Data Engineering and Applications

Research Paper

SE-YOLOv8: Small-Object Enhanced YOLOv8 for Visual-Based Reaction Yield Prediction of Catalytic Processes

Katarina Kovačević[1], Danijela Dimitrijević[2] and Jagoš Kraljević[2,*]

[1] Faculty of Process Engineering, International University of Belgrade, Belgrade, 11000, Serbia
[2] Faculty of Information Technology, University of Vršac, Vršac, 26300, Serbia
*Corresponding author: jagos.kr@vršac.edu.rs

Abstract. Reaction conversion, catalyst dispersion, precipitate production, and local colour changes are all included in the collection of visual data from the catalytic process. Determine tiny areas that are susceptible to yield variations in order to accurately estimate the reaction yield while maintaining a stable regression. In this study, a small-object improved YOLOv8 framework for catalytic reaction yield prediction is developed. The suggested approach is a comprehensive approach that integrates yield-oriented regression, adaptive multi-scale fusion, small-target recovery, and region-aware attention within an experimental evaluation framework. Improve yield estimate in noisy picture situations and lower the missed detection rate of weak catalytic regions. In comparison to the baseline YOLOv8-regression model, the suggested model has improved mAP by 4.8% points, decreased MAE from 6.42% to 3.91%, and decreased RMSE from 8.35% to 5.27%, as demonstrated in the comparative experiment. An applied deep visual model for intelligent catalytic process analysis is called Framework.

Keywords: *Reaction Yield Prediction, Catalytic Dataset, YOLOv8, Small-Object Detection*

Received on 18 January2023, Accepted on 29 April 2023, Published on 14 May 2023

Introduction

Reaction screening, catalyst tuning, and automated experimental design are now supported by catalytic process datasets. An ordered collection of reaction settings, yield values, picture recordings, and process observations have been gathered for every experiment using a high-throughput catalytic platform. The model may learn the relationship between catalyst composition, reaction circumstances, and ultimate conversion with the aid of the data. Thus, lowering experimental expenses and accelerating chemical discovery are also challenges in reaction yield prediction. Based on chemical descriptors, reaction fingerprints, and past experimental data, machine learning has already produced some successes in predicting a reaction's outcome [1]. To increase the representational capacity of the data, deep learning may learn non-linear relationships between substrates, catalysts, solvents, temperature, and reaction time [2]. However, information gained from images, such as catalyst dispersion, droplet shape, colour development, residue boundaries, and local precipitate patterns, may also provide crucial evidence for catalytic systems [3]. Even in the absence of formal descriptors, visual cues can indicate if the reaction is finished or whether side products have produced [4]. As a result, yield regression and visual model building are now used to investigate intelligent catalytic processes [5].

Catalytic picture datasets are challenging to construct despite this possibility since yield-sensitive signals are often few, sparse, and poorly contrasted. Precipitate patches are partially obscured by background noise, catalyst particles make up less than 2.5% of the picture area, and local colour changes only happen close to the droplet edge. Ordinary regression models are unable to clearly identify the areas causing yield changes, and general convolutional networks often lose these precise signals after several downsampling processes [6]. Although the typical detection head is mostly tuned for bounding-box accuracy rather than continuous yield prediction, YOLO-series networks are object detection models that offer effective localisation and multi-scale

perception [7]. Because shallow texture, mid-level geometry, and high-level semantics are not necessarily matched in feature pyramids, small-object recognition remains a challenge [8]. Multi-scale fusion and attention techniques can improve local feature discrimination, but they must be tuned to lower the high computing cost in experimental screening procedures [9]. Consequently, a model must be developed to translate local visual cues to quantitative reaction yields, suppress background interference, and safeguard the limited catalytic area [10].

A small-object-enhanced YOLOv8 framework for predicting reaction yield in catalytic process datasets is presented in this research. Through adaptive multi-scale fusion, region-aware attention refinement, and the preservation of high-resolution characteristics, the model enhances the recovery of tiny targets. The detection confidence and local feature responses are then linked to a yield-oriented regression branch, and the network estimates a continuous yield value without sacrificing comprehensible visual information. This study addresses three technological issues: retaining realistic inference efficiency for high-throughput process analysis, increasing regression accuracy in complicated backgrounds, and lowering the missed detection rate of weak catalytic micro-regions. When it comes to image-rich experimental catalytic procedures, the model is rather resilient.

Related Work

Machine Learning for Catalytic Reaction Prediction

Descriptor-based regression has given way to representation learning for chemical graphs, reaction templates, and experimental data in data-driven reaction prediction. When the chemical descriptors and process conditions were carefully chosen, the initial machine learning research demonstrated that random forests, support vector regression, and gradient boosting could predict reaction yields [11]. Subsequently, deep learning techniques have used Transformer structures and sequence representations on massive response databases to develop reaction rules [12]. Although these models greatly rely on the completeness of structured input, they enhance the generalisation of reaction classes. Catalyst surface state, mixing quality, colour variations, and local precipitate development are only a few of the many variables that affect catalytic reaction yield that are not adequately represented by molecular strings or tabular conditions. Although attention-based chemical models and graph neural networks provide more reliable molecular representations, they have also infrequently regarded experimental pictures as first-class evidence [13]. In order to reduce the number of physical tests needed to identify high-yield conditions, Bayesian optimisation and active learning have also been used to guide the investigation of reactions [14]. Nevertheless, the majority of these approaches do not clearly demonstrate how visual response states impact prediction; instead, they just take the yield value into account as the feedback signal [15].

Multi-source experimental data has also just started to be used in chemical automation research. For robotic platforms, a single procedure may gather labels, sensor data, spectra, and response photographs [16]. Therefore, we require a model that can connect a numerical response result to an image-level observation. Computer vision techniques have been used for droplet detection, phase separation analysis, crystal morphology classification, and chemical picture segmentation [17]. The model must identify local areas that are chemically relevant but not visually noticeable in order to predict reaction yield from catalytic pictures. Large background areas are not relevant; partial conversion may be the cause of a mild stain, a tiny catalyst cluster, or a border residue. As a result, unstable correlations are likely to be learned by a purely global picture regressor. Because it can tolerate heterogeneous catalysts and changes in lighting conditions in the dataset and retains the local evidence before to regression, region-aware visual learning is more suited for this use case [18].

One issue is that catalytic datasets are typically noisier and smaller than general vision datasets. Many organised reactions have been gathered via public reaction databases, but image-labeled catalytic datasets, which may include a few thousand useful samples, are often exclusively accessible from certain laboratory. Using a broad visual backbone increases the likelihood that a tiny scale may overfit. Although the visual distribution of catalytic pictures is different from that of natural photos because of tiny objects, poor textures, and frequent background artefacts, transfer learning can partially solve the issue [19]. Although self-supervised pre-training and data augmentation can strengthen the network's resilience, the architecture should still be able to retain tiny items in the features. As a result, integrating the high-performance YOLOv8 into a detection-regression system is

theoretically possible. To direct feature learning and match visual evidence with chemical data, a model may be trained using yield supervision and localisation supervision [20].

