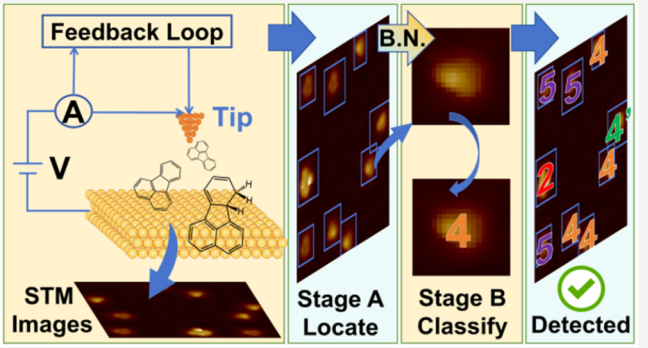
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Two-Stage Machine Learning Framework for Accurate Discrimination of Isomers and Very-Similar Molecules on Surfaces

Zixuan Wei, Qigang Zhong,* Jinbo Pan, Fang Han Lim, Lifeng Chi,* and Shixuan Du*

Cite This: J. Am. Chem. Soc. 2025, 147, 35232−35243
Read Online

ABSTRACT: The accurate detection and discrimination of on-surface organic isomers and very similar molecules are crucial for monitoring chemical reaction processes and analyzing various reaction mechanisms and molecular properties. Despite its importance, nano- and surface science communities still lack an efficient, robust, precise, and automated detection approach for on-surface isomers and highly similar molecules. Here, we present ReSTOLO, a convolution neural network (CNN)-based framework for precise detection and identification of multiple types of sparsely distributed molecules on surfaces, particularly designed for scanning tunneling microscopy (STM) images containing numerous molecules with analogous features. To address challenges arising from molecular shape and size similarities, we implemented a two-stage framework comprising two CNN models: YOLO v5.m was used for molecular localization, and ResNet-101 for classification. The framework optimally harnesses the advantages of both models by applying a box normalization connection. We demonstrated the framework's effectiveness by applying it to analyze a surface reaction process involving six molecules with nearly identical STM signatures. The training process employed an STM image database of single molecules augmented with physical and experimental tools constructed using standardized image boxes. This two-stage approach achieved approximately ∼20% improvements in performance metrics, including precision, recall, and accuracy, compared to conventional frameworks. The framework exhibits robust capabilities in automatically and efficiently pinpointing and discriminating between molecular species with similar configurations in complex surface reactions. This automated molecular discriminator represents a significant advance in facilitating STM tip-manipulated chemical reactions on surfaces.



INTRODUCTION

On-surface organic reactions and low-dimensional characterizations have attracted significant attention in recent years, driven by their unique properties, including diverse functional groups, low-dimensionality, confinement effects, carbon chain plasticity, and electronic delocalization effects in aromatic molecules.[1] Scanning probe microscopy (SPM) has enabled the observation and manipulation of chemical processes on metal surfaces under high vacuum and well-defined interfaces, complementing traditional solution and three-dimensional systems. These on-surface methodologies show huge potential in synthesizing novel low-dimensional nanostructures unattainable via conventional methods,[2,3] revealing reaction mechanisms,[2−8] discovering novel physical properties,[5−7,9−15] and developing advanced technical applications.[11,16−18] Across these fields, precise classification of specific entities, such as chiral molecules, isomers, or very-similar molecules, has become increasingly crucial.

Imaging techniques have emerged as the cornerstone of on-surface nanoscientific discoveries and industrial applications.[9,19] SPM techniques, particularly scanning tunnel microscopy (STM) and atomic force microscopy (AFM),[10−12] have proven powerful and indispensable for molecular observation and manipulation.[2,3,5−8,13−18,20−26] Their atomic-level resolution capabilities have positioned them at the central role of nanoscience research, enabling both direct observation[27] and precise manipulation[2,12,25,28] of molecular entities. The analysis of SPM images often requires accurate determination of the positions, structures, and categories of molecules, many of which might share similar characteristics, posing a fundamental challenge to the community. While traditional visual analysis by naked eye is time and labor-intensive, some processes, including molecular detection and enumeration, could be partially streamlined with proper choices of algorithms and data mining methods,[29,30] thanks to the recent advances in computer vision technology.

Received: March 3, 2025
Revised: September 8, 2025
Accepted: September 9, 2025
Published: September 18, 2025

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

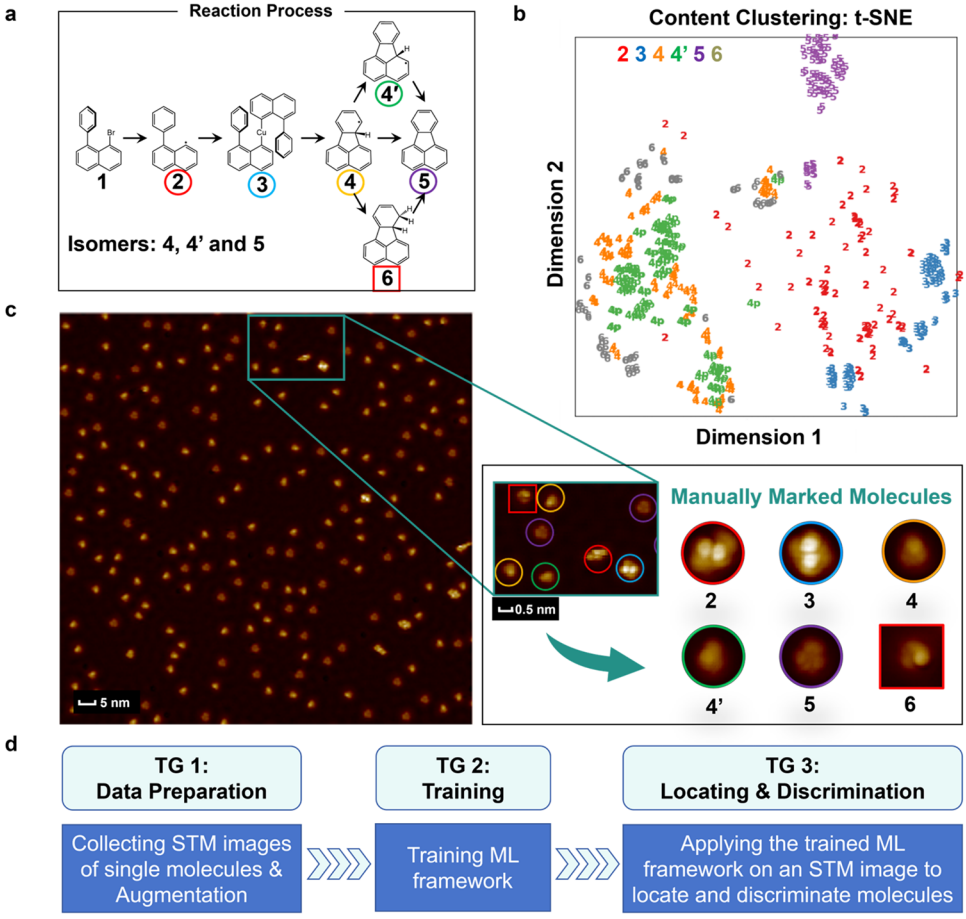


Figure 1. Experimental characterization and identification of on-surface dehydrogenative C−C coupling reactions. (a) Reaction scheme showing intermediate and final products on Cu(111). (b) Content clustering analysis of single-molecule images through t-SNE. (c) STM image of molecular reaction at 297 K; scanning parameters: 100 mV, 10 pA, with CO tips. The right panel displays the zoomed-in image of the rectangular region. The STM images used in this article are adapted with permission. Reproduced from ref 51 Copyright [2024] American Chemical Society. (d) Deployment workflow of molecular detection based on the STM image obtained from experiments. The deployment workflow includes three task groups (TGs): data preparation, model training, and detection (locating and classification).

In the past decade, machine learning (ML), particularly deep learning-based computer vision, methods have blossomed with significant application to material sciences.[31−37] Among them, the convolution neural network (CNN) has emerged as a transformative deep learning method, demonstrating significant capabilities in many important real-life tasks, including image classification, localization, segmentation, and reconstruction.[38−45] Within the surface and nanoscience domain, CNNs are widely applied to surface atom and defect detection,[32,38] phase formation analysis,[39] experimental image denoising,[40] ferroelectric domain detection,[41] and atomic characterization.[33] Notably, the flexibility of the shape and size of the bounding boxes have been the key factor to the robustness and versatility of these models.[42] While direct interpretation of on-surface molecular structures with deep learning approaches has been reported in the past few years,[43] previous studies mainly focused on limited molecular systems: single-molecule types with different chiralities,[44] or with defined orientations,[45] and two or more distinct molecular structures.[46] These systems exhibit readily distinguishable features, which can be identified through dimensionality reduction methods like t-SNE or PCA feature maps.[44,47,48] To enhance the performance of models, researchers have either incorporated spatial information (like specific patterns from assembled or periodic structures)[44,45,47,48] or integrated domain knowledge such as molecular structural information into Bayesian belief networks and Markov-based denoising methods.[49,50] However, precise localization and classification of similar molecules randomly distributed on surfaces remain significant challenges, especially when STM images of distinct molecules appear to be "very similar", such as when the differences between molecules involve only the quantity or position of hydrogen atoms.

