Research Article

Computer vision for efficient object detection and segmentation in molecular image analysis

Shaoxuan Yuan[1,2], Zhiwen Zhu[1], Jiayi Lu[1], Liangliang Cai[1,*], Qiang Sun[1]

[1]Materials Genome Institute, Shanghai University, Shanghai 200444, China.
[2]Faculty of Science, Katholieke Universiteit Leuven, Leuven 3001, Belgium.

[*]**Correspondence to:** Prof. Liangliang Cai, Materials Genome Institute, Shanghai University, Shanghai 200444, China. E-mail: cailiangliangmgi@shu.edu.cn

ORCID: Liangliang Cai (0000-0002-3112-815X)

How to cite this article: Yuan, S.; Zhu, Z.; Lu, J.; Cai, L.; Sun, Q. Computer vision for efficient object detection and segmentation in molecular image analysis. *J. Mater. Inf.* 2025;5:[Accepted]. https://dx.doi.org/10.20517/jmi.2025.78

Received: 11 September 2025 | **Revised:** 27 November 2025 | **Accepted:** 1 December 2025

Abstract

Image recognition, classification, and analysis of large sets of high-resolution molecular images are time-consuming and labor-intensive, even for human experts, owing to a lack of standardized approaches. In recent years, machine learning has emerged as a powerful tool for automating image data analysis in materials science. In this work, we developed a computer vision program for efficient object detection and instance segmentation, offering a fast alternative to manual molecular image analysis. By integrating YOLOv9 with an incremental learning strategy and hyperparameter optimization, the system enables accurate detection, classification, and segmentation of molecular species across diverse STM datasets. Our results demonstrate robust performance and minimal forgetting rates across multiple molecular categories, enabling scalable and updatable

surface image analysis workflows. We anticipate that computer vision methods will see increasing applications in the image data analysis within the field of on-surface chemistry.

Keywords: YOLOv9 object detection, instance segmentation, incremental learning, surface chemistry

INTRODUCTION

In recent years, the rise of artificial intelligence (AI) technologies and advances in materials science have brought about profound transformations in traditional materials design, synthesis and characterization[1-5]. Scanning tunneling microscopy (STM), a vital technique for investigating surface structures and adsorbates, has become widely employed in fields such as catalysis, molecular self-assembly, on-surface synthesis, and nanomaterials preparation[6-12], primarily due to its ability to provide atomic level resolution[13-18]. However, the high-resolution images obtained from STM measurements typically exhibit considerable complexity and contain rich microscopic information. Consequently, the analysis and interpretation of such images heavily depend on human experts' experience and subjective judgment. This reliance not only reduces research efficiency but also restricts the broader adoption and application of the cutting-edge technology.

To overcome the critical dependence on prior knowledges of human experts in high-resolution image analysis, artificial intelligence techniques have increasingly been introduced[19-28]. Recent studies have demonstrated that deep learning and computer vision algorithms can effectively automate the recognition and analysis of STM images, such as automatic detection and repair of imaging artifacts, classification of similar molecules, molecular counting, and automated length measurements of polymer chains. These studies highlight the significant potential of integrating AI and STM technology to improve analytical accuracy and reduce manual intervention.

Nevertheless, current researches remain largely limited to isolated applications tailored

to specific scenarios, typically focusing only on achieving a single functionality or particular target. Specifically, most existing studies employ small-scale, customized datasets for training and testing, lacking systematic data preparation and a general framework for predictive modeling and result analysis. Therefore, systematically integrating the entire AI pipeline, which includes data acquisition and preprocessing, model training, automated prediction and multi-scale data analysis, is one of the key challenges for achieving efficient, automated, and intelligent STM-based research.

In this work, we have developed a comprehensive framework that integrates different aspects of the machine learning workflow, including image data annotation, dataset creation, model training, inference, and data analysis. Based on the YOLOv9 model, a state-of-the-art approach that combines both object detection and instance segmentation capabilities, our program provides a one-stop solution for annotating image data, generating high-quality datasets, and training advanced models. Additionally, the program includes tools for data analysis based on prediction results, providing insights that can guide decision-making and further refine model performance. A key feature of the program is its support for incremental learning, allowing models to be updated and improved continuously as new data becomes available. This approach avoids for complete retraining from scratch and ensures that the models remain accurate and up to date.

MATERIALS AND METHODS`

Data preparation and annotation

To develop a robust model for molecular image analysis, we first curated a dataset of molecular images containing various molecular assemblies. The images were pre-processed to standardize resolution and contrast, ensuring uniformity across the dataset. We utilized an annotation tool to manually label molecular structures, providing bounding boxes and segmentation masks for training the object detection and instance segmentation models. Each molecular species was assigned a unique class ID, and the dataset was divided into training, validation, and test sets following an 80:10:10 split. All STM images used in this study were obtained from our group's experiments on molecular

self-assembly. To enhance dataset diversity, 2-3 original high-resolution STM images were augmented using operations such as rotation, flipping, cropping, and contrast adjustment, resulting in approximately 500 images in total. These augmentations preserve the intrinsic physical and structural characteristics of the molecular features while improving the robustness of model training. This dataset was then used for object detection labeling, where molecular positions were annotated using bounding boxes to form the complete training and validation sets.

Model selection and architecture

Our program is built on YOLOv9, the latest iteration in the YOLO (You Only Look Once) family, known for its efficient object detection and segmentation capabilities. Two model configurations were implemented: 1. Object detection which identifies molecular species using bounding boxes. 2. Instance segmentation which provides pixel-wise segmentation for precise molecular boundary identification.

Model training and evaluation

The YOLOv9 model was trained using a stochastic gradient descent (SGD) optimizer with momentum and weight decay applied for regularization. The model was trained for 100 epochs with an adaptive learning rate schedule. We used the following performance metrics for evaluation: F1 Score, which balances precision and recall for classification accuracy; mAP@0.95, which assessing detection accuracy across different confidence thresholds; forgetting rate, which measures the model's ability to retain knowledge of previously learned molecular categories. The incremental learning strategy was validated by training the model in multiple stages and measuring the degradation in performance across molecular categories.

The programming language is based on Python. We use PyTorch as the deep learning framework. GPU hardware is based on NVIDIA 3070Ti with 8GB VRAM. Training time is approximately 20 min per full training cycle.