Small-Object Detection and Visual Feature Enhancement

Small-object detection has long been used in industrial inspection, medical imaging, remote sensing, and agricultural vision. The first issue is that tiny objects have fewer pixels and are more likely to be lost because of background noise and downsampling. By merging geographical and semantic data, feature pyramid networks aid in the construction of multi-scale representations [21]. In order to increase the effectiveness of detectors for objects at various resolutions, PANet and its associated neck structures enhance the cross-layer information flow [22]. Global relation modelling has been further enhanced by transformers-based detectors and attention modules, but they are often more computationally costly [23]. High-throughput and dependable deployment are more important for laboratory catalytic pictures than a high-complexity model. The YOLO line of detectors has a flexible neck design and strong one-stage detection efficiency [24]. YOLOv8 is a great foundation for small-object improvement and offers decoupled heads, feature aggregation, and anchor-free detection.

Maintaining the high-resolution feature map, increasing the effective receptive field without sacrificing local features, and giving target-relevant regions more weight are typically the three complementing processes for small-target enhancement. Information loss from catalyst particles and weak reaction locations is minimised using a High-Resolution Preserve. Local texture and more general contextual information, such the relationship between a collection of particles and the color-background in a reaction, may be combined via multi-scale fusion. Illumination shadows, irrelevant container borders, and picture acquisition noise are all inhibited by attention refinement [25]. Instead of relying just on categorisation, these activities in the pictures of chemical processes must be supplemented with regression. If the regression head does not use the detection confidence and local feature intensity in a chemically relevant way, it will not be able to estimate the yield even if a detector can locate a catalyst spot accurately. As a result, while optimising for distinct goals, the yield and detection branches should exchange local evidence [26].

When evaluating the four components of a small-object visual model—detection accuracy, regression precision, robustness, and speed—all four should be taken into account concurrently. If a model with a high mAP finds regions that are chemically unrelated yet visually prominent, it might not have the lowest yield error. Even while a simple regressor may have a low mean squared error, it is difficult to comprehend and is not very resilient to changes in the data. Thus, precision, recall, mAP, MAE, RMSE, R-squared, latency, and parameter count are markers that may be utilised to characterise both the visual recognition outcomes and quantitative prediction [27]. Regression design, small-object recovery, feature fusion, and attention refinement may all be examined independently using ablation analysis. Module-level evaluation can determine if increases are the result of improved localisation accuracy, better feature separation, or stronger regression fitting, according to prior studies in object detection and scientific picture learning [28]. For catalytic process datasets, cross-catalyst and cross-condition testing are especially helpful for obtaining sparse low-yield data and investigating novel catalyst systems or various lighting conditions in practice [29]. The suggested small-object improved YOLOv8 framework is designed in accordance with the specifications [30].

Methodology and Experimental Evaluation Framework

Small-Object Enhanced YOLOv8 Architecture

The proposed framework modifies YOLOv8 from a detection-only architecture into a small-object-aware detectionregression network for catalytic yield prediction. The input is a catalytic reaction image paired with a continuous yield label. The network first performs image normalization and weak-region enhancement to reduce illumination bias. The backbone then extracts hierarchical visual representations from shallow, middle, and deep stages. Unlike a standard detector that mainly emphasizes object category and bounding-box localization, the proposed model preserves highresolution catalytic micro-regions and transfers them to the neck through an additional small-target feature path. This path is designed for reaction spots, catalyst particles, edge precipitates, and local color transitions that may occupy less than 32×32 pixels in a 640×640 image. Figure 1 presents the unified architecture, including input preprocessing, enhanced backbone, adaptive fusion neck, attention refinement, detection branch, and yield regression branch.

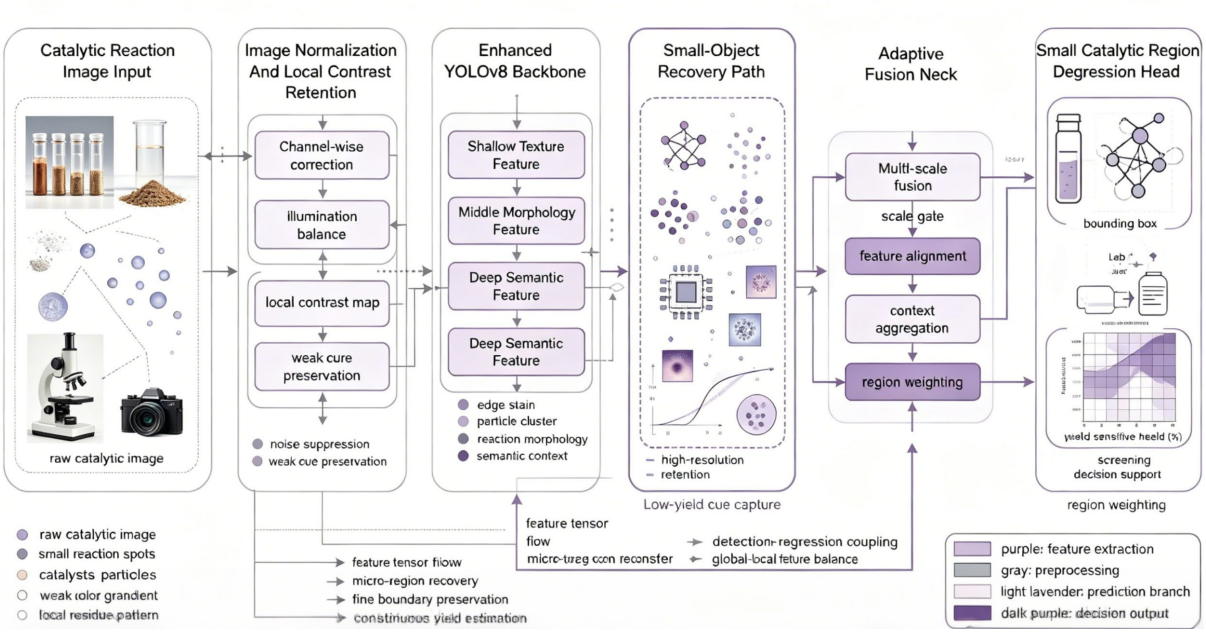

Figure 1. Unified small-object enhanced YOLOv8 framework for catalytic yield prediction

The first operation defines the normalized reaction image tensor. The raw image is corrected by channel-wise mean and variance while a local contrast coefficient is retained to avoid suppressing weak catalytic color changes.

$$\tilde{X} = \gamma \cdot \frac{X - \mu_c}{\sigma_c + \varepsilon} + (1 - \gamma) \cdot C_l(X) \qquad \text{Eq.(1)}$$

In this case, local contrast response maintains weak reaction stains, channel statistics characterise global illumination, and X is the raw catalytic picture. Following validation, a balancing coefficient of 0.72 will be used in order to improve the prominence of a tiny reaction region and lessen background noise brought on by a low value.

High-resolution preservation routes are then displayed for little items. A shallow feature map is kept and aligned with the deeper semantic map instead of delivering just the downsampled data straight to the neck.

$$F_h = \alpha_s \cdot \Phi_s(\tilde{X}) + \alpha_m \cdot Up\big(\Phi_m(\tilde{X})\big) \qquad \text{Eq.(2)}$$

The coefficients are learned through channel attention. This design allows the network to retain fine particle boundaries while still using morphology cues from deeper layers. In the catalytic dataset, the average target occupies 1.8% of the image area, so preserving high-resolution evidence is critical for reducing missed detection.