Traditional object detectors struggle in scenarios where multiple classes share highly similar shapes/textures, precisely the situation encountered with structural isomers on surfaces. Single-stage object detection models like YOLO, which simultaneously locate and identify targets using adaptive bounding boxes in one pass, particularly excel when working with data sets containing objects with diverse or distinct visual features. However, this approach becomes problematic when different molecule types produce detection boxes of similar size and shape; YOLO might retain multiple high-confidence boxes for visually similar species, leading to misclassification errors. On the other hand, two-stage detectors like Faster R-CNNs contain two stages: region proposal (RP) and classification and regression (C&R) connected by the region of interest pooling (ROI-P), which locate and identify targets separately. Faster R-

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

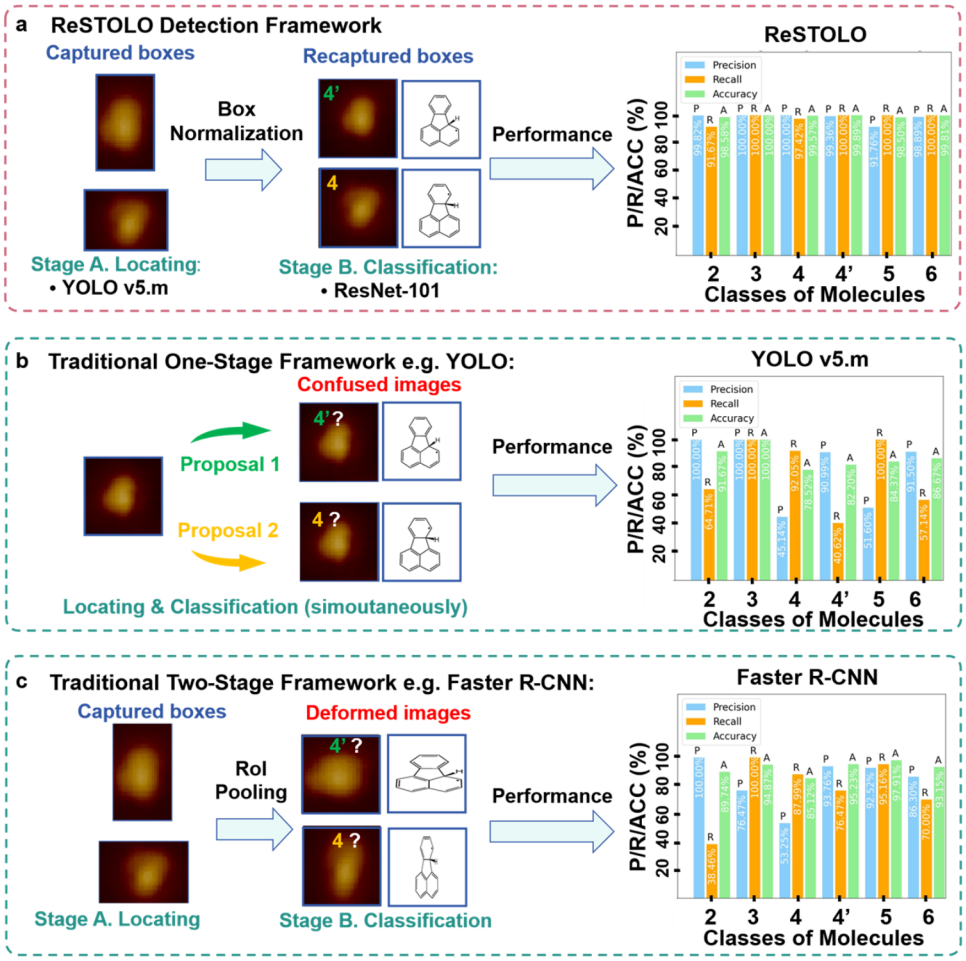


Figure 2. Performance comparison among our proposed ReSTOLO detection framework and traditional one- and two-stage frameworks: YOLO and Faster R-CNN. (a) Our proposed ReSTOLO detection framework. The operation mechanisms and performance of (b) YOLO and (c) Faster R-CNN. The single-molecule images in Stage A of (a) and (c) represent the located-out single molecule surrounded by detection boxes (bounding/anchor box). Single-molecule images in Stage B of (a) and (c) are size- and shape-unified through box normalization and recapture (a) and RoI pooling (c), respectively; the latter method may lose details. YOLO locates and classifies molecules simultaneously. For very similar molecules, two proposal boxes of similar molecule types with shapes alike could both have high scores and hence be kept, which possibly leads to misclassification. On the right side of the panel, three critical detection indexes are considered to evaluate the performance, including precision, recall, and accuracy.

CNNs first propose regions of interest before classifying features, following the ROI-P, which normalizes box dimensions to facilitate the classification. However, the second stage (C&R) after ROI-P faces limitations dealing with highly similar molecular structures, especially without obvious exploitable spatial relationships. The pooling process ROI-P can compromise, average out, or drop subtle shape details that are essential for distinguishing isomers.

Hereby, we present a CNN-based two-stage deep-learning STM detection framework named ReSTOLO, designed specifically for localizing and classifying sparsely distributed isomers and very similar molecules in multispecies STM overview images. The framework integrates two CNN models with a box normalization connection: YOLO version 5.m for initial localization (stage A) and ResNet-101 for subsequent classification (stage B). In stage A, YOLO v5.m employs adaptive bounding boxes in various sizes and shapes to robustly locate molecules, followed by standardization to fixed-size boxes for consistent image capture. Then, in stage B, ResNet-101 is utilized for precise molecular classification of these standardized images. The workflow of ReSTOLO is formulated into three task groups (TG): preparation, training, and detection. In the preparation task group, a single-molecule type STM image data set is constructed by incorporating physically and experimentally based augmentation methods, including SRResNet resolution enhancement, etc. Next, we optimized both YOLO v5.m and ResNet-101 independently in the training task group. Finally, ReSTOLO is applied to discriminating molecules with similar STM signatures. To showcase the effectiveness of our model, we applied ReSTOLO upon a complex surface reaction process on the Cu(111) surface, which involved six molecular species, four among them with very similar STM signatures; the framework achieved precise molecular localization and classification, demonstrating approximately 20−25% improvement in precision and recall compared to traditional structures, for example, Faster R-CNN and YOLO v5.m.

RESULTS AND DISCUSSION

Surface Reaction Process with Several Similar Molecules. Our investigation focuses on a surface reaction process involving six distinct organic aromatic molecules,

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

representing the intermediate and final states (designated as 2, 3, 4, 4′, 5, and 6) of 1-bromo-8-phenylnaphthalene (1) reaction on a Cu(111) surface, as shown in Figure 1a. The reaction begins with breaking the C−Br bond in state 1, generating an intermediate state 2 with an unpaired electron. Subsequently, two 2 molecules combine to form a copper-mediated dimer. Then, the reaction progresses by breaking the C−Cu bonds, yielding products 4, 4′, 5, and 6. Among these products, 4, 4′, and 5 are isomers, distinguished only by the location of the hydrogen atom, while 6 contains two extra H atoms compared to 4, 4′, and 5. Notably, molecule 2 exhibits structural features similar to 4, 4′, 5, and 6 except for the missing C−C bond. This dehydrogenative on-surface C−C coupling reaction serves as an ideal case study for developing automated discrimination of similar intermediate and final structures, crucial for elucidating the underlying physical and chemical mechanisms.[51]

The experimental procedure involved evaporating the precursor molecules on a Cu(111) substrate, followed by annealing at 297 K. STM measurements were performed, and all STM images acquired are shown in Figure 1c. Since debromination happens spontaneously upon precursor adsorption on the substrate, precursor 1 was excluded from subsequent research. The STM images reveal numerous bright spots with similar features (Figure 1c, left panel). Through noncontact AFM (nc-AFM) analysis, we identified several characteristic bright spots, with their corresponding STM images presented in Figure 1c, left panel. The similarity in STM signatures among intermediate and final states makes visual analysis by the naked eye impractical, even with extensive experimental expertise. The traditional, widely used dimension reduction and clustering algorithm t-SNE, frequently used in pattern discovery and data mining,[52] are implemented to analyze the features of our single-molecule images in Figure 1b. Molecules cannot be divided into distinctly distinguishable independent groups on the 2D feature plan, particularly for **4**, **4′**, and **6**. The dispersive distribution indicates the high similarity of molecules **4**, **4′**, and **6**, corresponding to the STM images in Figure 1c, in which molecules **4**, **4′**, and **6** exhibit three protrusions, with one brighter than the others. This limitation motivated our development of a deep learning framework specifically designed to differentiate these molecular species solely on the basis of their STM signatures. We proposed a deep-learning framework, ReSTOLO, containing a two-stage structure that locates and classifies molecules separately and sequentially (see Figure 2a). The deployment workflow of the ReSTOLO detection framework is shown in Figure 1d.