RESULTS AND DISCUSSION

Our framework consists of four main components: data annotation, dataset generation, model training/prediction, and data processing. In the training section, users can choose

between two models: object detection or instance segmentation, and both of them are based on the YOLOv9 model. The overall framework architecture is illustrated in Figure 1.

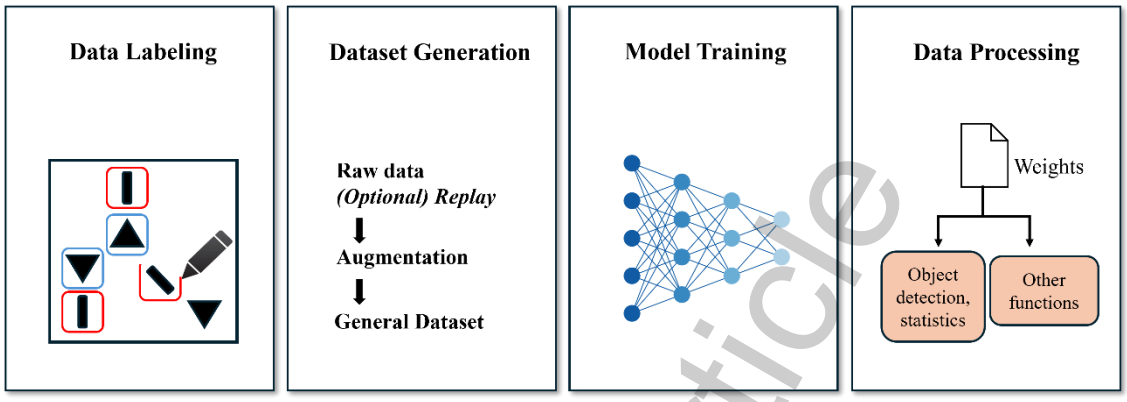

Figure 1. Workflow of the computer vision program for molecular image analysis.

The process begins with the "Data Labeling" stage, where the data is annotated to provide ground truth for model training. Following this, the system checks whether the labeled data corresponds to a new category. If a new category is detected, the system performs the "Load Class ID from Index Table" step to assign a unique identifier to the new category, which is then used in the "Dataset Generation" step. The generated dataset serves as the foundation for model training, ensuring that all relevant categories are included. Additionally, when generating the dataset, users can enable the incremental learning feature to create a more generalized dataset. In the incremental learning phase, the model is updated with newly added data without the complete retraining from scratch, significantly enhancing its adaptability, especially as new types of species and patterns are discovered from the experimental image data.

After dataset creation, the "Train & Detect" step utilizes the YOLOv9 model to perform both object detection and instance segmentation, identifying key features in the image data. YOLOv9, the latest version of the "You Only Look Once" family of models[29], combines advanced detection algorithms with high efficiency, enabling accurate and fast object detection even for complex datasets. It is improved upon its predecessors by

incorporating more robust backbone networks, enhanced anchor-free mechanisms, and dynamic label assignment strategies, which makes it particularly suitable for detecting small-scale molecular features in high-resolution images. Once the detection is completed, the results are processed by the "Data Processing" module which extracts valuable insights such as the identified classes and the average area of detected instances. These insights are crucial for understanding surface nanostructures, molecular self-assembly, and the distribution of reactants and products. The seamless integration of incremental learning ensures that the model evolves continuously as new data becomes available, maintaining accuracy of the model as experimental conditions change as well as allowing researchers to adapt their analysis to emerging surface phenomena. This combination of YOLOv9's advanced detection capabilities and incremental learning provides a powerful framework for analyzing molecular systems.

During the training process of this program, we found that hyperparameters have a significant impact on the model's performance. Improper hyperparameter configurations can drastically reduce training efficiency, potentially wasting several hours without yielding satisfactory results. In contrast, an appropriate combination of hyperparameters can accelerate convergence, improve model stability, and significantly enhance the overall performance. This effect becomes particularly evident in deep learning tasks where training is computationally expensive, as a poorly tuned learning rate or batch size can either cause divergence or lead to excessively slow convergence. Similarly, suboptimal values of augmentation or regularization parameters may result in underfitting or overfitting, making the training process inefficient and unstable. To address this challenge, it is crucial to adopt a systematic and automated hyperparameter optimization strategy that goes beyond manual trial-and-error. Given the large number of hyperparameters, manually adjusting each value is extremely time-consuming and practically impossible. Therefore, we employed a Bayesian optimization algorithm to search for the optimal combination of hyperparameters. Compared with conventional methods such as grid search and random search, Bayesian optimization can find near-

optimal solutions with fewer iterations, especially in high-dimensional hyperparameter spaces. It builds a surrogate model (e.g., Gaussian Process) to predict the behavior of the objective function and leverages an acquisition function (such as Expected Improvement, Probability of Improvement, or Upper Confidence Bound) to balance exploration of unknown regions and exploitation of promising areas, thereby avoiding inefficient or blind searches. Unlike grid search, which exhaustively evaluates every point in a predefined parameter grid, or random search, which samples configurations uniformly at random, Bayesian optimization intelligently selects the next point to evaluate based on prior knowledge from past observations. This makes it particularly efficient when each model evaluation is computationally expensive, as is the case in our training pipeline where a single training run can take several hours.

The surrogate model we utilized, Gaussian Process (GP), offers a probabilistic interpretation of the function being optimized. GP provides not only predictions of the mean performance for any hyperparameter configuration but also an estimate of uncertainty, which allows the acquisition function to determine whether to explore new regions or exploit areas that are already known to yield good results. The Expected Improvement (EI) criterion, in particular, evaluates the expected gain over the current best-known performance, guiding the search towards configurations that have the highest probability of improving the model metric (mAP@0.95 in our case). This dynamic balance between exploration and exploitation is the key to achieving both efficiency and accuracy in the hyperparameter search process. Furthermore, Bayesian optimization is well-suited for non-convex, noisy, and expensive objective functions, which are common in deep learning scenarios. The objective landscape of neural networks often contains many local minima and plateaus, making random or uniform search strategies highly inefficient. By contrast, Bayesian optimization builds a smooth approximation of the objective surface and can better navigate these complex landscapes. In our experiments, we observed that Bayesian optimization was able to identify optimal hyperparameter configurations within 30 iterations, while random search with the same number of trials failed to achieve a comparable performance. Therefore, we employed a Bayesian

optimization algorithm to search for the optimal combination of hyperparameters.