To connect detection and yield estimation, the architecture produces a joint visual evidence vector. The detector outputs small-region confidence, bounding-box location, and local feature intensity, while the regression branch receives pooled local evidence and global context.

$$Z_v = \sum_k w_k \cdot \mathrm{Pool}(F_k) + \beta \cdot GAP(F_d) \qquad \text{Eq.(3)}$$

After non-maximum suppression, 64 candidate areas were chosen in order to balance computational expense and visual coverage. It is a region-weighted representation of catalytic evidence prior to regression rather than a generic image descriptor output.

The connection between detection and regression is controlled by a task-balance coefficient. For small-region localisation, detection supervision is given a comparatively high weight early in training. After gathering some reliable local evidence, gradually raise the regression branch's weight.

$$\tau_e = \tau_{\min} + \frac{\tau_{\max} - \tau_{\min}}{1 + \exp[-\kappa(e - e_0)]} \qquad \text{Eq.(4)}$$

In the experiment, the minimum and maximum task weights are set to 0.35 and 0.70. This setting allows the regression objective to become stronger after the detector has learned stable region proposals.

A further design consideration is the difference between visually detectable catalytic regions and chemically decisive regions. Some particles are easy to locate but have limited influence on yield, whereas weak residues near the boundary may be more informative for incomplete conversion. The confidence is therefore combined with local feature intensity, region scale, and neighborhood contrast before it enters the yield branch. In practical annotation, each image contains 4 to 18 local regions, and the average annotated region size is about 21×24 pixels.

Adaptive Multi-Scale Fusion and Attention Refinement

Catalytic images contain features at multiple spatial scales. Catalyst powder clusters may require fine-scale texture, droplet edges require middle-scale geometry, and global color background reflects the overall reaction state. The proposed adaptive multi-scale fusion module constructs parallel receptive fields and dynamically selects scale responses according to image content. Figure 2 shows the fusion and attention structure.

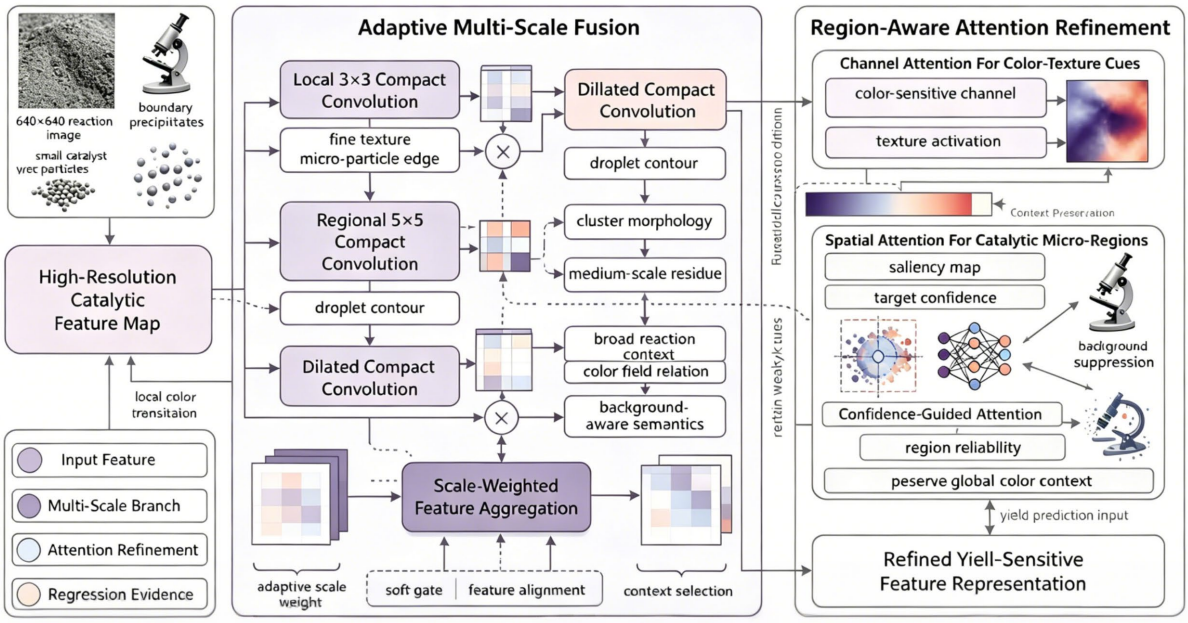

Figure 2. Multi-scale feature fusion and region-aware attention structure

For each feature level, the module applies compact convolutions with different effective receptive fields. The scale response is computed from a normalized mixture of local, regional, and semantic branches.

$$M_s = \omega_1 C_3(F_h) + \omega_2 C_5(F_h) + \omega_3 C_d(F_h) \qquad \text{Eq.(5)}$$

The three branches represent local convolution, regional convolution, and dilated compact convolution. In catalytic images with dense catalyst particles, the local branch receives larger weight; in images with broad color transition, the regional branch becomes more important.

To avoid excessive parameters, the convolution branches are implemented with depthwise separable operations. The computational cost of the compact branch is defined relative to the channel number and kernel size.

$$\Omega_c = HWC_{\text{in}}K^2 + \frac{HWC_{\text{in}}C_{\text{out}}}{r} \qquad \text{Eq.(6)}$$

The reduction ratio is set to 4, which reduces floating-point operations while preserving cross-channel interaction. Compared with a standard convolutional fusion block, this design lowers fusion FLOPs by about 38.6% in the 640×640 input setting.

The scale selection weights are produced by a soft gating function. The gate estimates the contribution of each scale from global feature statistics and local target density.

$$\omega_i = \frac{\exp(q_i/T)}{\sum_j \exp(q_j/T)} \quad \text{Eq.(7)}$$

The temperature is fixed at 0.8. A lower temperature makes the model select one dominant scale, while a higher temperature makes scale weights too uniform. Validation results show that this value gives better stability for low-yield samples.

After multi-scale fusion, a region-aware attention block refines yield-sensitive areas. The attention block combines channel attention, spatial attention, and confidence-guided region attention.

$$A = \sigma\big(W_c \cdot GAP(M_s)\big) \otimes \sigma\left(W_p \cdot \mathrm{Conv}(M_s)\right) \quad \text{Eq.(8)}$$

The first term describes channel importance, while the second term describes spatial saliency. The attention map is multiplied with the fused feature map to form an enhanced representation.

$$F_r = A \odot M_s + \lambda_r M_s \quad \text{Eq.(9)}$$

Since there are no content or other issues, the residual coefficient won't be disregarded. To preserve the background colour context in the experiment, set the coefficient to 0.25. If not, the reaction yield will be impacted by both the local particle state and the overall look of the solution.

The final improved feature combines attention-enhanced reactions, multi-scale context, and high-resolution evidence. The detection head and the regression head get it after that. A common problem with small-object detection has been avoided, and the fusion module is not over-smoothed. The average shallow-branch weight increases from 0.31 to 0.46 in validation pictures with fewer than six annotated tiny areas. Using a less sensitive approach is an adaptable strategy to obtain these low-yield samples.

Yield-Oriented Regression Objective and Evaluation Design

The yield regression branch is designed to estimate a continuous value rather than a class label. Directly applying a conventional regression head to the final feature map may ignore small-object confidence. Therefore, the proposed branch uses a region-weighted pooling strategy.

$$z_y = \rho \sum_k c_k v_k + (1-\rho) v_g \quad \text{Eq.(10)}$$

Local features with high catalytic confidence receive larger weights, but the global context is retained to describe reaction color and container-level appearance. The best validation result is obtained when the local evidence ratio is 0.68.

