's applicability across various imaging modalities was assessed. The NSYOLO model was initially trained on bright-field TEM (BF-TEM), high-angle annular dark-field STEM (HAADF-STEM), and secondary electron SEM images, where nanoparticles typically display strong

intensity contrast and well-defined edges. To further evaluate the model's performance under different imaging modalities, additional experiments were conducted, with visualization results presented in Supplementary Figure S8. As illustrated in Supplementary Figure S8a,d, the model maintains effective segmentation performance on backscattered electron (BSE) SEM images, indicating good transferability to similar contrast conditions. For BF-TEM images (Supplementary Figure S8b,e), accurate segmentation is also achieved due to the presence of clear particle boundaries. However, for dark-field TEM images (Supplementary Figure S8c,f), the reduced contrast between particles and the background results in missed detections, although particles with relatively higher contrast can still be partially identified. Furthermore, additional tests on high-resolution TEM (HRTEM) images reveal that the model fails to reliably segment nanoparticles. This is primarily due to the fundamentally different contrast mechanism in HRTEM, where lattice fringes and phase contrast dominate image formation, resulting in blurred particle boundaries and ambiguous contours. These findings demonstrate that model performance is primarily governed by boundary clarity and contrast quality rather than the imaging modality itself. Imaging conditions that produce strong contrast and well-defined edges are favorable, whereas complex contrast mechanisms or reduced edge definition lead to performance degradation.

Third, the generalization capability of the proposed model was assessed using a small out-of-distribution (OOD) dataset comprising triangular and pentagonal nanoparticles, which were not included in the training data. As demonstrated in Supplementary Table S4, the detection rate declines as the confidence threshold increases, indicating greater uncertainty with structures absent from the training dataset. Nonetheless, the model successfully detects a significant proportion of nanoparticles. The visual results are depicted in Supplementary Figure S9. The original test images are presented in Supplementary Figure S9a,e,i,m. As illustrated in Supplementary Figure S9b,f,j,n, at a lower confidence threshold of 0.25, most triangular and pentagonal nanoparticles are successfully detected. When the threshold is raised to 0.5 (Supplementary Figure S9c,g,k,o), the majority of OOD morphologies remain detectable, suggesting a degree of generalization capability. However, further increases in the threshold (Supplementary Figure S9d,h,l,p) result in a marked decrease in the detection rate of triangular nanoparticles, while pentagonal nanoparticles maintain relatively higher detection performance. This phenomenon may be attributed to the geometric similarity between pentagonal nanoparticles and the predominantly spherical particles in the training dataset, leading to higher-confidence predictions. In contrast, triangular nanoparticles exhibit more distinctive structural features, resulting in lower confidence scores and greater sensitivity to the threshold choice. These findings indicate that the model's generalization ability is influenced not only by the presence of non-trained data but also by the structural similarity between the target morphology and the training distribution. This analysis mitigates concerns regarding potential overfitting and provides practical insights into the model's applicability to other nanoparticle systems, including guidance on selecting appropriate confidence thresholds in real-world applications. Furthermore, it is noted that the developed analysis system allows users to adjust the confidence threshold for segmentation, enabling researchers to balance detection sensitivity and prediction reliability according to specific application requirements. More details can be found in Figure 13a under the "Segmentation Details" panel.

Finally, it is important to note that the proposed approach is based on supervised learning, which inherently requires annotated training data and may limit generalization when applied to significantly different imaging conditions or particle morphologies. In contrast, recent approaches based on pre-trained

vision foundation models, such as the Segment Anything Model (SAM) [45], offer strong zero-shot generalization and flexibility. For instance, in the workflow proposed by Genc *et al.*, an object detection model (YOLOv8) is first used to localize nanoparticles and generate bounding box prompts, which are then used by SAM to perform segmentation [46]. However, such methods typically rely on user-provided prompts or auxiliary detection models. In comparison, the proposed method provides a fully automated and consistent solution, which is particularly advantageous for high-throughput nanoparticle analysis involving large datasets. Therefore, these approaches should be considered complementary, with the choice depending on the specific application scenario.