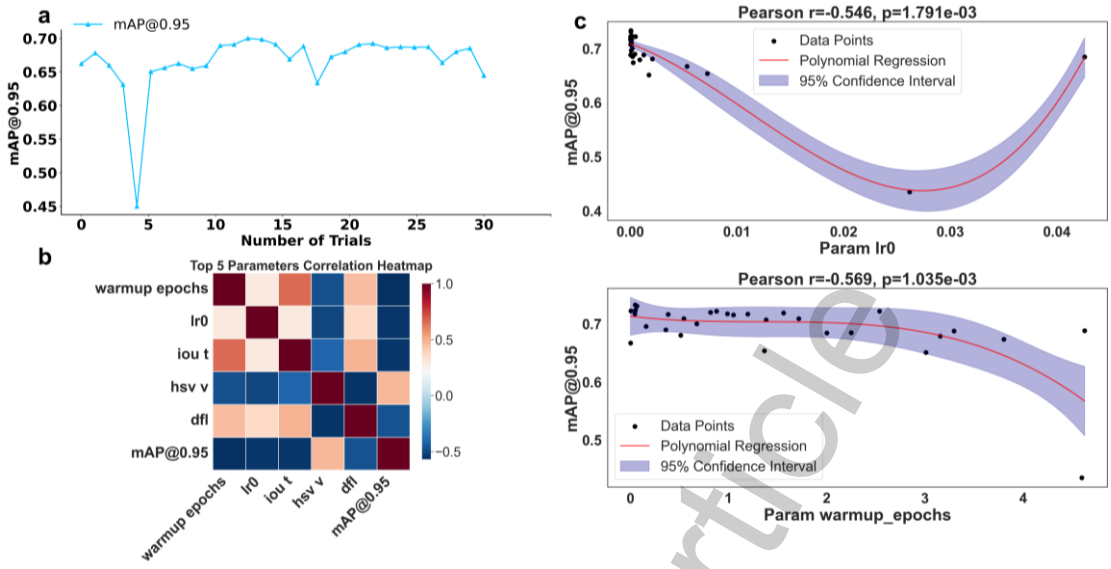

Figure 2. Hyperparameter optimization and its impact on model performance. (a) Model performance (mAP@0.95) over 30 iterations by Bayesian optimization, where each iteration represents a new set of hyperparameters; (b) The Pearson correlation heatmap showing the correlations among top 5 hyperparameters and mAP@0.95. Darker color indicates stronger correlations, providing valuable insights into which hyperparameter should be prioritized for further optimization; (c) The relationships between lr0 (top panel), warmup epochs (bottom panel), and mAP@0.95, with Pearson correlation coefficients and p-values displayed on top of the panels. These values help to identify which hyperparameters most strongly influence the model performance.

We constructed a search space from the selected hyperparameters and performed 30 iterations of Bayesian optimization. The selected hyperparameters[30] were chosen based on their relevance to model performance and training stability. From Figure 2a, we observed that the optimal configuration was achieved at the 13th iteration, where mAP@0.95 reached its maximum value. The heatmap in Figure 2b further illustrates the correlations among key hyperparameters, revealing that the initial learning rate (lr0) shows a strong negative correlation with performance, while warmup epochs exhibit a moderate positive correlation. These observations suggest that proper balance among learning rate, regularization, and warmup duration is crucial for optimizing training

stability and generalization. Detailed mathematical formulations and correlation analyses have been moved to the Supplementary Information (Section S4). In Figure 2c, two most influential hyperparameters were selected based on their Pearson correlation coefficients with the model performance metric. The scatter plots illustrate the relationships between each selected hyperparameter (e.g., initial learning rate Ir0, warmup epochs) and mAP@0.95. Each point represents a result from a training session, where the x-axis shows the hyperparameter value and the y-axis shows the corresponding mAP@0.95 score. In these plots, by fitting a polynomial regression curve, we can visually observe the trends between each hyperparameter and the model performance. For example, it is clear to see that the initial learning rate shows a significant negative correlation until lr0 ~ 0.03 with the model performance, meaning that increasing the learning rate leads to a decrease in mAP@0.95. This trend is clearly demonstrated in both the scatter plot and the regression curve, where the shape of the curve is closely related to the distribution of the data.

In our program, we have specifically incorporated the feature of incremental learning, and among the various incremental learning methods, we selected the most stable and convenient replay mechanism[31]. The principle of the replay mechanism is to save a portion of historical data during the model training process and periodically retrain the model with both the historical data and new data in subsequent training sessions. This approach allows the model to retain memory of the old data while learning new information, thus preventing it from forgetting previously learned knowledge.

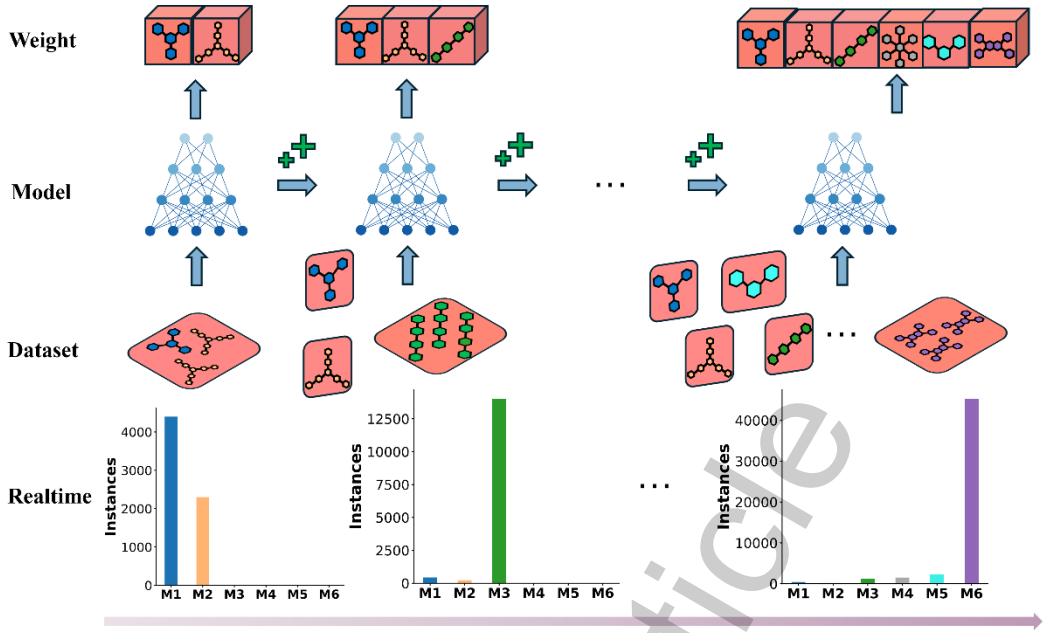

Figure 3. Illustration of the replay mechanism in incremental learning. The top row (Weight) represents the weight adjustments with more data being included. The second row (Model) illustrates the model training steps. The third row (Dataset) shows the dataset replay strategy, where historical data is incorporated into the training process. The bottom row (Realtime) presents the change in the number of instances from each category (M1 to M6) that participate in the training as a result of applying the replay strategy.