ReSTOLO Detection Framework and Deployment Workflow. Traditional object detection paradigms such as YOLO architectures or Faster R-CNN are optimized for real-world scenarios characterized by (i) variable observer distances, (ii) intraclass heterogeneity across dimensions, and (iii) multimodal discriminative features. These frameworks employ adaptive bounding boxes to accommodate such variability, which is essential for applications such as autonomous navigation systems. However, detection of surface-bound sparsely distributed molecules presents a distinct set of constraints: (i) planar geometries with uniform z-axis positioning, (ii) shape/structure homogeneity within molecular species, and (iii) shape/size-dependent classification parameters. Hence, these fundamental differences from macroscopic detection scenarios could cause conventional flexible-boundary approaches to be detrimental to accuracy. Within this molecular context, geometric transformation of a detected molecule A whether scaling or compression might induce misidentification with its isomeric counterpart B, since affine and scaling invariance do not apply as in macroscopic object detection. Meanwhile, conventional crystallographic analysis methods are also not applicable due to the stochastic, nonperiodic distribution of surface molecules. These unique constraints point to the necessity of a specialized detection framework optimized for molecular imaging applications.

As shown in Figure 2b, a one-stage framework like YOLO locates and classifies targets simultaneously. YOLO generates many boxes of different shapes with type label and their probability (bounding/anchor box) in the same time and afterward rules out unlikely boxes using the non-maximum suppression (NMS) algorithm. For highly similar targets, the shape variability of bounding boxes could seriously interfere the classification. More concretely, two or more bounding boxes of different molecular types with similar shapes could both be kept due to the intrinsic similarity between different molecular types. However, for two-stage frameworks like Faster R-CNN in Figure 2c, a Region of Interest (RoI) pooling is used to unify the shape and size of bounding boxes. However, the details and shape nuances could be erased and dropped by the RoI pooling process. These shortages are evident, especially in cases of high similarity (Figure 2b,c). Hence, applying both frameworks to our molecular system would lead to a significant decrease in precision and accuracy index for certain molecule types (Figure 2b,c right).

Our framework ReSTOLO, inspired by the proposal-classification strategy of the deep vision model Faster R-CNN and R-FCN,[53] integrates YOLO v5.m and ResNet-101 in a two-stage detection pipeline, as presented in Figure 2a. This design leverages YOLO v5.m's structural lightness and variable bounding box capabilities for initial molecular localization while mitigating classification ambiguity through ResNet-101's fixed-size image analysis (224×224 pixels^2). The YOLO v5.m and ResNet-101 are integrated through an intermediate box normalization algorithm that recalibrates detected regions to standardized dimensions (150×150 pixels^2) and centers molecules using a weighted pixel square sum average method. This design successfully avoided the loss of image shape information and texture details and reduced the similarity problem, resulting in good detection effectiveness. The deployment workflow of the ReSTOLO detection framework includes data preparation (TG 1), training (TG 2), and locating and classification (TG 3), as shown in Figure 1d. It is worth noting that the improvements from our model design are clearly observable when comparing different frameworks trained in parallel on identical original data sets (see Figure 2).

TG 1. Data Preparation. In data preparation, each single-molecule image of various types was captured using 300×300 pixel^2 boxes at 200% magnification (that is, magnifying the original STM image with a ratio of 200% to display the fine detail) from experimentally collected STM images. Incorporation of images at various resolutions aims to enhance model robustness by encouraging learning of resolution-independent molecular features. This data acquisition and cleaning process is rather key to our deployment.

A data set comprising approximately ∼40 molecules per molecular species, extracted from both high- and low-resolution STM images exhibiting varying background intensities, is obtained after the STM experiments and data

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

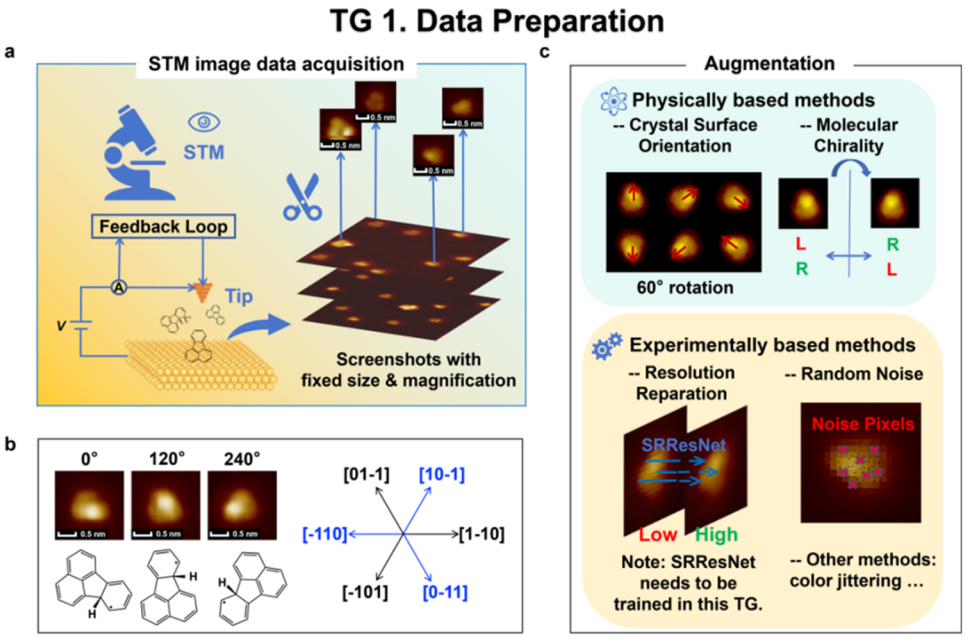


Figure 3. Data preprocessing and feature extraction for molecular STM images. (a) STM experiments and data acquisition. Single-molecule images were extracted by screenshotting with a fixed size and magnification. (b) STM images of the single molecule (screenshots) with three equivalent orientations on the Cu(111) surface (blue or black), with an additional three angular bisector orientations used for data augmentation. (c) Two main kinds of augmentation methods: physically based methods and experimentally based methods. The physically based methods mainly consider the inherent physical and chemical properties of our surface system. Single-molecule images were rotated according to the three equivalent directions of the Cu(111) surface and their angular bisectors. Mirror reflections were applied to consider the existence of chirality. Experimentally based methods indicate the difference in experimental details. Low-resolution images were repaired by the super resolution (SR) algorithm to compensate for the lack of high-resolution images. SRResNet is trained using high-resolution single-molecule images from the acquired data, with training and verification data sets randomly divided with a 0.1 verification ratio. Random white noise and other random pixel change methods were applied to account for potential changes in the imaging condition and environment (see Supporting Information 5.1).

cleaning process. Then, the data set is augmented by incorporating material physical properties and experimental perspective, including instrumental and methodological limitations, for a more domain-specific enhancement. Previous studies have demonstrated the effectiveness of data augmentation strategies when applying deep machine learning to surface systems with limited training data.[44,47] In addition, multiple studies across various machine learning applications have also validated this approach for small data sets.[54,55] The detailed process is shown in Figure 3a–c. Considering the physical and experimental background of our augmentation strategies, their theoretical sights are distinct. The details of augmentation methods were introduced as follows.

The physically based augmentations include 60 ° rotation and mirror reflection. The Cu(111) surface exhibits three equivalent crystal orientations, imposing corresponding molecular alignment constraints (Figure 3b). We also performed a principal component analysis (PCA) to help further discern the surface crystal orientations (see Supporting Information 2). Based on these physical constraints, we implemented rotational augmentation along these three primary directions and their angular bisectors, rather than arbitrary angular rotations, to optimize computational cost while maintaining physical relevance (top panel, Figure 3c). Also, given that **4**, **4′**, **5**, and **6** are the intermediate compounds without complex stereochemical transformation, we treated chiral variants as equivalent. This equivalence was implemented through horizontal and vertical mirror transformations of molecular screenshots, effectively doubling the available training data for chiral species, while maintaining physical accuracy. The rationale behind these physically based augmentations can also be interpreted in a more mathematical perspective; the classification of molecules can be conceptualized as a geometrical transform-sensitive function from molecular image (independent variable) to molecular type (dependent variable) that maintains invariance under the aforementioned specific symmetry operations. These symmetry operations should not influence the identification of the molecular type.