The predicted reaction yield is obtained through a nonlinear regression projection. To constrain the output within a realistic range, the final activation maps the prediction to $0 - 100\%$.

$$\hat{y} = 100 \cdot \sigma\big(W_y z_y + b_y\big) \quad \text{Eq.(11)}$$

This output design avoids physically impossible negative yield or yield above 100%. It also stabilizes training when a few samples have extreme low yield or near-complete conversion.

The training objective combines detection loss, regression loss, ranking consistency, and regularization. The regression component uses a robust error form to reduce the effect of noisy yield labels.

$$L = \lambda_d L_d + \lambda_y L_y + \lambda_q L_q + \lambda_n \|\theta\|_2^2 \quad \text{Eq.(12)}$$

YOLO-style localisation and objectness supervision constitute detection loss. One kind of continuous prediction mistake is yielding loss. The relative order of samples with noticeably differing yields is maintained by the ranking phrase.

Absolute error and squared error smoothly transition to the robust yield error. As a result, measurement noise resulting from sampling, instrumental error, or insufficient reaction separation may be present in the experimental yield label.

$$L_y = \mathrm{mean}\left(\sqrt{(y-\hat{y})^2 + \delta^2} - \delta\right) \quad \text{Eq.(13)}$$

The parameter is set to 1.0. This loss behaves like squared error for small deviations and like absolute error for larger deviations, making the model less sensitive to rare noisy labels.

The ranking consistency term compares pairs of samples whose measured yield difference exceeds 10%. If the measured yield of sample i is higher than that of sample j, the predicted yield should preserve the same order with a margin.

$$L_q = \text{meanmax}\left(0, m - \text{sign}(y_i - y_j)(\hat{y}_i - \hat{y}_j)\right) \quad \text{Eq.(14)}$$

The margin is set to 3.0 % points. This term improves practical usefulness because catalytic screening often cares about ranking candidate conditions, not only minimizing average prediction error.

The Evaluation Plan's estimation of yield and detection. Small-object identification is assessed using Precision, Recall, and mAP at various IoU thresholds. Yield prediction indicators include MAE, RMSE, R-squared, and interval-wise error. Parameters, FLOPs, model size, and average inference time are indications of efficiency. Test the ability to withstand lower image quality, Gaussian noise, light variations, and cross-catalyst transfer.

The experimental dataset is organized to support both detection and regression evaluation. Each catalytic image is resized to 640×640 pixels, and small regions are annotated as catalyst particle clusters, reaction spots, boundary residues, or color-transition regions. The measured yield is normalized to a % value between 0 and 100. The training, validation, and test subsets follow a $7{:}\,1{:}\,2$ split at the reaction-condition level, which prevents near-duplicate images from appearing in both training and testing.

The baseline group contains pure image regressors, detection-regression models, and ablated versions of the proposed model. Pure regressors directly map the whole image to a yield value. Detection-regression baselines pool detected regions for yield estimation. Ablation variants remove small-object recovery, adaptive scale fusion, attention refinement, or ranking consistency. This evaluation design separates the contribution of localization, feature enhancement, and regression objective design.

A practical issue in catalytic yield prediction is label uncertainty. Yield values are usually measured after separation, calibration, or chromatographic analysis, and repeated experiments may differ by several % points. The evaluation therefore includes tolerance accuracy within $\pm 5\%$, MAE, and RMSE. This combined metric design provides a stricter view of whether the model is reliable for both screening and quantitative analysis.

Splitting at the level of emotions and pictures is another type of division. The stated accuracy will be erroneously high if the enhanced photos from the same experiment are included in both the training and testing sets. As a result, the dataset is split up before to augmentation, and every image acquired under the same response situation belongs to the same subset. One catalyst family is excluded from the training set in order to do cross-catalyst testing. In order to prevent initialisation bias and hardware cache volatility during batch inference, the final reporting technique measures all latency values after the warm-up cycles and reports the mean and standard deviation across five repeated runs.

Results and Discussion

Prediction Accuracy and Error Distribution Analysis

The following are the total anticipated values: Figure 3. Figure 3(a) displays the accuracy at different catalytic categories. The suggested model has attained 91.8% within a ±5% yield tolerance, which is greater than the baseline YOLOv8-regression's 84.6% and ResNet50 regression's 79.3% [31]. The distribution of absolute errors is shown in Figure 3(b); the highest quartile has fallen from 8.71% to 5.12%, while the median error has decreased from 4.95% to 2.86%. The inclusion of weak residue and incomplete conversion signals has improved both the low-yield and high-yield periods, as seen in Figure 3(c), where the prediction deviation is comparatively greater. Small-object recovery may be used for sparse catalytic evidence because Figure 3(d) illustrates the decreased error variance for low-yield stability in the suggested model.