From Figure 3, we can see the replay mechanism selects a small batch of historical data and combines it with the current training data for the model training. This not only helps the model continuously adapt to new data but also ensures the stability of the model, as it "reviews" previous data during training. This helps to avoid overfitting to the current samples while maintaining learning from past knowledge. In this way, the model can gradually enhance its generalization ability and robustness during the incremental learning process, leading to more effective knowledge updates and applications.

The advantage of the replay mechanism is that it allows the model to be continuously updated and optimized as new data becomes available, without the need for complete retraining from scratch. This greatly improves training efficiency and the flexibility of data handling. Specifically for atomic-scale STM image analysis, the replay mechanism

offers distinct advantages over other incremental learning strategies such as knowledge distillation. STM images contain highly detailed pixel-level information - fine variations in local contrast, adsorption geometry, and tip-induced electronic effects that are physically meaningful and critical for molecular recognition. Distillation-based methods, which transfer feature distributions between teacher and student models, often compress or smooth out these subtle spatial and intensity variations, leading to a partial loss of nanoscale structural information. In contrast, replay revisits a small portion of real historical STM data during training, directly preserving the original pixel intensity distribution and spatial correlations. This approach enables the model to maintain sensitivity to atomic-scale morphological and electronic features while learning new molecular systems, effectively mitigating catastrophic forgetting. As a result, the replay strategy not only aligns with the physical nature of STM imaging but also achieves superior stability and accuracy in nanoscale feature recognition.

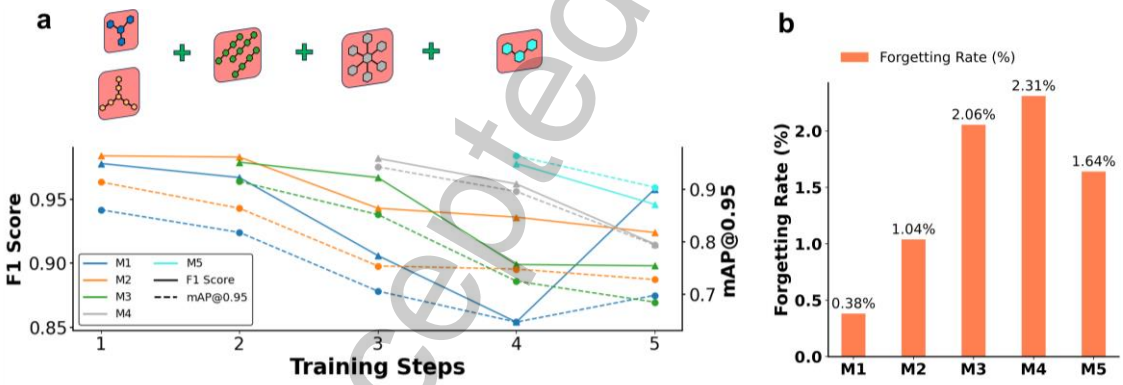

Figure 4. Model performance and forgetting rate across training steps. (a) The performances of the model over five training steps, represented by F1 scores and mAP@0.95 for different categories (M1 to M5). Each line corresponds to the F1 score or mAP@0.95 of a specific category; (b) The forgetting rates for each category (M1 to M5) as a percentage, with category M4 exhibiting the highest forgetting rate (2.31%) compared to the other categories.

As shown in Figures 4 and 5a, we selected five molecular categories (M1 to M5) to evaluate the performance of the model under the replay mechanism strategy. The model underwent five rounds of training after continuously updating with new categories of

molecular image data, during which we observed the changes in F1 scores and mAP@0.95 scores for each category as training progressed. F1 score and mAP@0.95 were chosen as the primary evaluation metrics due to their complementary nature: F1 score reflects the classification accuracy by balancing precision and recall, making it ideal for evaluating the model's ability to distinguish among molecular categories, while mAP@0.95 focuses on the spatial precision of bounding boxes, providing insights into the model's localization accuracy. By combining these metrics, we gained a comprehensive understanding of the model's performance in both classification and recognition tasks. Based on the changes in these metrics, we calculated the forgetting rate for each category to measure the model's ability to retain knowledge as new training data was introduced.

$$Forgetting\ Rate = \frac{P_{initial} - P_{final}}{P_{initial}}$$

Where $P_{initial}$ represents the model's performance on a specific category or task before incremental training (e.g., F1 score). P_{final} represents the model's performance on the same category or task after the training is completed.

From Figure 4b, we can clearly see that the forgetting rates for all categories have been reduced to very low levels, with all rates below 2.5%, demonstrating the effectiveness of the replay mechanism in mitigating catastrophic forgetting and maintaining the model performance across multiple categories. This demonstrates the impact of training on the model's ability to retain learned knowledge, with the forgetting rate indicating the loss of performance on previously learned categories.