3.5. Development of a geometric–morphological analysis and visualization system

To overcome the practical usability constraints associated with current deep learning–based nanoparticle analysis tools, particularly those related to complex environmental configuration and the necessity for programming expertise, a web-based system for geometric and morphological analysis and visualization has been developed utilizing Python Flask. This system facilitates fully automated processing, quantitative analysis, and visualization of electron microscopy images through an intuitive graphical interface, eliminating the need for coding or manual parameter adjustments.

As illustrated in Figure 13a, users are able to execute the complete analysis workflow—encompassing image upload, nanoparticle instance segmentation, scale calibration, morphological parameter extraction, and result visualization—through straightforward interactive operations. The system incorporates key metric visualization, automated statistical analysis, and the capability for one-click downloading of editable analysis reports, thereby substantially lowering the technical barrier for researchers in materials science. Figures 13b,c display representative quantitative analysis results for spherical nanoparticles, while comprehensive analysis results for spherical, rod-shaped, and cubic nanoparticles are provided in Supplementary Figures S10–S12, respectively.

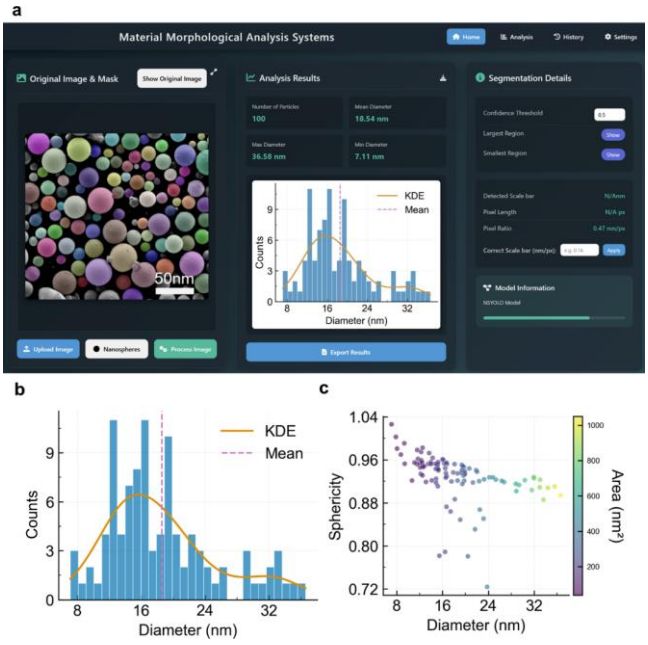

Figure 13. Nanoparticle microscopy image geometric analysis system. **(a)** Main interface of the nanoparticle analysis system; **(b)** Diameter distribution histogram of spherical nanoparticles; **(c)** Relationship between sphericity and area under different diameters for spherical nanoparticles.

4. Conclusion

This study introduces NSYOLO, a deep learning framework designed for the automated instance segmentation and geometric and morphological analysis of nanoparticles in SEM/TEM images. It also constructs a high-quality SEM/TEM image dataset encompassing three representative nanoparticle morphologies: nanospheres, nanorods, and nanocubes. By integrating the BADSConv and BRA mechanisms into the YOLOv11 architecture, NSYOLO effectively addresses critical challenges in nanoparticle imaging, such as complex boundary structures, dense particle distributions, and inadequate global context modeling. Experimental results indicate that NSYOLO significantly enhances segmentation accuracy compared to the baseline model and surpasses several existing open-source nanoparticle segmentation methods. Furthermore, the developed web-based application augments the practicality and scalability of the proposed approach for materials characterization workflows.

Despite these promising results, several limitations persist. The model's performance may deteriorate in scenarios characterized by significant particle overlap or inadequate boundary contrast, and its generalization capability is inherently affected by the diversity of the training data. Furthermore, the current framework has been validated on a limited set of representative morphologies, potentially restricting its direct applicability to more complex nanoparticle systems. Additionally, the integration of advanced attention mechanisms introduces further computational overhead, which may constrain deployment in resource-limited or real-time analysis scenarios.

Future research will concentrate on enhancing the robustness and generalizability of the model by integrating more diverse datasets, encompassing a broader spectrum of nanoparticle morphologies and imaging conditions. Additionally, the exploration of data-efficient learning strategies, such as transfer learning or few-shot learning, may further augment the model's adaptability to untrained structures. Finally, model optimization techniques, including network pruning, knowledge distillation, and lightweight attention designs, will be examined to decrease computational costs while preserving high segmentation performance.