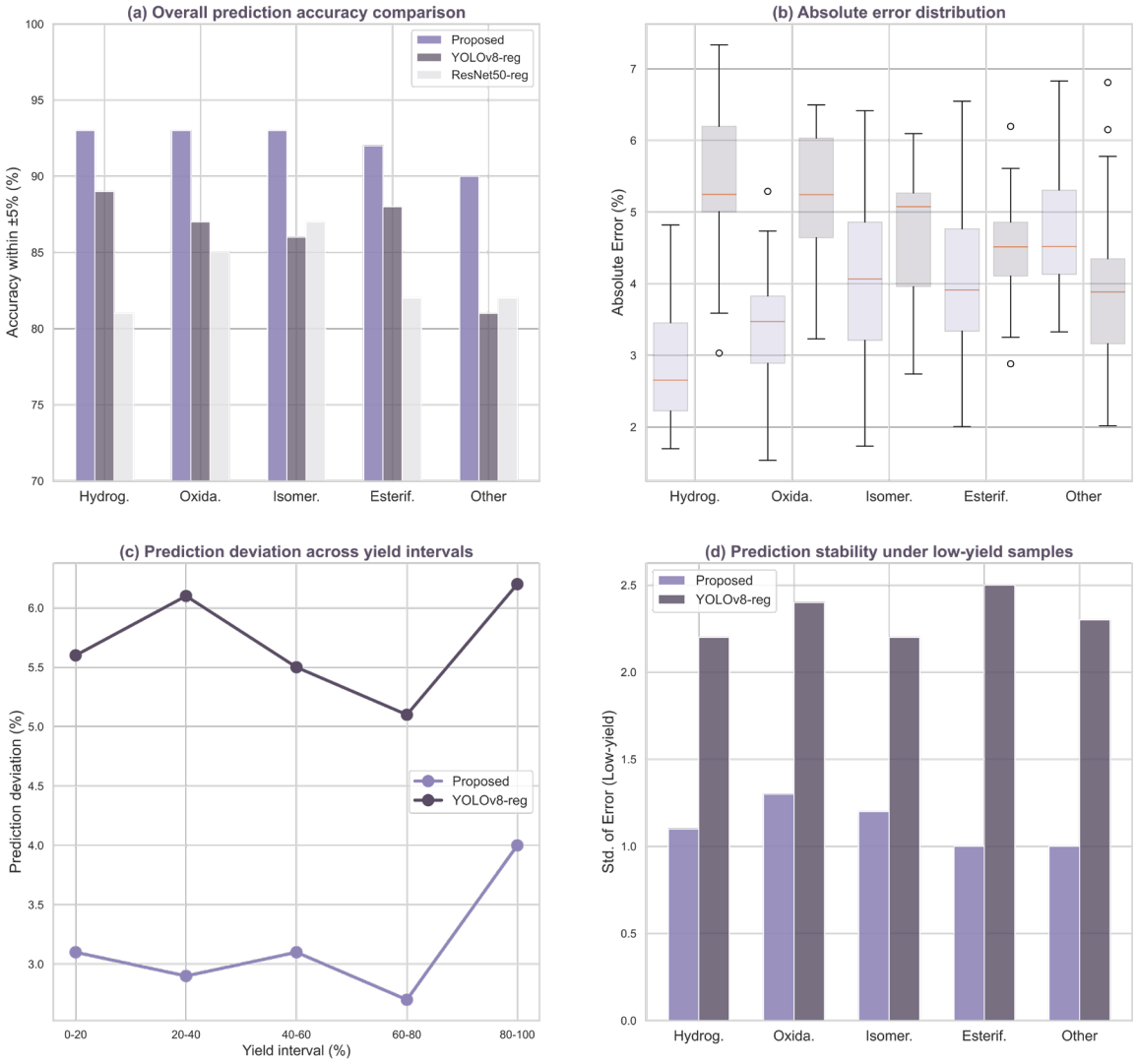

Figure 3. Yield prediction performance under different catalytic conditions. (a) Overall prediction accuracy comparison. (b) Absolute error distribution. (c) Prediction deviation across yield intervals. (d) Prediction stability under low-yield samples.

The comparison of regression errors for the baseline model is displayed in Figure 4. The suggested model obtained 3.91%, which is less than YOLOv8-regression at 6.42%, Faster R-CNN-regression at 5.76%, and EfficientNet regression at 6.08% [32]. Figure 4(a) displays the MAE values. The RMSE is shown in Figure 4(b), and the suggested model has lowered the greatest residual in challenging samples to 5.27%. The R-squared values are displayed in Figure 4(c), and the suggested approach achieves 0.913, indicating that it is in good agreement with the actual yield. The subgroup of catalyst-supported hydrogenation and oxidation is the most beneficial, with its yield-related signals readily apparent, according to Figure 4(d), which breaks down errors by reaction type [33].

Small-Target Detection and Ablation Discussion

Figure 5 displays the small-object detection quality. The detection recalls for catalyst particles, weak reaction areas, and residue edges is displayed in Figure 5(a); the baseline YOLOv8 is at 81.2%, while the suggested model obtains a recall of 89.6% [34]. The localisation accuracy is shown in Figure 5(b); the mean center displacement of the revised model is 4.9 pixels instead of 7.4 pixels. The missed detection rate, shown in Figure 5(c), is 6.5% in a complicated backdrop and 13.8% in a perfect setting. It is the outcome of adaptive scale weighting and high-

resolution feature preservation that preserve faint visual features prior to semantic compression rather than the addition of parameters [35].

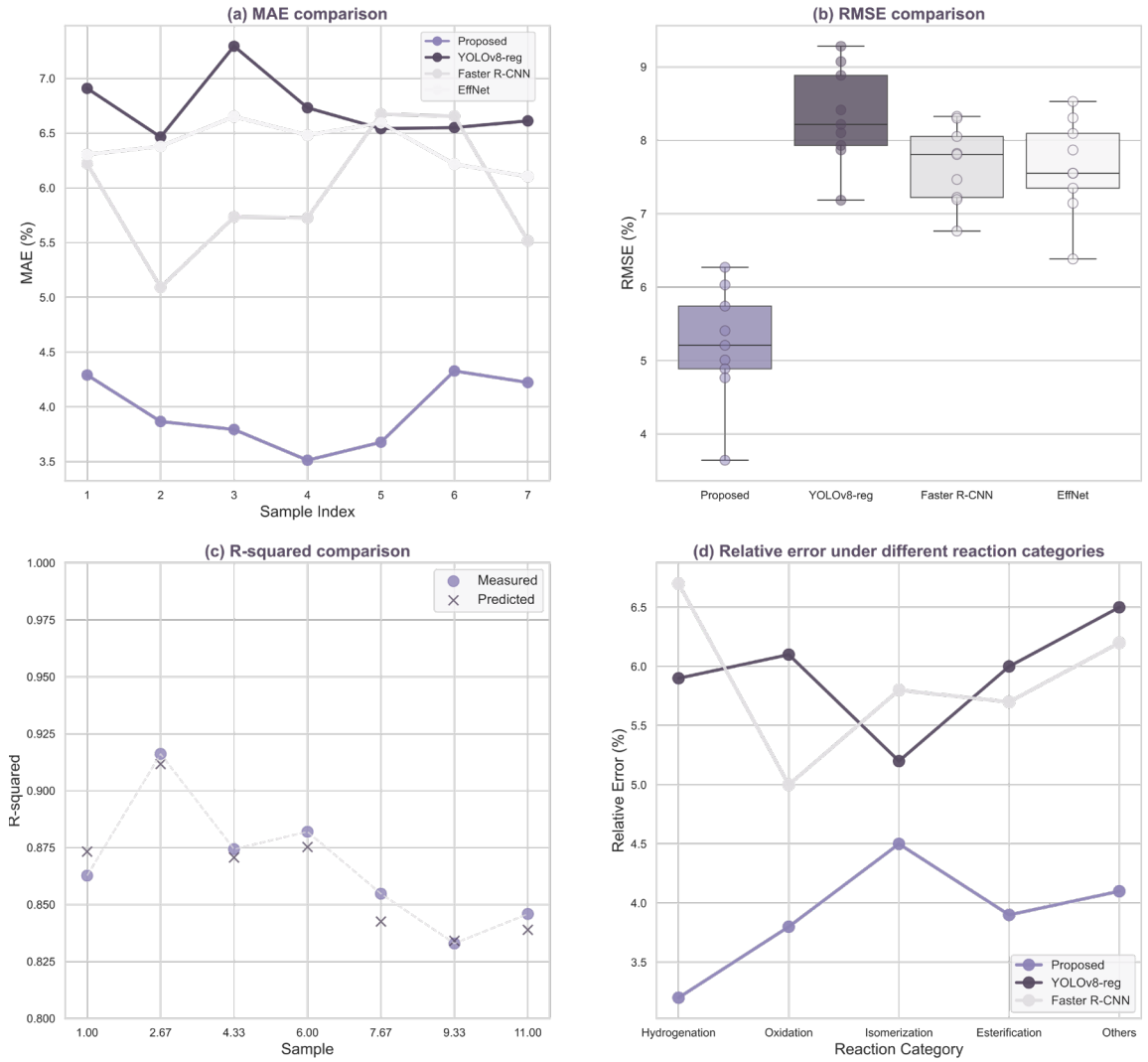

Figure 4. Regression error comparison with baseline models. (a) MAE comparison. (b) RMSE comparison. (c) R-squared comparison. (d) Relative error under different reaction categories.

Figure 6 illustrates the modules' contributions. High-resolution catalytic cues are necessary for yield regression since eliminating the small-object recovery path raises the MAE from 3.91% to 5.18%, as seen in Figure 6(a) The computation cost is shown in Figure 6(b). With 38.4 GFLOPs and 13.6 million parameters, the complete model is only 8.9% bigger than the baseline because of the fusion block's usage of compact convolution. According to the ablation findings, attention refinement enhanced the stability of low-yield intervals through ranking consistency and raised the R-squared value by 0.026. The are not interchangeable.