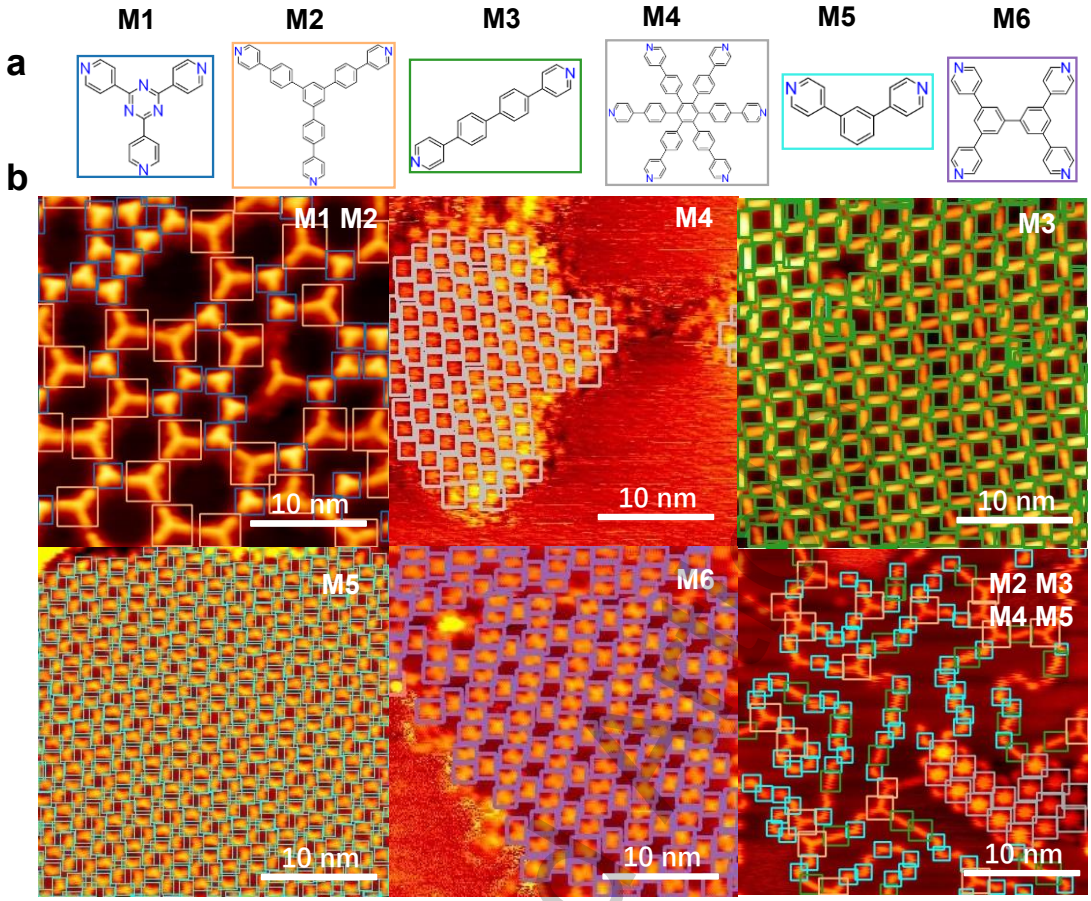

Figure 5. STM molecular images and prediction results. (a) Chemical structures of six molecular categories (M1 to M6); (b) STM images and the prediction results obtained using the well-trained models. The inferred molecules are highlighted by markers with corresponding colors. Blue squares denote molecule M1, orange squares denote molecule M2, green squares denote molecule M3, grey squares denote molecule M4, cyan squares denote molecule M5, purple squares denote molecule M6.

Through Figure 5, we can more intuitively access the effectiveness of incremental learning approach. This result is based on the model after five rounds of incremental learning, where the weight file integrates the learned representations of six molecular categories (M1 to M6). Figure 5a displays six representative molecular categories (M1-M6), each corresponding to a specific precursor or coordination unit used in STM studies of on-surface molecular self-assembly, i.e., M1: 2,4,6-tri(pyridin-4-yl)-1,3,5-triazine; M2: 4,4'-(5'-(4-(pyridin-4-yl)phenyl)-[1,1':3',1''-terphenyl]-4,4''-diyl)dipyridine; M3: 4,4'-di(pyridin-4-yl)-1,1'-biphenyl; M4: hexakis[4-(4'-pyridyl)phenyl]benzene; M5: 4,4'-(1,3-phenylene)bis-pyridine; and M6: 3,3',5,5'-tetra(pyridin-4-yl)-1,1'-biphenyl.

Using the trained weights, a single prediction was performed, and the results in Figure 5b show a clear and accurate identification of the corresponding molecular structures. The predicted patterns align well with the corresponding molecular structures in Figure 5a, demonstrating that the model effectively captures and retains the features of all six categories without significant degradation in performance. To further quantify this, the overall performance metrics indicate a precision of 0.959, recall of 0.931, mAP@0.5 of 0.976, and mAP@0.5:0.95 of 0.792, confirming the model's robustness across classes. Specifically, the per-class mAP@0.5:0.95 values for M1 through M6 are 0.698, 0.728, 0.685, 0.794, 0.904, and 0.945, respectively. These results highlight the model's strong generalization ability and its capability to maintain high accuracy and localization precision across diverse molecular categories, even after multiple stages of incremental updates.

This highlights the strength of the incremental learning strategy, as it enables the model to continuously learn new molecular categories while retaining knowledge of previously learned categories. The predictions indicate that the model can handle the complexity of molecular assemblies, accurately detect and classify molecules, thus illustrating its robustness and generalization capability after incremental learning. The clear patterns and accurate bounding boxes in the predictions reflect the model's ability to integrate diverse molecular features, making it a powerful tool for analyzing high-resolution molecular images of complex molecular arrangements.

Batch effects arising from different STM imaging environments, such as variations in tunneling current stability, tip geometry, or detector sensitivity, may introduce distributional bias in the image data. To minimize this effect, our dataset includes STM images collected under diverse experimental conditions on Au(111), Ag(111), and Cu(111) substrates. By incrementally introducing data from these different imaging environments while replaying earlier samples, the model continuously calibrates itself

across multiple acquisition domains, thereby reducing the risk of model bias. In addition to the experiments discussed above, we further evaluated the model's performance on images of the same molecular species acquired on three different substrates, namely Au(111), Ag(111), and Cu(111) [Figure 6]. The results show that the incremental learning strategy not only allows the model to continuously adapt to new data but also exhibits strong generalization across diverse imaging conditions. Despite variations in surface structures, noise, and contrast levels among different substrates, the model successfully identified the target molecule with no noticeable degradation in detection accuracy. This outcome indicates that the replay mechanism effectively preserves the essential features of the learned molecule, enabling robust recognition across heterogeneous datasets. The ability to adapt to different substrates highlights the strong transferability and generalization power of the incremental learning framework. In scanning tunneling microscopy (STM) imaging, the type of substrate plays a critical role in determining the overall image characteristics. For example, Au(111), Ag(111), and Cu(111) surfaces differ in their lattice constants and electronic structures, which leads to significant variations in background textures, contrast levels, and noise patterns in the resulting images. Conventional deep learning detection models are often sensitive to these differences; when trained exclusively on data from one substrate, their performance usually drops significantly when tested on another[32]. To overcome this issue, we adopted the aforementioned incremental learning strategy combined with a replay mechanism.