The experimentally based augmentations concern the variations and limitations of the experimental conditions. Our STM data set included both low- and high-resolution images (bottom panel, Figure 3c), reflecting practical experimental constraints including measurement limitations and computational resource optimization. Low-resolution images, while efficient to acquire, inherently contain less identifiable features, potentially compromising neural network effectiveness.[46] Inspired by the generative residual networks previously reported for simulation-to-realistic nanostructure image conversion, we adopted the SRResNet architecture to compensate for this limitation.[56] After compiling the initial training data set, several high-resolution images in the data set were selected to train the SRResNet reparation model, leveraging the interpixel relationship of molecule screenshots. The training and verification loss of SRResNet showed no sign of overfitting (see Supporting Information 3.2, Figure S17). Notably, such an implementation required careful consideration of true pixel resolution, as both original STM images and magnified screenshots often contain redundant color information across adjacent pixels. To determine optimal sampling parameters and true image resolution, we employed

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

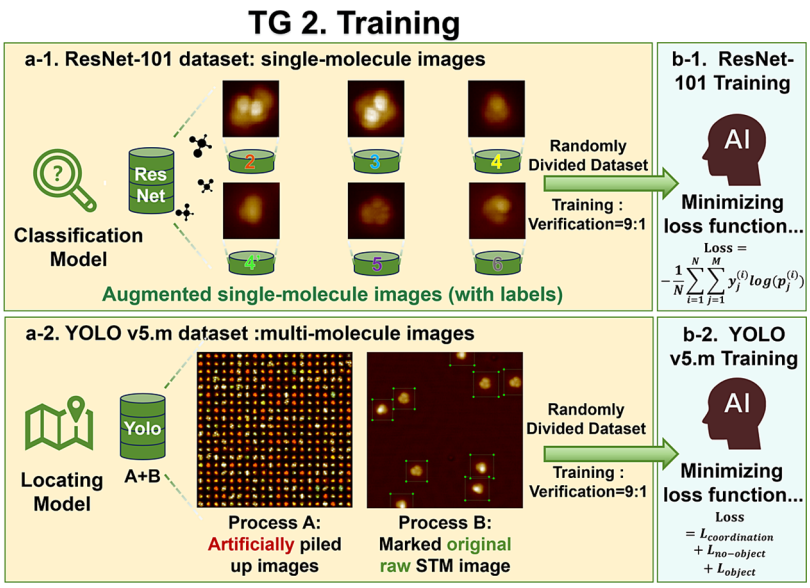


Figure 4. Training pipeline for the ReSTOLO framework (TG 2). The training pipeline can be divided into two branches according to the model type: YOLO v5.m and ResNet-101. (a)-1 ResNet-101 model is trained directly using the data set from TG 1, using augmented single-molecule images with type labels. (a)-2 YOLO v5.m requires additional data set preprocessing. As a target detection model, YOLO v5.m needs to be trained with multimolecule images. Here, we implied two different methods: A (piling up augmented single-molecule screenshot images) and B (marking molecules with a plug-in program LABELIMG in the original raw multimolecule STM images). (b) Training of two models. Models are trained by minimizing the loss functions. Training and verification data sets were divided randomly with a ratio of 9:1 for both models. Note that SRResNet also needs to be trained, which was included in TG 1.

fast Fourier transformation (FFT)[57] to quantitatively assess the spatial frequency content (detailed SRResNet training protocols are provided in Supporting Information 3). Furthermore, random white noise was introduced to simulate thermal-dynamic fluctuations and system perturbations of the experimental instruments.

We have also considered other widely applied augmentation methods, which may not have a very straightforward background but could be helpful to the enhancement of model performance. The jittering of brightness, contrast, and hue of molecular images was applied. This might partially reflect different STM imaging conditions in real experiments, although some of such jittered values could be "non-physical" (see Supporting Information 5.1). These augmentation methods are also included in the experimentally based augmentations.

TG 2. Training. The training task group involved optimizing YOLO version 5 for molecular localization and ResNet-101 for subsequent classification. While potentially less precise than more complex architectures, YOLO v5.m's lightweight design proved ideal for initial molecular detection. On the other hand, ResNet-101, chosen for its sophisticated classification capabilities enabled by residual connections, effectively mitigated network degradation during training.

The data set used in training the locating model contains STM images with multiple molecules to simulate realistic STM imaging conditions. We developed a Python-based subworkflow to generate composite images by combining 15 × 15 pixel^2 screenshots at different densities (Figure 4a-2, Process A) and simultaneously generating corresponding label files, which contain precise molecular locations. Also, we supplemented the synthetic data set with manually annotated raw STM images using LABELIMG software (Figure 4a-2, Process B), to enhance performance and account for substrate features such as surface structures. Since geometric precision was less critical for the localization stage, only square-format boxes were used to optimize the detection accuracy.

The classification training pipeline directly utilized the augmented single-molecule data set from TG 1, requiring no additional preprocessing (Figure 4a-1). The training is realized by minimizing the loss function, as in other supervised learning tasks. During the training (TG 2), there is no training priority requirement for these two models since their training does not depend on each other. That is, they can be trained simultaneously (Figure 4b). Comprehensive training protocols, loss functions, hyperparameter configurations, and convergence metrics are detailed in Supporting Information 6.2 and 7.2. Our 10-fold cross validation of both models further confirms the absence of overfitting in our framework (see Supporting Information 6.3 and 7.4). No specific hyperparameters or training protocols were particularly critical for converging the YOLO and ResNet models. Complete training protocols and optimized hyperparameter configurations are detailed in Supporting Information 4, 5.1, 6.1−6.2, 7.1−7.2, and 9.1.

TG 3. Detection. The final task group focused on the framework validation using previously unseen STM images. The complete testing subworkflow architecture is illustrated in Figure 5a(1−5). Following initial localization, molecular centers were adjusted using a weighted average method (box normalization), where pixel weights were determined by the square sum of the RGB channel values. Bounding boxes were subsequently standardized to 150 × 150 pixels^2 (corresponding to the 300 × 300 pixels^2 training data set at 200% magnification) to maintain geometric consistency crucial for accurate classification. This standardization step proved essential, as the classification stage is highly sensitive to molecular shape and size, where even the modest geometric distortions could lead to significant misclassification. The

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

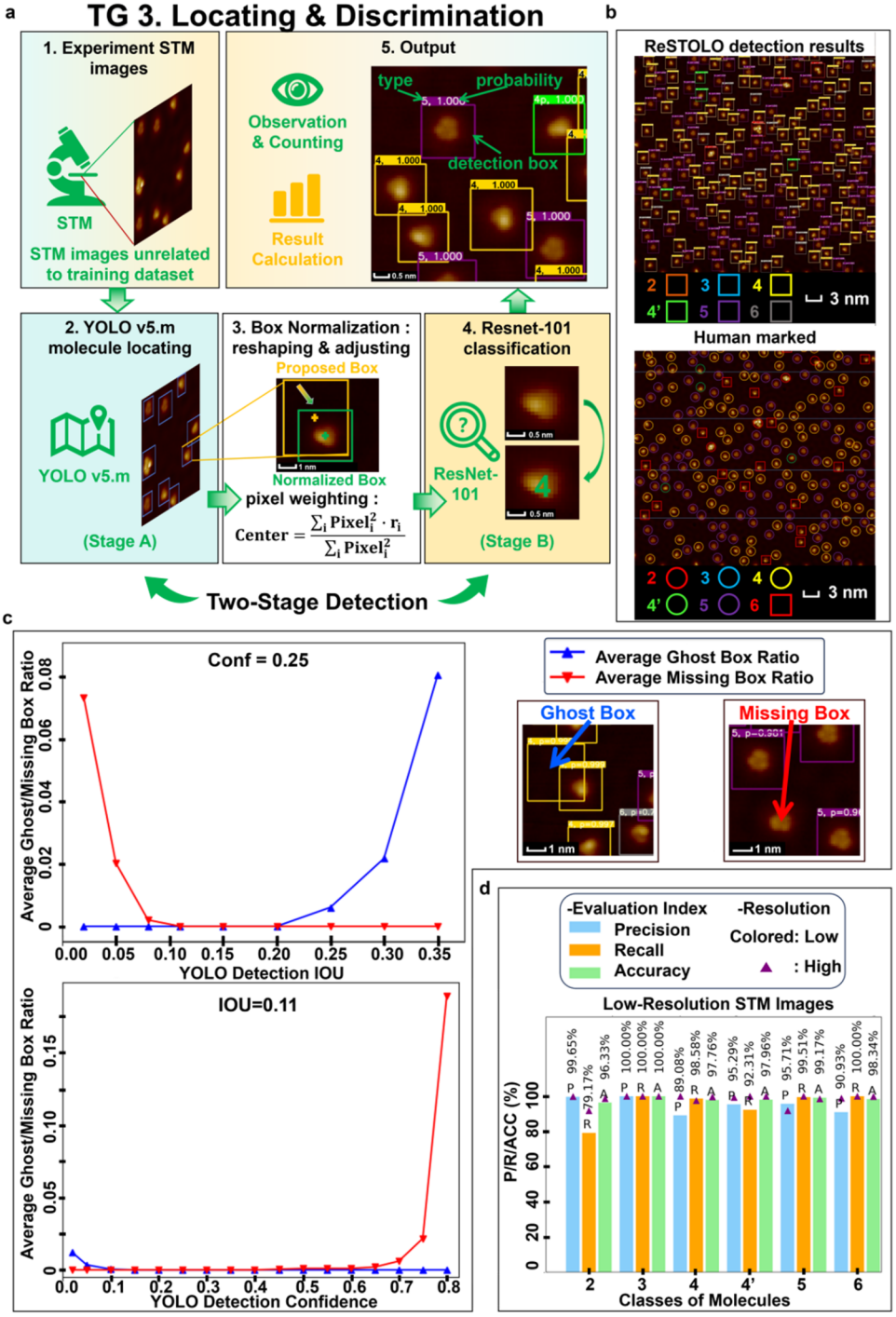


Figure 5. Workflow and performance of the ReSTOLO framework for molecular detection (TG 3). (a) Testing TG implements a sequential two-stage approach: Stage A for molecule localization and Stage B for feature classification. Between stages, bounding box geometries are standardized using a pixel-weighted optimization algorithm called box normalization. (b) AI detection (up) vs naked-eye manual annotation (down) using a high-resolution image (YOLO Confidence = 0.25, IoU = 0.11). Molecular type classification scheme: Type **2**: red (square/round), Type **3**: blue (square/round, not present), Type **4**: yellow (square/round), Type **4′**: green (square/round), Type **5**: purple (square/round), and Type **6**: gray square/red square. (c) Localization error analysis in Stage A via IoU and confidence metric. Subplots illustrate examples of false positives (ghost boxes, left) and false negatives (missing boxes, right). This hyperparameter sweep also helps to determine the best YOLO threshold hyperparameter selections. (d) Classification performance metrics in Stage B upon low-resolution (23 × 23 pixels^2, colored columns) and high-resolution (45 × 45 pixels^2, purple triangular) STM images.

principles and implementation details of the box normalization are described in Supporting Information 10.1.