Supplementary data

This article provides supplementary figures and tables. Supplementary materials include representative SEM/TEM nanoparticle datasets (Figure S1), ablation studies of DSConv and BADSConv with training loss analysis (Figure S2 and Table S1), visual segmentation comparisons among different models (Figures S3–S6), robustness evaluation under challenging imaging conditions and applicability assessments across different electron microscopy imaging modes (Figures S7–S8), out-of-distribution generalization analyses under different confidence thresholds (Figure S9 and Table S4), model training configurations, hardware/software environments, and hyperparameter settings (Tables S2–S3), as well as comprehensive quantitative analysis results generated using the developed interactive nanoparticle analysis system (Figures S10–S12).

Data availability statement

The data will be made available upon request from the corresponding author.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools (ChatGPT) only to improve language and readability. The authors take full responsibility for the content of the manuscript.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (grant no. 22472155) and the Scientific Research Fund of Zhejiang Provincial Education Department (grant no. Y202456921).

Authors' contribution

Data curation, investigation, visualization, software, writing—original draft, Xiao Han; validation, project administration, funding acquisition, Linfeng Jin; resources, formal analysis, supervision, Yijun Zhong; conceptualization, methodology, writing—review and editing, funding acquisition, Changfa Guo; All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare that they have no competing financial interests or personal relationships that could have influenced the work reported in this study.

References

[1] Afolabi RO. A comprehensive review of nanosystems' multifaceted applications in catalysis, energy, and the environment. *J. Mol. Liq.* 2024, 397:124190.

[2] Singh NB, Kumar B, Usman UL, Susan MABH. Nano revolution: exploring the frontiers of nanomaterials in science, technology, and society. *Nano-Struct. Nano-Objects* 2024, 39:101299.

[3] Nyabadza A, McCarthy É, Makhesana M, Heidarinassab S, Plouze A, *et al*. A review of physical, chemical and biological synthesis methods of bimetallic nanoparticles and applications in sensing, water treatment, biomedicine, catalysis and hydrogen storage. *Adv. Colloid Interface Sci.* 2023, 321:103010.

[4] Shahzadi S, Fatima S, Shafiq Z, Janjua MRSA. A review on green synthesis of silver nanoparticles (SNPs) using plant extracts: a multifaceted approach in photocatalysis, environmental remediation, and biomedicine. *RSC Adv.* 2025, 15(5):3858–3903.

[5] Huang J, Liu J, Wang J. Optical properties of biomass-derived nanomaterials for sensing, catalytic, biomedical and environmental applications. *TrAC Trends Anal. Chem.* 2020, 124:115800.

[6] Su D. Advanced electron microscopy characterization of nanomaterials for catalysis. *Green Energy Environ.* 2017, 2(2):70–83.

[7] Karatzia A, Loza K, Prymak O, Heggen M, Epple M. Thermal behavior of silver–gold core–shell nanocubes: *In situ* X-ray diffraction and *in situ* electron microscopy (SEM and TEM). *J. Phys. Chem. C* 2023, 127(39):19620–19628.

[8] Merugu R, Gothalwal R. Microscopic techniques for characterisation of nanomaterials: a minireview. *Mater. Today Proc.* 2021, 47:4753–4757.

[9] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to Image J: 25 years of image analysis. *Nat. Methods* 2012, 9(7):671–675.

[10] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern*. 1979, 9(1):62–66.

[11] Canny J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2009, 8(6):679–698.

[12] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.* 1991, 13(6):583–598.

[13] Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* 2023, 622:178–210.

[14] Neven D, De Brabandere B, Georgoulis S, Proesmans M, Van Gool L. Towards end-to-end lane detection: an instance segmentation approach. In *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, June 26–30, 2018, pp. 286–291.

[15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39(4):6408–651.

[16] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, October 5–9, 2015, pp. 234–241.

[17] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: a nested U-Net architecture for medical image segmentation. *IEEE Trans. Med. Imaging* 2019, 39(6):1856–1867.

[18] Huang H, Lin L, Tong R, Hu H, Zhang Q, *et al.* UNet 3+: a full-scale connected U-Net for medical image segmentation. *arXiv* 2020, arXiv:2004.08790.