In our experiments, we progressively introduced STM images of the target molecule obtained on Au(111), Ag(111), and Cu(111). At the initial stage, the model was trained solely on Au(111) data and achieved high detection accuracy. We then incrementally incorporated images from Ag(111) and Cu(111) substrates while simultaneously replaying Au(111) images to reinforce the retention of its previously acquired knowledge. This strategy ensured that the model maintained strong recognition performance for Au(111) molecules while adapting to the unique imaging characteristics of Ag(111) and Cu(111). The final evaluation results demonstrated that the model maintained

consistently high accuracy across all three substrates (with only minor variations in mAP@0.5), thereby validating the effectiveness of the incremental learning approach.

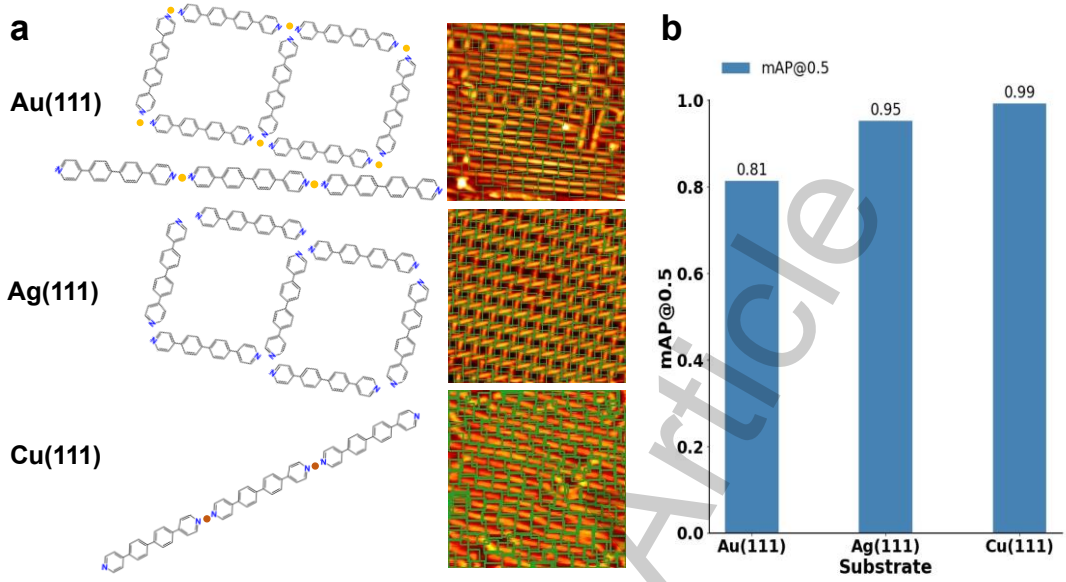

Figure 6. Performance of the incremental learning framework on the sample molecule on multiple metal substrates. (a) Visualization of molecular detection results for 4,4'-di(pyridin-4-yl)-1,1'-biphenyl on STM images acquired from Au(111), Ag(111), and Cu(111) surfaces, respectively. All STM images span 20 nm × 20 nm. Bounding boxes highlight the successfully identified molecules. In the model, yellow and orange represent gold (Au) and copper (Cu) atoms, respectively; (b) Quantitative performance comparison across substrates using mAP@0.5, showing the model's ability to maintain high detection accuracy (0.81, 0.95, and 0.99) under varying imaging conditions.

As shown in Figure 6, the model demonstrates robust molecular recognition even when the substrate type changes, which typically introduces differences in background texture, surface electronic structures, and imaging noise. Despite these challenges, the incremental learning approach enables the model to retain the discriminative features of the molecule while adapting to variations caused by different substrates. This robustness is largely attributed to the fact that our incremental training does not simply overwrite previously learned weights with new data, but instead uses a balanced training approach where data from both old and new substrates are sampled in each training epoch, with

old datasets comprising 10% of the training data. This allows the model to maintain an equilibrium between previously learned features and new information, resulting in stable recognition performance under diverse imaging conditions.

Our experiments confirmed that even when Cu(111) images presented higher noise levels and lower contrast, the model was still able to accurately detect and classify the molecule based on its essential structural features. Furthermore, variations in substrate properties not only affect image noise and contrast but can also influence the self-assembly behavior of molecules on the surface. For instance, 4,4'-di(pyridin-4-yl)-1,1'-biphenyl molecules may exhibit slightly different arrangements on Au(111), Ag(111), and Cu(111) due to differences in adsorption energy and intermolecular interactions, which affect the molecular packing density and orientation[33,34]. By incorporating data from multiple substrates, the incremental learning model learns to recognize key molecular structures and features regardless of such variations.

Lastly, we note an additional advantage of the incremental learning framework is its improved training efficiency. Traditional approaches would require training a separate model for each substrate, which is computationally costly and time-consuming. In contrast, our incremental approach dynamically updates the same model parameters, achieving “train-once, adapt-to-many” capability. The replay mechanism also reinforces the retention of key features learned in earlier stages, ensuring that the model does not forget older datasets while adapting to new conditions. This not only reduces the computational overhead but also streamlines the entire training pipeline. These results emphasize the transferability of the learned representations and confirm that the model can accurately detect and classify the molecule across heterogeneous environments without significant performance loss.

CONCLUSIONS

In this study, we developed a comprehensive YOLOv9-based program for object detection and instance segmentation of molecular STM images, especially addressing the

challenges posed by manual molecular image analysis in surface nanostructures. The program integrates several advanced features, including data labeling, dataset creation, model training, detection, and analysis, with a strong focus on supporting incremental learning. The inclusion of the replay mechanism for incremental learning has been proved highly effective in mitigating catastrophic forgetting, allowing the model to retain previously learned information while adapted to new data. By replaying a small portion of the old dataset, the system reduced the rate of forgetting while maintaining efficiency. Additionally, Bayesian optimization was employed to fine-tune hyperparameters, further enhancing the model's performance. The results demonstrate that our approach provides a scalable, efficient, and robust solution for the analysis of high-resolution molecular STM images, enabling researchers to accurately interpret complex surface reactions exceeding the limitations of manual analysis. Overall, the integration of machine learning techniques, including incremental learning and hyperparameter optimization, offers promising improvements in the accuracy, consistency, and scalability of molecular image analysis, paving the way for further innovations in the field of surface chemistry.