To minimize experimental artifacts, we implemented Otsu background subtraction[58] to reduce substrate- and condition-dependent effects. Given that STM scanning regions are adjustable during experiments, we implemented edge-detection criteria to exclude partially captured molecules. Specifically, molecular detection with centers close to image boundaries was automatically excluded. Furthermore, we applied physical constraints by rejecting detection boxes with areas outside the

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

range of two-thirds to three-halves of the standard 150 × 150 pixels^2 area, as such deviations would be physically implausible. Notably, while our framework was optimized for a consistent magnification ratio, it maintains adaptability to varying magnification levels through simple scalar adjustment of the standardized box dimensions, implemented via multiplication by the relative magnification ratio; see Supporting Information 10.1. We also conducted sensitivity analysis on ResNet-101's performance relative to the shape and size variation of the normalized box; details are presented in Supporting Information 10.2 and 10.3, Figure S23.

For each detected molecule, the ResNet-101 model automatically calculates the probability distribution across all possible molecular types (e.g., P(4|5), the probability of molecule **5** being classified as type **4**, expressed in Bayesian notation). These probability values are generated by the six output channels in the model's final fully connected layer (FCC), indicating the uncertainty of the model's discrimination between different molecule types. More details are provided in Supporting Information 7.4.

Detection Details and Hyperparameters. Our ReSTOLO framework is validated by employing independent test sets comprising 12 low-resolution images (L) and five high-resolution images (H) with detailed data set specifications provided in Supporting Information 4.1. Figure 5b presents a qualitative comparison between AI-based detection and human naked-eye annotation using a high-resolution STM image. The framework demonstrated exceptional performance in molecular localization and classification, achieving near-perfect correspondence with human expert annotations. Additional detection results across multiple STM images are documented in Supporting Information 8.

The true positives, true negatives, false positives, and false negatives were manually verified and normalized to evaluate the precision, recall, and accuracy of our framework. Given our two-stage architecture, traditional single-stage evaluation metrics such as average precision/recall (AP/AR) indices are not applicable here. Instead, we implemented stage-specific evaluation protocols: YOLO v5.m's localization performance was assessed through missing/ghost-box robustness analysis (Stage A), while ResNet-101's classification performance was evaluated using normalized molecular-type-specific precision, recall, and accuracy metrics (Stage B). Due to the significant imbalances between population of molecules, the number of molecules was normalized for metric calculation (detailed in Supporting Information 1.1). These evaluations, displayed in Figure 5c,d, provide a comprehensive framework for performance assessment.

The localization stage demonstrated high sensitivity to Intersection over Union (IoU) variations compared to confidence thresholds. Optimal performance was achieved with IoU = 0.11 and confidence = 0.25, resulting in near-perfect localization of the molecules. The threshold values were determined through a systematic parameter sweep. The optimal intervals are shown in Figure 5c. In the classification stage, the precision, recall, and accuracy all exceeded 90% across the very similar molecules **4**, **4′**, **5**, and **6**, and molecule **3** achieved perfect classification in several metrics. High-resolution STM images consistently outperformed low-resolution counterparts by 3−5% across metrics. The lowest performance was observed for molecule **2** in low-resolution images (79.2%, precision), which could be attributed to the ambiguity of molecule **2** with molecule **4** and its polymorphological relatives underrepresented in the training set. The framework inherently extracts and learns contour information about the different molecule types, as observed from the final convolutional layer of the ResNet-101 feature map (see Supporting Information 7.3, Figure S18).

We tested the performance of our framework over a differently divided training and testing data set setting, which gives evidence to the reliability of the model's performance (see revised Supporting Information 4.2, Figure S8).

Note that the tip geometry and electronic structure could have critical influences on the appearance of STM images. To mitigate the tip effects on molecular recognition, all of the STM images in this work were acquired with a CO tip. The ReSTOLO users are required to maintain a well-defined tip condition throughout imaging to ensure consistent STM image quality at a level perceptible to the human eye.

Detection Effectiveness. The pipeline of skillfully concatenating two models, namely, YOLO and ResNet-101, is intuitively straightforward, as we attempt to fully exploit the advantages of the two models as much as possible. In other words, to make them suit their corresponding subtasks: locating and classification, respectively. To benchmark our framework's effectiveness, we compared two widely adopted architectures: (i) stand-alone YOLO v5.m and (ii) Faster R-CNN, with our ReSTOLO framework. To ensure fair comparison, we maintained consistent training data preparation methodologies across all frameworks, utilizing the protocols established for ReSTOLO in TG 2. Results presented in Figure 2 demonstrate significant performance differentials. YOLO v5.m exhibited substantial degradation in precision for molecules 4, 5, and 6 and in recall for molecules 2, 4′, and 6. While YOLO v5.m showed improved performance on high-resolution data sets compared to low-resolution ones, its overall performance remained markedly inferior to ReSTOLO. Similarly, Faster R-CNN demonstrated notably lower precision for molecules 4, 5, and 6 and diminished recall for molecules 2, 4, 4′, and 6. For Faster R-CNN evaluation, input images were segmented into 640 × 640 pixel^2 regions to accommodate model constraints. Supplementary analyses, including ResNet-101 box shape and size robustness studies (Supporting Information 9.3, Figure S23), support our initial hypothesis regarding the limitations of single-stage frameworks in this application domain. Benchmark YOLO v5 and Faster R-CNN were optimized to their fullest potential as much as possible (Supporting Information 9.2 for details). ReSTOLO, benchmark YOLO v5, and Faster R-CNN were all trained in parallel by using identical molecular image data.

Our two-stage design is not a simple direct connection of the locating and classification model. Instead, a box normalization is creatively integrated between the two stages. The key innovation of our framework is the design of a two-stage architecture with a box normalization method. The contribution of the framework itself can be estimated by parallel comparisons between different frameworks (see Figure 2 and Supporting Information 5.3, Figure S11).

Contribution of Data Augmentation. In order to achieve convergence, we applied various augmentation strategies to our two-stage framework. A detailed quantitative analysis of different augmentation strategies can be found in Supporting Information 5.2, Figures S9 and S10. The results indicate that random color permutation and random white noise serve as critical baseline components for effective detection, enabling successful classification of type 5 molecules

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

and substantial portions of other molecular species. Furthermore, physical augmentation (rotation, reflection) and the contrast/hue/lightness modulations proved essential for achieving the highly accurate discrimination between types 4, 4′, and 6. All of the augmentation methods contributed to model convergence and performance of the model across both low- and high-resolution images. The convergence of the two-stage architecture fundamentally requires the strategic application of data augmentation.

When implementing limited augmentation strategies, we observed significant performance disparities between low- and high-resolution images; results for high-resolution images are consistently better. As we progressively incorporated more augmentation methods, the performance gap narrows substantially, though high-resolution images are maintained slightly better. A detailed quantitative analysis of the effects of different augmentations across low- and high- resolution images is presented in Supporting Information 5.2 (Figures S9 and S10).

General Applicability of ReSTOLO. In order to display the universality and generalizability of our framework ReSTOLO and the independence of the substrate of choice and specific molecular interactions, we additionally investigated the performance of ReSTOLO on two other systems: acenaphthylene relatives (AN) on Au(111) and the hexaphenyl-substituted hexabenzocoronene (HBC-Ph) on the Au(111) surface. Our framework displayed good performance in both systems. Four types of molecules were detected with high accuracy and recall for ANs. Target molecules HBC-Ph were detected with high confidence from other impurities (see Supporting Information 12). By training with the STM data of other systems under specific conditions, one can easily apply ReSTOLO to discriminate molecules.

Robustness of Box Normalization. Due to the fact that molecules typically appear brighter than the substrate in STM imaging, the pixel-weighted square sum method of box normalization demonstrates strong resilience against imaging challenges, including contrast, noise, resolution limitations, and nonphysical distortions, as long as molecules remain distinguishable from the substrate background (see Supporting Information 10.4).