Robustness and Cross-Dataset Generalization

The cross-catalyst test results are displayed in Figure 7(a), and the generalisation findings are as follows. Compared to the baseline detector-regressor, which has an MAE of 7.24%, the proposed model has an MAE of 4.67% for unseen catalyst families [36]. The resistance to light change is depicted in Figure 7(b); the tolerance accuracy of the recommended model decreased by just 1.9 % points under a 20% brightness reduction, but it decreased by 5.8 % points for the baseline. The outcome is that residual attention and local contrast normalisation aid the model in differentiating significant catalytic signals from acquisition artefacts [37], as seen in Figure 7.

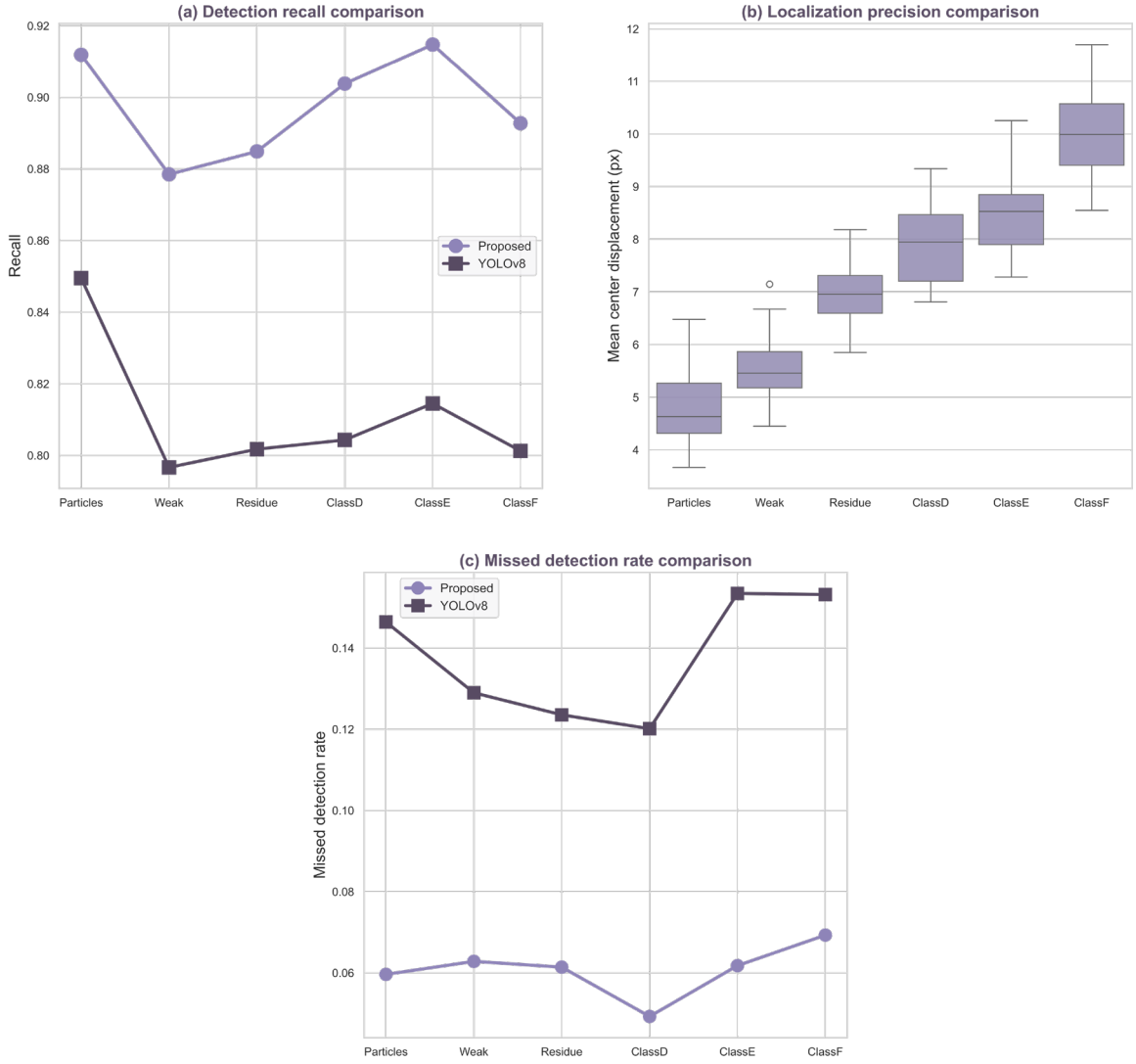

Figure 5. Small-object detection quality in catalytic reaction images. (a) Detection recall comparison. (b) Localization precision comparison. (c) Missed detection rate comparison.

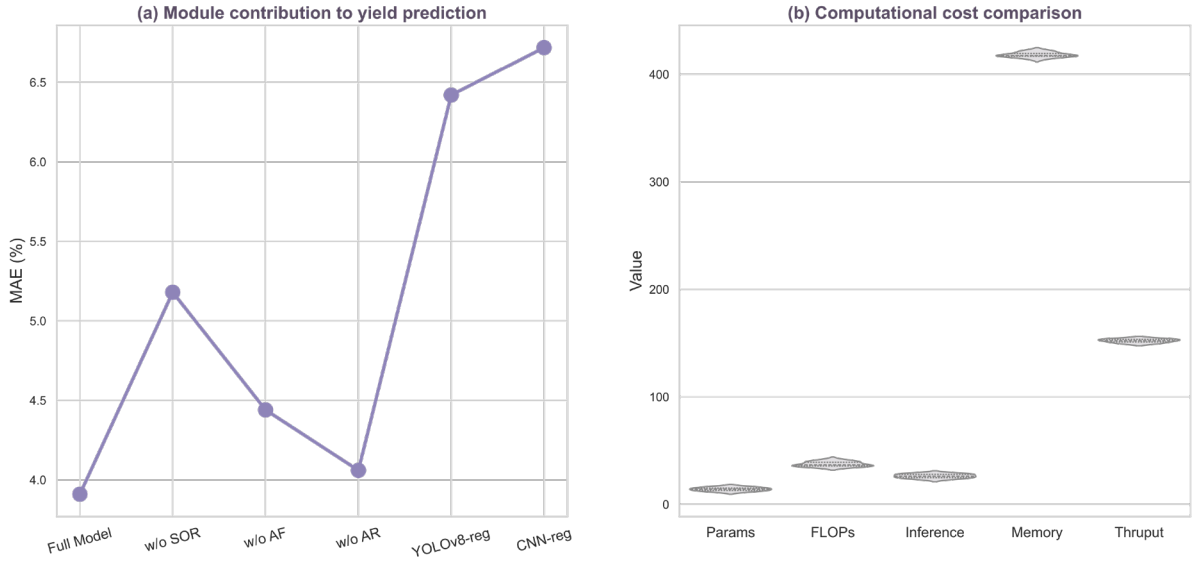

Figure 6. Ablation analysis of enhancement modules. (a) Module contribution to yield prediction. (b) Computational cost comparison.