DECLARATIONS

Authors' contributions

Research conception and study design, data analysis & interpretation, and software framework development & machine-learning model implementation: Shaoxuan Yuan, Zhiwen Zhu

Data acquisition, annotation pipeline development, and result visualization: Jiayi Lu

Manuscript drafting, critical revision, and important intellectual content review: Shaoxuan Yuan, Liangliang Cai

Administrative, technical, and material support; project supervision; and funding acquisition: Qiang Sun

Availability of data and materials

The code and trained models supporting the findings of this study have been made publicly available in a GitHub repository: https://github.com/yuanke75/yolo_program_for_spm. All datasets generated and analyzed during the current study are included within the repository as processed files and annotation scripts.

Financial support and sponsorship

This work was financially supported by the National Natural Science Foundation of China (No. 22302120).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Lo, S.; G. Baird, S.; Schrier, J.; Blaiszik, B.; Carson, N.; Foster, I.; Aguilar-Granda, A.; V. Kalinin, S.; Maruyama, B.; Politi, M.; et al. Review of low-cost self-driving laboratories in chemistry and materials science: the "frugal twin" concept. *Digit. Discov.* **2024**, *3*, 842-868 doi:10.1039/D3DD00223C.

2. Wang, Z.; Qin, M.; Zhang, P.; Xu, Y.; Que, S.; Yan, F.; Xiang, X.D. High throughput characterization method of electrical and phonon properties by dielectric resonant spectroscopy. *MGEA* **2025**, doi:10.1002/mgea.70010.

3. Xiang, J.; Li, Y.; Zhang, X.; He, Y.; Sun, Q. Local large language model-assisted literature mining for on-surface reactions. *MGEA* **2025**, *3*, e88 doi:10.1002/mgea.88.

4. Li, Y.; Liu, X.; Wang, X.; Xie, W.; Qiu, D.; Yang, J. Effect of material properties on the thermal responses of the carbonization and pyrolysis layers of polymer matrix composites for charring-ablators. *J. Mater. Inform.* **2025**, *5*, 31, doi:10.20517/jmi.2024.104.

5. Wang, W.Y.; Zhang, S.; Li, G.; Lu, J.; Ren, Y.; Wang, X.; Gao, X.; Su, Y.; Song, H.;

Li, J. Artificial intelligence enabled smart design and manufacturing of advanced materials: The endless Frontier in AI+ era. *MGEA* **2024**, *2*, e56 doi:10.1002/mgea.56.

6. Jung, T.A.; Schlittler, R.R.; Gimzewski, J.K. Conformational identification of individual adsorbed molecules with the STM. *Nature* **1997**, *386*, 696-698, doi:10.1038/386696a0.

7. Meng, T.; Lu, Y.; Lei, P.; Li, S.; Deng, K.; Xiao, X.; Ogino, K.; Zeng, Q. Self-Assembly of Triphenylamine Macrocycles and Co-assembly with Guest Molecules at the Liquid-Solid Interface Studied by STM: Influence of Different Side Chains on Host-Guest Interaction. *Langmuir* **2022**, *38*, 3568-3574, doi:10.1021/acs.langmuir.2c00188.

8. Wyrick, J.; Wang, X.; Namboodiri, P.; Kashid, R.V.; Fei, F.; Fox, J.; Silver, R. Enhanced Atomic Precision Fabrication by Adsorption of Phosphine into Engineered Dangling Bonds on H-Si Using STM and DFT. *ACS Nano* **2022**, *16*, 19114-19123, doi:10.1021/acsnano.2c08162.

9. Wang, L.; Xia, Y.; Ho, W. Atomic-scale quantum sensing based on the ultrafast coherence of an H2 molecule in an STM cavity. *Science* **2022**, *376*, 401-405 doi:10.1126/science.abn9220.

10. Moreno, D.; Parreiras, S.O.; Urgel, J.I.; Muniz-Cano, B.; Martin-Fuentes, C.; Lauwaet, K.; Valvidares, M.; Valbuena, M.A.; Gallego, J.M.; Martinez, J.I.; et al. Engineering Periodic Dinuclear Lanthanide-Directed Networks Featuring Tunable Energy Level Alignment and Magnetic Anisotropy by Metal Exchange. *Small* **2022**, *18*, e2107073, doi:10.1002/smll.202107073.

11. Lyu, C.-K.; Gao, Y.-F.; Gao, Z.-A.; Mo, S.-Y.; Hua, M.-Q.; Li, E.; Fu, S.-Q.; Chen, J.-Y.; Liu, P.-N.; Huang, L.; Lin, N. Synthesis of Single-Layer Two-Dimensional Metal–Organic Frameworks M3(HAT)2 (M=Ni, Fe, Co, HAT=1,4,5,8,9,12-hexaazatriphenylene) Using an On-Surface Reaction. *Angew. Chem.* **2022**, *134*, e202204528 doi:10.1002/ange.202204528.

12. Liu, J.; Li, J.; Xu, Z.; Zhou, X.; Xue, Q.; Wu, T.; Zhong, M.; Li, R.; Sun, R.; Shen, Z.; et al. On-surface preparation of coordinated lanthanide-transition-metal clusters. *Nat. Commun.* **2021**, *12*, 1619, doi:10.1038/s41467-021-21911-z.

13. Di Giovannantonio, M.; Fasel, R. On-surface synthesis and atomic scale characterization of unprotected indenofluorene polymers. *J. Polym. Sci.* **2022**, *60*, 1814-1826, doi:10.1002/pol.20210902.

14. Wang, J.; Niu, K.; Xu, C.; Zhu, H.; Ding, H.; Han, D.; Zheng, Y.; Xi, J.; You, S.; Deng, C.; et al. Influence of Molecular Configurations on the Desulfonylation Reactions on Metal Surfaces. *J. Am. Chem. Soc.* **2022**, *144*, 21596-21605, doi:10.1021/jacs.2c08736.