Influence of Super-Resolution Reparation. SRResNet plays solely the role of augmentation and participates in neither the verification of ReSTOLO nor the testing after training. SRResNet itself does not appear to be overfitting during its training process (Supporting Information 3.2). Besides, the verification loss of ReSTOLO decreases monotonically (Supporting Information 6.2 and 7.2). Furthermore, the whole detection performance increases after applying the SR reparation augmentation (Supporting Information 5.2), and the potential introduction of nonphysical characteristics by SR does not appear to influence the detection.

Alternatives of Models and Strategies. As a two-stage framework, the performance of our ReSTOLO framework depends on the effectiveness of both YOLO v5.m (stage A) and ResNet-101 (stage B), as illustrated in Figures 2 and 5. Given YOLO has already demonstrated robust performance (see Figure 5 and Supporting Information 6.3, Figure S15), classification accuracy in stage B becomes the determining factor. Theoretically, ResNet-101 could be replaced with any high-performing model. Here, we focused our comparative analysis on four different classification models: VGG-16[59] (∼13.8 million parameters), ResNet-101[60] (∼44.5 million parameters), Vision Transformer ViT B/16[61] (∼86 million parameters, the smallest standard ViT size), and Efficient Net B6[62] (∼43 million parameters, selected to approximate the complexity of ResNet-101). As shown in Figures S33 and S34, ViT B/16 and Efficient Net B6 achieve good performance comparable to ResNet-101, while VGG-16 demonstrates limited performance and insufficient performance to be used for our classification requirements (Figure S31). These results suggest that ResNet-101, Vision Transformer ViT B/16, and Efficient Net B6 can be interchangeably used for our two-stage framework; see Figures S32−S34. We conclude that replacing ResNet-101 in the current implementation is unnecessary, as all of the three models demonstrate similar effectiveness. Apart from performance, the deployment requirements for ResNet-101, ViT, and Efficient-Net are also similar, while the training of ResNet-101 is relatively faster (see Supporting Information 12) for details. However, a newer architecture like ViT and Efficient Net might offer advantages for other surface systems, and users are encouraged to experiment with them. Details can be found in Supporting Information 12.

Our implemented YOLO v5 has consistently demonstrated near-perfect performance in localizing isolated molecules (Figure 5 and Supporting Information 6.3, Figure S15), suggesting limited potential improvement from newer detection models. Nonetheless, users are still encouraged to use the recent model design. For challenging cases where molecules exhibit extreme proximity, newer models like YOLO v8 might be considered to further enhance detection capabilities.

The prohibitive resource requirements for constructing universal STM image databases or libraries, given the vast chemical space of surface molecular species and experiment-specific configurations, necessitate adopting a specified machine learning data set for targeted molecular recognition tasks.

For framework design and training strategies, our direct training strategy involved a box normalization operation. In addition to our current training strategy, we acknowledge that an end-to-end iterative training strategy could also potentially optimize box proposals to maximize the final classification accuracy. However, the possible iterative training approach involves substantial trade-offs: considerably more training time and more computational resources. More importantly, the interdependent model interactions in iterative training introduce optimization complexities similar to those observed in the generative adversarial network (GAN), potentially trapping the system in a local minimum problem during training.[63,64] Nonetheless, the effect and complexity of this method will be investigated in the future.

CONCLUSIONS

The present work has demonstrated an efficient two-stage surface molecular detection framework, integrating YOLO v5.m and ResNet-101 architectures with comprehensive data augmentation strategies. The framework's key innovation lies in its intermediate box normalization mechanism, which effectively addresses molecular similarity challenges by decoupling localization and classification tasks. This architectural decision enables each model to focus on its respective strengths: robust detection for YOLO v5.m and precise classification for ResNet-101. Our physics-informed data augmentation pipeline incorporates surface crystal symmetry, molecular chirality, and realistic STM experimental conditions.

https://doi.org/10.1021/jacs.5c03730
J. Am. Chem. Soc. 2025, 147, 35232−35243

Multiple extra tests proved the robustness, reliability, and the universality of ReSTOLO. This framework achieved successful discrimination among 'very-similar' isomers (**4**, **4′**, **5**, and **6**) with high precision and accuracy. Comparative analyses against established Faster R-CNN and YOLO v5.m validated our framework's superior performance. The framework's robustness across varying resolution scales and experimental conditions, coupled with adaptive box size adjustment, effectively overcomes the limitations inherent to single-stage approaches.

EXPERIMENTAL SECTION

YOLO (You Only Look Once). YOLO is a state-of-the-art, real-time object detection system that has gained considerable attention in the field of computer vision.[65,66] Unlike traditional object detection methods that employ a two-step process involving region proposal and classification, YOLO performs both tasks simultaneously by using the anchor box and non-maximum suppression (NMS). This unified approach allows YOLO to achieve a high-speed performance. The latest versions of YOLO have demonstrated remarkable efficiency in detecting objects in real-world scenarios, making them suitable for applications like video surveillance, autonomous driving, manufacturing industry, and augmented reality.[45,66−69] More details can be found in Supporting Information 5.

ResNet-101. ResNet-101, short for Residual Network with 101 layers, is a deep convolutional neural network designed to overcome the vanishing gradient problem that plagues very deep networks.[60,70] By introducing residual learning, ResNet-101 allows the training of extremely deep networks without a significant drop in performance. The network's architecture includes skip connections, or shortcuts, that enable the gradient to be directly back-propagated to earlier layers, thus facilitating easier optimization. ResNet-101 has been widely adopted in various image recognition tasks and has set new standards for accuracy in the ImageNet challenge. More details can be found in Supporting Information 6.

Super-Resolution. Super Resolution (SR) is a technique that aims to enhance the resolution of images beyond the capabilities of the original capturing device. SR methods can produce high-resolution images from low-resolution sources, which are particularly useful in scenarios where high-resolution imagery is desired but not feasible to capture. Deep learning-based SR algorithms, such as those using CNNs, have shown remarkable results in restoring fine details and textures in images. Applications of SR include medical imaging, satellite imagery, and video upscaling, where the enhanced clarity and detail can be critical for accurate analysis and interpretation. Two well-known SR models are SRGAN and SRResNet.[71] More details can be found in Supporting Information 3.

Otsu Algorithm. The Otsu algorithm is a popular method for automatic image thresholding, which is the process of separating an image into two or more regions based on a clustering algorithm according to the intensity levels of its pixels. Developed by Nobuyuki Otsu,[58] this algorithm aims to find the optimal threshold value that minimizes the intraclass variance of the foreground and background pixels while maximizing the interclass variance. The Otsu algorithm is computationally efficient and does not require prior knowledge of the image distribution, making it a versatile tool for image segmentation in various fields, including medical imaging, text recognition, and object detection.

Box Normalization. In most real STM experiment cases, the magnification ratio of the vision field is fixed in order to facilitate manipulation, observation, and molecule comparison. Considering this, in our ReSTOLO framework, we adjust the center location, shape, and size of the bounding box generated in the locating stage by the weighted pixel method and recapture the single molecule using a new, normalized, fixed-size box. Unlike Region Proposal and RoI pooling of Faster R-CNN,[72] this method principally does not lose any details of the image because there is no additional mathematical averaging process like RoI pooling. The weighted pixel square sum average method is presented in Figure 5. Complete algorithmic details are available in Supporting Information 10.1.

ASSOCIATED CONTENT

Data Availability Statement

The key findings of this study are supported by data available within the article and the Supporting Information files. All codes are freely available on GitHub (https://github.com/ApolloniusWei/ReSTOLO). All the relevant data is available from the authors upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.5c03730.

> Python packages for establishing, deploying, and analyzing the whole framework; the model performance calculation method; PCA analysis of surface molecule direction; FFT analysis of image resolution and SRResNet training; data set division and reversed validation; data augmentation methods and their contributions; the structure, training details, and 10-fold validation of YOLO v5.m, ResNet-101; amelioration and hyperparameter selections for ReSTOLO, benchmark YOLO v5.m, and Faster R-CNN; other STM detection images; principals and robustness of the box normalization method; discussions about alternatives of models and strategies; and detection performance on two other surface systems (PDF)

AUTHOR INFORMATION

Corresponding Authors

Qigang Zhong − *State Key Laboratory of Bioinspired Interfacial Materials Science, Institute of Functional Nano and Soft Materials (FUNSOM), Soochow University, Suzhou 215123, China;* orcid.org/0000-0001-7100-7363; Email: qgzhong@suda.edu.cn

Lifeng Chi − *State Key Laboratory of Bioinspired Interfacial Materials Science, Institute of Functional Nano and Soft Materials (FUNSOM), Soochow University, Suzhou 215123, China;* orcid.org/0000-0003-3835-2776; Email: chilf@suda.edu.cn

Shixuan Du − *Institute of Physics and University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China;* orcid.org/0000-0001-9323-1307; Email: sxdu@iphy.ac.cn

Authors

Zixuan Wei − *Institute of Physics and University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China*

Jinbo Pan − *Institute of Physics and University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China;* orcid.org/0000-0003-2612-8232

Fang Han Lim − *Institute of Physics and University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.5c03730

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by funding from the National Key Research and Development Program of China (2022YFA1204100), the National Natural Science Foundation of China (Nos. U24A20496, 22472115 and 62488201), Collaborative Innovation Center of Suzhou Nano Science & Technology, the 111 Project, Suzhou Key Laboratory of Surface and Interface of Intelligent Matter (SZS2022011), the Gusu Innovation and Entrepreneurship Talent Program Major Innovation Team (ZXD2023002), Innovative Center for Molecular Science of Surface and Interface (MOSSI), Soochow University.