[19] Chen L. Papandreou G. Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *arXiv* 2017, arXiv:1606.00915.

[20] Chen L, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv* 2017, arXiv:1706.05587.

[21] Chen L, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv* 2017, arXiv:1802.02611.

[22] Bolya D, Zhou C, Xiao F, Lee YJ. CT: real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Republic of Korea, October 27–November 2, 2019, pp. 9157–9166.

[23] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *arXiv* 2017, arXiv:1703.06870.

[24] Sun Z, Shi J, Wang J, Jiang M, Wang Z, *et al.* A deep learning-based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale* 2022, 14(28):10761–10772.

[25] Kim H, Han J, Han TYJ. Machine vision-driven automatic recognition of particle size and morphology in SEM images. *Nanoscale* 2020, 12(37):19461.

[26] Glaubitz C, Bazzoni A, Ackermann-Hirschi L, Baraldi L, Haeffner M, *et al.* Leveraging machine learning for size and shape analysis of nanoparticles: a shortcut to electron microscopy. *J. Phys. Chem. C* 2024, 128(1):421–427.

[27] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. *arXiv* 2024, arXiv:2410.17725.

[28] Qi Y, He Y, Qi X, Zhang Y, Yang G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, October 2–6, 2023, pp.6047–6056.

[29] Zhu L, Wang X, Ke Z, Zhang W, Lau RWH. BiFormer: vision transformer with bi-level routing attention. *arXiv* 2023, arXiv:2303.08810.

[30] Park YH, Choi H, Lee KB, Kim H. Curriculum learning-driven YOLO for tumor detection in ultrasound using hierarchically zoomed-in images. *Appl. Sci.* 2025, 15(19):10337.

[31] Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* 2008, 77(1):157–173.

[32] Dai J, Qi H, Xiong Y, Li Y, Zhang G, *et al.* Deformable convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 22–29, 2017, pp. 764–773.

[33] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc. IEEE* 2002, 86(11):2278–2324.

[34] Li Q, Jia X, Zhou J, Shen L, Duan J. Rediscovering BCE loss for uniform classification. *arXiv* 2024, arXiv:2403.07289.

[35] Zheng Z, Wang P, Ren D, Liu W, Ye R, *et al.* Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* 2021, 52(8):8574–8586.

[36] Girshick R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, December 7–13, 2015, pp. 1440–1448.

[37] Li X, Wang W, Wu L, Chen S, Hu X, *et al.* Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *arXiv* 2006, arXiv:2006.04388

[38] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis*. 2010, 88(2):303–338.

[39] Lin T, Maire M, Belongie S, Hays J, Perona P, *et al*. Microsoft COCO: common objects in context. In *proceedings of the European Conference on Computer Vision 2014*, Zurich, Switzerland, September 6–12, 2014, pp. 740–755.

[40] Wang C, Wang Y, Graser J, Zhao R, Gao F, *et al*. Solution-based carbohydrate synthesis of individual solid, hollow, and porous carbon nanospheres using spray pyrolysis. *ACS Nano* 2013, 7(12):11156–11165.

[41] Mishra SR, Tracy JB. Sequential actuation of shape-memory polymers through wavelength-selective photothermal heating of gold nanospheres and nanorods. *ACS Appl. Nano Mater.* 2018, 1(7):3063–3067.

[42] Pan J, Wang W, Ji M, Xing X, Lu Z. Robust synthesis of silver nanocubes in oil phase. *Cryst. Growth Des.* 2023, 23(4):2203–2208.

[43] Huang Q, Zhang D, Yu H, Ding Y, Xia Y. Seeing is believing: how does the surface of silver nanocubes change during their growth in an aqueous system. *Nano Lett.* 2025, 25(17):7115–7120.

[44] Gao F, Li H, Chen Z, Yi Y, Nie S, *et al.* A chemical autonomous robotic platform for end-to-end synthesis of nanoparticles. *Nat. Commun.* 2025, 16(1):7558.

[45] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, *et al.* Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, Paris, France, October 2–6, 2023, pp. 4015–4026.

[46] Genc A, Marlowe J, Jalil A, Belzberg D, Kovarik L, *et al.* A versatile machine learning workflow for high-throughput analysis of supported metal catalyst particles. *Ultramicroscopy* 2025, 271:114116.