15. Kinikar, A.; Di Giovannantonio, M.; Urgel, J.I.; Eimre, K.; Qiu, Z.; Gu, Y.; Jin, E.; Narita, A.; Wang, X.-Y.; Müllen, K.; et al. On-surface polyarylene synthesis by cycloaromatization of isopropyl substituents. *Nat. Synth.* **2022**, *1*, 289-296 doi:10.1038/s44160-022-00032-5.

16. Liu, L.; Klaasen, H.; Witteler, M.C.; Schulze Lammers, B.; Timmer, A.; Kong, H.; Monig, H.; Gao, H.Y.; Neugebauer, J.; Fuchs, H.; Studer, A. Polymerization of silanes through dehydrogenative Si-Si bond formation on metal surfaces. *Nat. Chem.* **2021**, *13*, 350-357, doi:10.1038/s41557-021-00651-z.

17. Mallada, B.; de la Torre, B.; Mendieta-Moreno, J.I.; Nachtigallova, D.; Matej, A.; Matousek, M.; Mutombo, P.; Brabec, J.; Veis, L.; Cadart, T.; et al. On-Surface Strain-Driven Synthesis of Nonalternant Non-Benzenoid Aromatic Compounds Containing Four- to Eight-Membered Rings. *J. Am. Chem. Soc.* **2021**, *143*, 14694-14702, doi:10.1021/jacs.1c06168.

18. Zhu, X.; Liu, Y.; Pu, W.; Liu, F.-Z.; Xue, Z.; Sun, Z.; Yan, K.; Yu, P. On-Surface Synthesis of C144 Hexagonal Coronoid with Zigzag Edges. *ACS Nano* **2022**, *16*, 10600-10607 doi:10.1021/acsnano.2c02163.

19. Hellerstedt, J.; Cahlík, A.; Švec, M.; Stetsovych, O.; Hennen, T. Counting molecules: Python based scheme for automated enumeration and categorization of molecules in scanning tunneling microscopy images. *Softw. Impacts* **2022**, *12*, 100301, doi:10.1016/j.simpa.2022.100301.

20. Krull, A.; Hirsch, P.; Rother, C.; Schiffrin, A.; Krull, C. Artificial-intelligence-driven scanning probe microscopy. *Commun. Phys.* **2020**, *3, 1*, doi:10.1038/s42005-020-0317-3.

21. Milošević, D.; Vodanović, M.; Galić, I.; Subašić, M. Automated estimation of chronological age from panoramic dental X-ray images using deep learning. *Expert Syst. Appl.* **2022**, *189*, 116038, doi:10.1016/j.eswa.2021.116038.

22. Li, J.; Telychko, M.; Yin, J.; Zhu, Y.; Li, G.; Song, S.; Yang, H.; Li, J.; Wu, J.; Lu, J.; Wang, X. Machine Vision Automated Chiral Molecule Detection and Classification in Molecular Imaging. *J. Am. Chem. Soc.* **2021**, *143*, 10177-10188 doi:10.1021/jacs.1c03091.

23. Gordon, O.M.; Hodgkinson, J.E.A.; Farley, S.M.; Hunsicker, E.L.; Moriarty, P.J. Automated Searching and Identification of Self-Organized Nanostructures. *Nano Lett.* **2020**, *20*, 7688-7693 doi:10.1021/acs.nanolett.0c03213.

24. Kang, J.; Yoo, Y.J.; Park, J.-H.; Ko, J.H.; Kim, S.; Stanciu, S.G.; Stenmark, H.A.; Lee, J.; Mahmud, A.A.; Jeon, H.-G.; Song, Y.M. Deepgt: Deep Learning-Based Quantification of Nanosized Bioparticles in Bright-Field Micrographs of Gires-Tournois Biosensor. *Nano Today* **2023**, 52, 101968 doi:10.2139/ssrn.4428599.

25. Faraz, K.; Grenier, T.; Ducottet, C.; Epicier, T. Deep learning detection of nanoparticles and multiple object tracking of their dynamic evolution during in situ ETEM studies. *Sci. Rep.* **2022**, *12*, 2484 doi:10.1038/s41598-022-06308-2.

26. Newby, J.M.; Schaefer, A.M.; Lee, P.T.; Forest, M.G.; Lai, S.K. Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 9026-9031 doi:10.1073/pnas.1804420115.

27. Zhu, Z.; Yuan, S.; Yang, Q.; Jiang, H.; Zheng, F.; Lu, J.; Sun, Q. Autonomous Scanning Tunneling Microscopy Imaging via Deep Learning. *J. Am. Chem. Soc.* **2024**, *146*, 29199-29206, doi:10.1021/jacs.4c11674.

28. Seifert, T.J.; Stritzke, M.; Kasten, P.; Moller, B.; Fingscheidt, T.; Etzkorn, M.; de Wolff, T.; Schlickum, U. Chirality Detection in Scanning Tunneling Microscopy Data Using Artificial Intelligence. *Small Methods* **2024**, *8*, e2400549, doi:10.1002/smtd.202400549.

29. Wang, Chien-Yao, I-Hau Yeh, and Hong-Yuan Mark Liao. "Yolov9: Learning what you want to learn using programmable gradient information." *European conference on*

computer vision. Cham: Springer Nature Switzerland, 2024, doi:10.48550/arXiv.2402.13616.

30. Gao, T.; Gao, J.; Zhang, J.; Song, B.; Zhang, L. Development of an accurate "composition-process-properties" dataset for SLMed Al-Si-(Mg) alloys and its application in alloy design. *J. Mater. Inf.* **2023**, *3*, 6. dx.doi.org/10.20517/jmi.2023.03

31. Zhou, Da-Wei, et al. Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024), doi:10.48550/arXiv.2302.03648.

32. Pan, S. J. & Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359 (2010). doi.org/10.1109/TKDE.2009.191

33. Otero, R., Gallego, J.M., de Parga, A.L.V., Martín, N. and Miranda, R. (2011), Molecular Self-Assembly at Solid Surfaces. Adv. Mater., 23: 5148-5176. doi.org/10.1002/adma.201102022

34. Voigt, J., Baljozović, M., Martin, K. *et al.* An aperiodic chiral tiling by topological molecular self-assembly. *Nat Commun* **16**, 83 (2025). doi.org/10.1038/s41467-024-55405-5