■ ABBREVIATIONS

TG, task group; SR, super resolution

■ REFERENCES

(1) Gao, H.-Y. Recent advances in organic molecule reactions on metal surfaces. *Phys. Chem. Chem. Phys.* **2024**, *26* (28), 19052−19068.
(2) Mali, K. S.; Pearce, N.; De Feyter, S.; Champness, N. R. Frontiers of supramolecular chemistry at solid surfaces. *Chem. Soc. Rev.* **2017**, *46* (9), 2520−2542.
(3) Frezza, F.; Matěj, A.; Sánchez-Grande, A.; Carrera, M.; Mutombo, P.; Kumar, M.; Curiel, D.; Jelínek, P. On-surface synthesis of a radical 2D supramolecular organic framework. *J. Am. Chem. Soc.* **2024**, *146* (5), 3531−3538.
(4) Zhou, Q.; Kaappa, S.; Malola, S.; Lu, H.; Guan, D.; Li, Y.; Wang, H.; Xie, Z.; Ma, Z.; Häkkinen, H.; et al. Real-space imaging with pattern recognition of a ligand-protected Ag_{374} nanocluster at sub-molecular resolution. *Nat. Commun.* **2018**, *9* (1), No. 2948.
(5) Lin, T.; Shang, X. S.; Adisoejoso, J.; Liu, P. N.; Lin, N. Steering on-surface polymerization with metal-directed template. *J. Am. Chem. Soc.* **2013**, *135* (9), 3576−3582. From NLM Medline.
(6) Shi, K. J.; Shu, C. H.; Wang, C. X.; Wu, X. Y.; Tian, H.; Liu, P. N. On-surface Heck reaction of aryl bromides with alkene on Au(111) with palladium as catalyst. *Org. Lett.* **2017**, *19* (11), 2801−2804. From NLM PubMed-not-MEDLINE.
(7) Song, X.; Liu, J.; Zhang, T.; Chen, L. 2D conductive metal-organic frameworks for electronics and spintronics. *Sci. China: Chem.* **2020**, *63* (10), 1391−1401.
(8) Shi, P.-P.; Lu, S.-Q.; Song, X.-J.; Chen, X.-G.; Liao, W.-Q.; Li, P.-F.; Tang, Y.-Y.; Xiong, R.-G. Two-dimensional organic−inorganic perovskite ferroelectric semiconductors with fluorinated aromatic spacers. *J. Am. Chem. Soc.* **2019**, *141* (45), 18334−18340.
(9) Khajetoorians, A. A.; Wegner, D.; Otte, A. F.; Swart, I. Creating designer quantum states of matter atom-by-atom. *Nat. Rev. Phys.* **2019**, *1* (12), 703−715.
(10) Jandt, K. D. Developments and perspectives of scanning probe microscopy (SPM) on organic materials systems. *Mater. Sci. Eng.* **1998**, *21* (5−6), 221−295.
(11) Bian, K.; Gerber, C.; Heinrich, A. J.; Müller, D. J.; Scheuring, S.; Jiang, Y. Scanning probe microscopy. *Nat. Rev. Methods Primers* **2021**, *1* (1), 36 DOI: 10.1038/s43586-021-00033-2.
(12) Giessibl, F. J. The qPlus sensor, a powerful core for the atomic force microscope. *Rev. Sci. Instrum.* **2019**, *90* (1), No. 011101.
(13) Narita, A.; Wang, X.-Y.; Feng, X.; Müllen, K. New advances in nanographene chemistry. *Chem. Soc. Rev.* **2015**, *44* (18), 6616−6643.
(14) Xu, K.; Urgel, J. I.; Eimre, K.; Di Giovannantonio, M.; Keerthi, A.; Komber, H.; Wang, S.; Narita, A.; Berger, R.; Ruffieux, P.; et al. On-surface synthesis of a nonplanar porous nanographene. *J. Am. Chem. Soc.* **2019**, *141* (19), 7726−7730. From NLM PubMed-not-MEDLINE.
(15) Ewen, P. R.; Sanning, J.; Doltsinis, N. L.; Mauro, M.; Strassert, C. A.; Wegner, D. Unraveling orbital hybridization of triplet emitters at the metal-organic interface. *Phys. Rev. Lett.* **2013**, *111* (26), No. 267401. From NLM PubMed-not-MEDLINE.
(16) Mayder, D. M.; Tonge, C. M.; Nguyen, G. D.; Hojo, R.; Paisley, N. R.; Yu, J.; Tom, G.; Burke, S. A.; Hudson, Z. M. Design of high-performance thermally activated delayed fluorescence emitters containing s-triazine and s-heptazine with molecular orbital visualization by STM. *Chem. Mater.* **2022**, *34* (6), 2624−2635.
(17) Müllen, K.; Rabe, J. P. Nanographenes as active components of single-molecule electronics and how a scanning tunneling microscope puts them to work. *Acc. Chem. Res.* **2008**, *41* (4), 511−520.
(18) Hu, Y.; Zhong, G.; Guan, Y. S.; Lee, N. H.; Zhang, Y.; Li, Y.; Mitchell, T.; Armstrong, J. N.; Benedict, J.; Hla, S. W.; Ren, S. Alkali-metal-intercalated percolation network regulates self-assembled electronic aromatic molecules. *Adv. Mater.* **2019**, *31* (11), No. e1807178. From NLM PubMed-not-MEDLINE.
(19) Jelínek, P. High resolution SPM imaging of organic molecules with functionalized tips. *J. Phys.:Condens. Matter* **2017**, *29* (34), No. 343002. From NLM PubMed-not-MEDLINE.
(20) Song, L.; Yang, B.; Liu, F.; Niu, K.; Han, Y.; Wang, J.; Zheng, Y.; Zhang, H.; Li, Q.; Chi, L. Synthesis of two-dimensional metal−organic frameworks via dehydrogenation reactions on a Cu(111) surface. *J. Phys. Chem. C* **2020**, *124* (23), 12390−12396.
(21) Zhang, H. Y.; Zhang, Z. X.; Song, X. J.; Chen, X. G.; Xiong, R. G. Two-dimensional hybrid perovskite ferroelectric induced by perfluorinated substitution. *J. Am. Chem. Soc.* **2020**, *142* (47), 20208−20215. From NLM PubMed-not-MEDLINE.
(22) Liu, J.; Abel, M.; Lin, N. On-surface synthesis: A new route realizing single-layer conjugated metal-organic structures. *J. Phys. Chem. Lett.* **2022**, *13* (5), 1356−1365. From NLM PubMed-not-MEDLINE.
(23) Hla, S.-W.; Rieder, K.-H. STM control of chemical reactions: single-molecule synthesis. *Annu. Rev. Phys. Chem.* **2003**, *54* (1), 307−330.
(24) Komeda, T.; Kim, Y.; Fujita, Y.; Sainoo, Y.; Kawai, M. Local chemical reaction of benzene on Cu(110) via STM-induced excitation. *J. Chem. Phys.* **2004**, *120* (11), 5347−5352.
(25) Avouris, P.; Lyo, I. W. Probing the chemistry and manipulating surfaces at the atomic scale with the STM. *Appl. Surf. Sci.* **1992**, *60−61*, 426−436.
(26) Dazzi, A.; Prater, C. B. AFM-IR: Technology and applications in nanoscale infrared spectroscopy and chemical imaging. *Chem. Rev.* **2017**, *117* (7), 5146−5173.
(27) Wu, X.; Delbianco, M.; Anggara, K.; Michnowicz, T.; Pardo-Vargas, A.; Bharate, P.; Sen, S.; Pristl, M.; Rauschenbach, S.; Schlickum, U.; et al. Imaging single glycans. *Nature* **2020**, *582* (7812), 375−378.
(28) Lorente, N.; Rurali, R.; Tang, H. Single-molecule manipulation and chemistry with the STM. *J. Phys.:Condens. Matter* **2005**, *17* (13), S1049.
(29) Alldritt, B.; Hapala, P.; Oinonen, N.; Urtev, F.; Krejci, O.; Federici Canova, F.; Kannala, J.; Schulz, F.; Liljeroth, P.; Foster, A. S. Automated structure discovery in atomic force microscopy. *Sci. Adv.* **2020**, *6* (9), No. eaay6913.
(30) 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. *Software Impacts* **2022**, *12*, No. 100301.
(31) Su, J.; Li, J.; Guo, N.; Peng, X.; Yin, J.; Wang, J.; Lyu, P.; Luo, Z.; Mouthaan, K.; Wu, J.; et al. Intelligent synthesis of magnetic nanographenes via chemist-intuited atomic robotic probe. *Nat. Synth.* **2024**, *3* (4), 466−476.
(32) Ziatdinov, M.; Ghosh, A.; Wong, C. Y.; Kalinin, S. V. AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nat. Mach. Intell.* **2022**, *4* (12), 1101−1112.
(33) Usman, M.; Wong, Y. Z.; Hill, C. D.; Hollenberg, L. C. L. Framework for atomic-level characterisation of quantum computer arrays by machine learning. *npj Comput. Mater.* **2020**, *6* (1), 19.
(34) Ghosh, A.; Sumpter, B. G.; Dyck, O.; Kalinin, S. V.; Ziatdinov, M. Ensemble learning-iterative training machine learning for