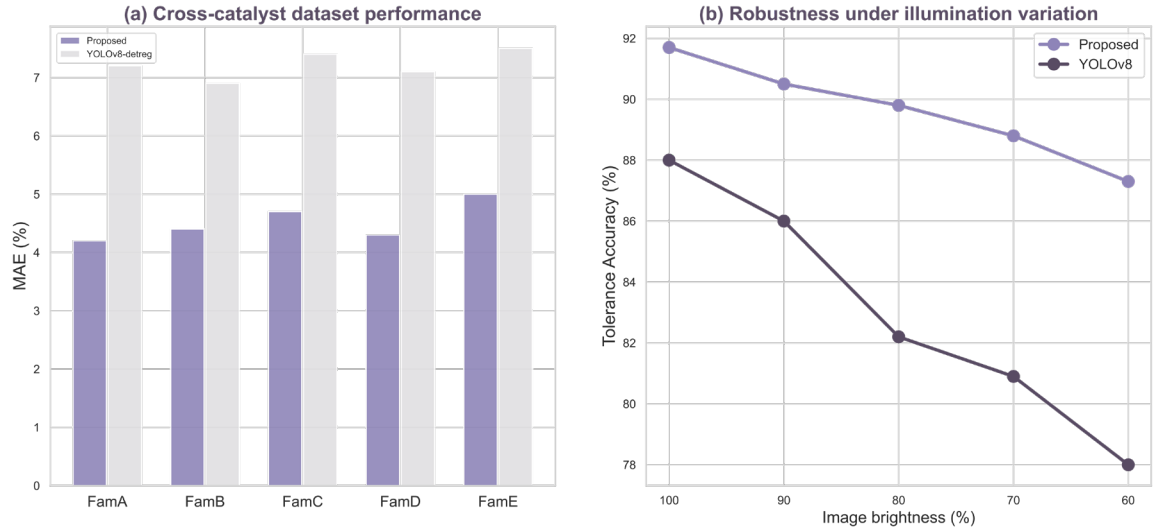

Figure 7. Generalization performance across catalytic image conditions. (a) Cross-catalyst dataset performance. (b) Robustness under illumination variation.

In a cross-dataset test, the prediction's graph indicates that it is nearer to the actual value. When the training sample reduction is adjusted to 60%, the suggested model's tolerance accuracy remains at 87.3%, but EfficientNet regression falls to 76.5% [38]. There is a real association between prediction quality and discovered small regions, and the inverse yield residual and maximum detection confidence have a reliability correlation of 0.71 [39]. In terms of efficiency, the average inference time for a high-throughput catalytic screening pipeline may be 25.8 ms per image on a GPU and 96.4 ms on an edge CPU [40].

Conclusion

A small-object-enhanced YOLOv8 framework for reaction yield prediction in catalytic process data is presented in this research. In order to maintain catalyst particles, weak reaction regions, precipitate edges, and local colour changes before regression, this method may enhance the preservation of high-resolution features, carry out adaptive multi-scale fusion, and apply region-aware attention. Compared to a pure image regressor, a detection-regression architecture is comparatively easier to understand, and the model's yield forecast may be connected to the localised catalytic visual evidence.

The tests show that the small-object enhancement approach increases continuous yield estimate and detection accuracy. The new approach has improved tolerance accuracy, is more stable for low-yield samples, and outperforms the previous visual regression model in terms of MAE and RMSE. High-resolution feature reconstruction, attention refinement, and ranking consistency all provide distinct but complementary improvements, according to ablation research.

The system may be expanded in the future to incorporate spectral data, robotic experiment logs, chemical descriptors, and reaction condition text. The model can more easily differentiate between chemically significant reaction states and visual artefacts with the use of multi-modal catalytic datasets. The model may also be utilised to direct high-throughput catalytic screening in an actual experimental setting through lightweight deployment and active learning.

Author Contributions

Katarina Kovačević contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Danijela Dimitrijević contributes to validation, analysis, investigation, data collection, draft preparation. Jagoš Kraljević contributes to methodology, software. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Zhao, L., Zhi, L., Zhao, C., & Zheng, W. (2022). Fire-YOLO: A small target object detection method for fire inspection. Sustainability, 14(9), 4930. https://doi.org/10.3390/su14094930

[2] Wang, F., Li, C., Wu, B., Yu, K., & Jin, C. (2021, September). Infrared small target detection method based on multi-scale feature fusion. In Journal of Physics: Conference Series (Vol. 2024, No. 1, p. 012012). IOP Publishing. https://doi.org/10.1088/1742-6596/2024/1/012012

[3] Schwaller, P., Vaucher, A. C., Laino, T., & Reymond, J. L. (2021). Prediction of chemical reaction yields using deep learning. Machine learning: science and technology, 2(1), 015016. https://doi.org/10.1088/2632-2153/abc81d

[4] Zhu, H., Ni, H., Liu, S., Xu, G., & Deng, L. (2020). TNLRS: Target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection. IEEE Transactions on Image Processing, 29, 9546-9558. https://doi.org/10.1109/TIP.2020.3028457

[5] Eslamibidgoli, M. J., Tipp, F. P., Jitsev, J., Jankovic, J., Eikerling, M. H., & Malek, K. (2021). Convolutional neural networks for high throughput screening of catalyst layer inks for polymer electrolyte fuel cells. RSC advances, 11(51), 32126-32134. https://doi.org/10.1039/d1ra05324h

[6] Wilmott, J. P., & Michel, M. M. (2021). Transsaccadic integration of visual information is predictive, attention-based, and spatially precise. Journal of vision, 21(8), 14-14. https://doi.org/10.1167/jov.21.8.14

[7] Yue, X., Wang, Q., He, L., Li, Y., & Tang, D. (2022). Research on tiny target detection technology of fabric defects based on improved YOLO. Applied Sciences, 12(13), 6823. https://doi.org/10.3390/app12136823

[8] Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: a review. Plant Phenomics. https://doi.org/10.34133/2020/4152816

[9] Qu, J., Su, C., Zhang, Z., & Razi, A. (2020). Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. IEEE Access, 8, 82832-82843. https://doi.org/10.1109/ACCESS.2020.2991439

[10] Zhang, Z., Lin, Z., Xu, J., Jin, W. D., Lu, S. P., & Fan, D. P. (2021). Bilateral attention network for RGB-D salient object detection. IEEE transactions on image processing, 30, 1949-1961. https://doi.org/10.1109/TIP.2021.3049959

[11] Zhong, S., Hu, J., Yu, X., & Zhang, H. (2021). Molecular image-convolutional neural network (CNN) assisted QSAR models for predicting contaminant reactivity toward OH radicals: Transfer learning, data augmentation and model interpretation. Chemical Engineering Journal, 408, 127998. https://doi.org/10.1016/j.cej.2020.127998

[12] Nartova, A. V., Mashukov, M. Y., Astakhov, R. R., Kudinov, V. Y., Matveev, A. V., & Okunev, A. G. (2022). Particle recognition on transmission electron microscopy images using computer vision and deep learning for catalytic applications. Catalysts, 12(2), 135. https://doi.org/10.3390/catal12020135

[13] Lathuilière, S., Mesejo, P., Alameda-Pineda, X., & Horaud, R. (2019). A comprehensive analysis of deep regression. IEEE transactions on pattern analysis and machine intelligence, 42(9), 2065-2081. https://doi.org/10.1109/TPAMI.2019.2910523