uncertainty quantification and automated experiment in atom-resolved microscopy. *npj Comput. Mater.* **2021**, *7* (1), 100.
(35) Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C. W.; Choudhary, A.; Agrawal, A.; Billinge, S. J. L.; et al. Recent advances and applications of deep learning methods in materials science. *npj Comput. Mater.* **2022**, *8* (1), 59.
(36) Narasimha, G.; Kong, D.; Regmi, P.; Jin, R.; Gai, Z.; Vasudevan, R.; Ziatdinov, M. Uncovering multiscale structure-property correlations via active learning in scanning tunneling microscopy. *npj Comput. Mater.* **2025**, *11* (1), 189.
(37) Volk, A. A.; Epps, R. W.; Yonemoto, D. T.; Masters, B. S.; Castellano, F. N.; Reyes, K. G.; Abolhasani, M. AlphaFlow: autonomous discovery and optimization of multi-step chemistry using a self-driven fluidic lab guided by reinforcement learning. *Nat. Commun.* **2023**, *14* (1), No. 1403.
(38) Ziatdinov, M.; Fuchs, U.; Owen, J. H.; Randall, J. N.; Kalinin, S. V. Robust Multi-Scale Multi-Feature Deep Learning for Atomic and Defect Identification in Scanning Tunneling Microscopy on H-Si (100) 2 × 1 Surface 2020 https://arxiv.org/abs/2002.04716. (accessed 11 Feb, 2020).
(39) Vasudevan, R. K.; Laanait, N.; Ferragut, E. M.; Wang, K.; Geohegan, D. B.; Xiao, K.; Ziatdinov, M.; Jesse, S.; Dyck, O.; Kalinin, S. V. Mapping mesoscopic phase evolution during E-beam induced transformations via deep learning of atomically resolved images. *npj Comput. Mater.* **2018**, *4* (1), 30.
(40) Sadri, A.; Petersen, T. C.; Terzoudis-Lumsden, E. W. C.; Esser, B. D.; Etheridge, J.; Findlay, S. D. Unsupervised deep denoising for four-dimensional scanning transmission electron microscopy. *npj Comput. Mater.* **2024**, *10* (1), 243.
(41) Ludacka, U.; He, J.; Qin, S.; Zahn, M.; Christiansen, E. F.; Hunnestad, K. A.; Zhang, X.; Yan, Z.; Bourret, E.; Kézsmárki, I.; et al. Imaging and structure analysis of ferroelectric domains, domain walls, and vortices by scanning electron diffraction. *npj Comput. Mater.* **2024**, *10* (1), 106.
(42) Krull, A.; Hirsch, P.; Rother, C.; Schiffrin, A.; Krull, C. Artificial-intelligence-driven scanning probe microscopy. *Commun. Phys.* **2020**, *3* (1), 54.
(43) Kurki, L.; Oinonen, N.; Foster, A. S. Automated structure discovery for scanning tunneling microscopy. *ACS Nano* **2024**, *18* (17), 11130−11138. From NLM PubMed-not-MEDLINE.
(44) 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* (27), 10177−10188.
(45) Yuan, S.; Zhu, Z.; Lu, J.; Zheng, F.; Jiang, H.; Sun, Q. Applying a deep-learning-based keypoint detection in analyzing surface nanostructures. *Molecules* **2023**, *28* (14), 5387.
(46) Huang, W.; Jin, Y.; Li, Z.; Yao, L.; Chen, Y.; Luo, Z.; Zhou, S.; Lin, J.; Liu, F.; Gao, Z.; et al. Auto-resolving the atomic structure at van der Waals interfaces using a generative model. *Nat. Commun.* **2025**, *16* (1), No. 2927.
(47) Zhu, Z.; Lu, J.; Zheng, F.; Chen, C.; Lv, Y.; Jiang, H.; Yan, Y.; Narita, A.; Müllen, K.; Wang, X. Y.; Sun, Q. A deep-learning framework for the automated recognition of molecules in scanning-probe-microscopy images. *Angew. Chem., Int. Ed.* **2022**, *61* (49), No. e202213503.
(48) Ziatdinov, M.; Maksov, A.; Kalinin, S. V. Learning surface molecular structures via machine vision. *npj Comput. Mater.* **2017**, *3* (1), 31.
(49) Pearl, J. Fusion, propagation, and structuring in belief networks. *Artif. Intell.* **1986**, *29*, 241−288.
(50) Li, S. Z. Markov Random Field Models in Computer Vision. In *European Conference on Computer Vision*; Springer, 1994; Vol. *2*, pp 361−370.
(51) Zhong, Q.; Jung, J.; Kohrs, D.; Kaczmarek, L. A.; Ebeling, D.; Mollenhauer, D.; Wegner, H. A.; Schirmeisen, A. Deciphering the mechanism of on-surface dehydrogenative C−C coupling reactions. *J. Am. Chem. Soc.* **2024**, *146* (3), 1849−1859.
(52) Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9* (11), 2579−2605.
(53) Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. *Adv. Neural. Inf. Process. Syst.* **2016**, *29*, 9.
(54) Wang, Y.; Yao, Q.; Kwok, J. T.; Ni, L. M. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Comput. Surv.* **2020**, *53* (3), 63.
(55) Lu, J.; Gong, P.; Ye, J.; Zhang, J.; Zhang, C. A survey on machine learning from few samples. *Pattern Recognit.* **2023**, *139*, No. 109480.
(56) Huang, W.; Jin, Y.; Li, Z.; Yao, L.; Chen, Y.; Luo, Z.; Zhou, S.; Lin, J.; Liu, F.; Gao, Z. Auto-resolving atomic structure at van der Waal interfaces using a generative model. *Nat. Commun.* **2025**, *16* (1), No. 2927.
(57) Cooley, J. W.; Lewis, P. A.; Welch, P. D. The fast Fourier transform and its applications. *IEEE Trans. Educ.* **1969**, *12* (1), 27−34.
(58) Otsu, N. A threshold selection method from gray-level histograms. *Automatica* **1975**, *11* (285−296), 23−27.
(59) Simonyan, K.; Zisserman, A.Very Deep Convolutional Networks for Large-Scale Image Recognition *Proc. Int. Conf. Learn. Representat.* 2015, pp 1−14.
(60) He, K.; Zhang, X.; Ren, S.; Sun, J.Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* 2016; pp 770−778.
(61) Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.et al.An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale *Int. Conf. Learn. Represent.* 2021, pp 611−631.
(62) Tan, M.; Le, Q. V.EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning*; PMLR, 2019; pp 6105−6114.
(63) Becker, E.; Pandit, P.; Rangan, S.; Fletcher, A. K. Instability and Local Minima in GAN Training with Kernel Discriminators. *Adv. Neural. Inf. Process. Syst.* **2022**, *35*, 20300−20312.
(64) Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *International Conference on Neural Information Processing Systems*; MIT Press, 2014; pp 2672−2680.
(65) Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A.You Only Look Once: Unified, Real-Time Object Detection. In *IEEE Conference on Computer Vision and Pattern Recognition* 2016; pp 779−788.
(66) Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **2023**, *11* (7), 677.
(67) Hatab, M.; Malekmohamadi, H.; Amira, A. Surface Defect Detection Using YOLO Network. In *Intelligent Systems and Applications*; Springer, 2021; Vol. *1*, pp 505−515.
(68) Benjumea, A.; Teeti, I.; Cuzzolin, F.; Bradley, A. YOLO-Z: Improving Small Object Detection in YOLOv5 for Autonomous Vehicles 2021 https://arxiv.org/abs/2112.11798. (accessed Dec 22, 2021).
(69) Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A comprehensive review of yolo architectures in computer vision: From YOLOv1 toYOLOv8 and YOLO-nas. *Mach. Learn. Knowl. Extr.* **2023**, *5* (4), 1680−1716.
(70) Balduzzi, D.; Frean, M.; Leary, L.; Lewis, J.; Ma, K. W.-D.; McWilliams, B. The Shattered Gradients Problem: If Resnets are the Answer, Then What is the Question?. In *International Conference on Machine Learning*; PMLR, 2017; pp 342−350.
(71) Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W.Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *IEEE Conference on Computer Vision and Pattern Recognition* 2017; pp 4681−4690.
(72) Girshick, R.Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* 2015; pp 1440−1448.