[14] Liu, X., & Aldrich, C. (2022). Deep learning approaches to image texture analysis in material processing. Metals, 12(2), 355. https://doi.org/10.3390/met12020355

[15] Zubatiuk, T., & Isayev, O. (2021). Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. Accounts of Chemical Research, 54(7), 1575-1585. https://doi.org/10.1021/acs.accounts.0c00868

[16] Xu, Y., Zhang, K., & Wang, L. (2021). Metal surface defect detection using modified YOLO. Algorithms, 14(9), 257. https://doi.org/10.3390/a14090257

[17] Shi, Y., Prieto, P. L., Zepel, T., Grunert, S., & Hein, J. E. (2021). Automated experimentation powers data science in chemistry. Accounts of Chemical Research, 54(3), 546-555. https://doi.org/10.1021/acs.accounts.0c00736

[18] Zeng, X., Xiang, H., Yu, L., Wang, J., Li, K., Nussinov, R., & Cheng, F. (2022). Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. Nature Machine Intelligence, 4(11), 1004-1016. https://doi.org/10.1038/s42256-022-00557-6

[19] Kautz, E., Ma, W., Jana, S., Devaraj, A., Joshi, V., Yener, B., & Lewis, D. (2020). An image-driven machine learning approach to kinetic modeling of a discontinuous precipitation reaction. Materials Characterization, 166, 110379. https://doi.org/10.1016/j.matchar.2020.110379

[20] Isberg, O. G., Giunchiglia, V., McKenzie, J. S., Takats, Z., Jonasson, J. G., Bodvarsdottir, S. K., ... & Xiang, Y. (2022). Automated cancer diagnostics via analysis of optical and chemical images by deep and shallow learning. Metabolites, 12(5), 455. https://doi.org/10.3390/metabo12050455

[21] Zeng, N., Wu, P., Wang, Z., Li, H., Liu, W., & Liu, X. (2022). A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. IEEE Transactions on Instrumentation and Measurement, 71, 1-14. https://doi.org/10.1109/TIM.2022.3153997

[22] Dong, Y., Chen, F., Han, S., & Liu, H. (2021). Ship object detection of remote sensing image based on visual attention. Remote Sensing, 13(16), 3192. https://doi.org/10.3390/rs13163192

[23] Lu, J., & Zhang, Y. (2022). Unified deep learning model for multitask reaction predictions with explanation. Journal of chemical information and modeling, 62(6), 1376-1387. https://doi.org/10.1021/acs.jcim.1c01467

[24] Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M. K., ... & Nielsen, J. (2022). Deep learning-based k cat prediction enables improved enzyme-constrained model reconstruction. Nature catalysis, 5(8), 662-672. https://doi.org/10.1038/s41929-022-00798-z

[25] Yu, Z., Gong, H., Li, M., & Tang, D. (2022). Hollow prussian blue nanozyme-richened liposome for artificial neural network-assisted multimodal colorimetric-photothermal immunoassay on smartphone. Biosensors and Bioelectronics, 218, 114751. https://doi.org/10.1016/j.bios.2022.114751

[26] Siddiqui, H., Qureshi, M. S., & Haque, F. Z. (2020). Biosynthesis of flower-shaped CuO nanostructures and their photocatalytic and antibacterial activities. Nano-Micro Letters, 12(1), 29. https://doi.org/10.1007/s40820-019-0357-y

[27] Zhu, H., Peng, H., Xu, G., Deng, L., Cheng, Y., & Song, A. (2021). Bilateral weighted regression ranking model with spatial-temporal correlation filter for visual tracking. IEEE Transactions on Multimedia, 24, 2098-2111. https://doi.org/10.1109/TMM.2021.3075876

[28] Reedha, R., Dericquebourg, E., Canals, R., & Hafiane, A. (2022). Transformer neural network for weed and crop classification of high-resolution UAV images. Remote sensing, 14(3), 592. https://doi.org/10.3390/rs14030592

[29] Chen, Z., Zhong, F., Luo, Q., Zhang, X., & Zheng, Y. (2022, November). Edgevit: Efficient visual modeling for edge computing. In International conference on wireless algorithms, systems, and applications (pp. 393-405). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19211-1_33

[30] Zou, J., & Zhang, Q. (2022). EyeSay: Brain visual dynamics decoding with deep learning & edge computing. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 2217-2224. https://doi.org/10.1109/TNSRE.2022.3193714

[31] Yamamura, H., Putri, E. U., Kawakami, T., Suzuki, A., Ariesyady, H. D., & Ishii, T. (2020). Dosage optimization of polyaluminum chloride by the application of convolutional neural network to the floc images captured in jar tests. Separation and Purification Technology, 237, 116467. https://doi.org/10.1016/j.seppur.2019.116467

[32] Han, X., Jia, M., Chang, Y., & Li, Y. (2022). An improved approach towards more robust deep learning models for chemical kinetics. Combustion and Flame, 238, 111934. https://doi.org/10.1016/j.combustflame.2021.111934

[33] Kim, S., Lu, P. Y., Mukherjee, S., Gilbert, M., Jing, L., Čeperić, V., & Soljačić, M. (2020). Integration of neural network-based symbolic regression in deep learning for scientific discovery. IEEE transactions on neural networks and learning systems, 32(9), 4166-4177. https://doi.org/10.1109/TNNLS.2020.3017010

[34] Chen, Y., Sundah, N. R., Ho, N. R., Natalia, A., Liu, Y., Miow, Q. H., ... & Shao, H. (2021). Collaborative Equilibrium Coupling of Catalytic DNA Nanostructures Enables Programmable Detection of SARS-CoV-2. Advanced Science, 8(18), 2101155. https://doi.org/10.1002/advs.202101155

[35] Zhang, W., Jiao, L., Li, Y., Huang, Z., & Wang, H. (2021). Laplacian feature pyramid network for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-14. https://doi.org/10.1109/TGRS.2021.3072488

[36] Preckel, F., Golle, J., Grabner, R., Jarvin, L., Kozbelt, A., Müllensiefen, D., ... & Worrell, F. C. (2020). Talent development in achievement domains: A psychological framework for within-and cross-domain research. Perspectives on Psychological Science, 15(3), 691-722. https://doi.org/10.1177/1745691619895030

[37] Shi, G., He, Y., & Zhang, C. (2020). Feature extraction and classification of cataluminescence images based on sparse coding convolutional neural networks. IEEE Transactions on Instrumentation and Measurement, 70, 1-11. https://doi.org/10.1109/TIM.2020.3023508

[38] Sulley, G. A., & Montemore, M. M. (2022). Recent progress towards a universal machine learning model for reaction energetics in heterogeneous catalysis. Current Opinion in Chemical Engineering, 36, 100821. https://doi.org/10.1016/j.coche.2022.100821

[39] Scalia, G., Grambow, C. A., Pernici, B., Li, Y. P., & Green, W. H. (2020). Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. Journal of chemical information and modeling, 60(6), 2697-2717. https://doi.org/10.1021/acs.jcim.9b00975

[40] Borrel, A., Huang, R., Sakamuru, S., Xia, M., Simeonov, A., Mansouri, K., ... & Kleinstreuer, N. C. (2020). High-throughput screening to predict chemical-assay interference. Scientific Reports, 10(1), 3986. https://doi.org/10.1038/s41598-020-60747-3