

Synthetic-data-driven You-Only-Look-Once–Based Detection of Computer Numerical Control Machining Chips: Comparative Analysis of Omniverse Generated and Manually Annotated Images

Ta-Jen Peng,* Jr-Rung Chen, En-Cheng Liou, Yen-Ling Lin, and Sheng-Wei Wang

Department of Intelligent Automation Engineering, National Chin-Yi University of Technology, Taiwan

(Received July 7, 2025; accepted November 11, 2025)

Keywords: CNC machining, chip detection, synthetic data, NVIDIA Omniverse, YOLO model

With the increasing automation of production lines, chip residues generated during computer numerical control (CNC) machining may result in tool damage, equipment malfunctions, unscheduled downtime, and increased maintenance costs if not promptly identified and managed. To address this problem, we propose a synthetic data generation framework that integrates 2D image-based modeling with the NVIDIA Omniverse simulation platform to rapidly produce large-scale annotated datasets of cutting tool chips. These synthetic datasets were subsequently employed for training a You Only Look Once (YOLO) object detection model. A comparative analysis was performed between synthetic images generated by Omniverse and manually annotated real images, evaluating their detection performance and labeling costs. Experimental results demonstrated that synthetic data significantly reduces manual labeling efforts while maintaining high detection accuracy. The proposed system facilitates the early detection of abnormal chipping or chip entanglement, providing actionable insights for preventive maintenance and adaptive machining scheduling. Consequently, this approach enhances intelligent tool condition monitoring and predictive maintenance capabilities, thus offering essential technological support for smart manufacturing and automated production lines.

1. Introduction

With the rapid development of production line automation and smart manufacturing technologies, computer numerical control (CNC) machining systems have become a fundamental component in precision manufacturing. However, during high-speed cutting processes, chip residues generated from machining—if not promptly detected and managed—may result in tool wear, machining defects, equipment malfunctions, and unplanned downtime. Such issues ultimately increase maintenance costs and compromise production quality and operational stability. Therefore, the effective identification and monitoring of chip conditions during

*Corresponding author: e-mail: TJPeng@ncut.edu.tw
<https://doi.org/10.18494/SAM5839>

machining are essential for achieving predictive maintenance and ensuring process reliability. Recent advances in object detection models, notably the You Only Look Once (YOLO) series, have been extensively employed in industrial vision applications.^(1–14) However, their practical performance often suffers owing to difficulties in obtaining adequate training datasets, especially for tool chips characterized by high variability and irregular geometries. Manual annotation of such data is not only labor-intensive but also inadequate in representing the full diversity of machining conditions. To overcome this challenge, in this study, we propose a synthetic data generation pipeline that integrates 2D image modeling techniques with the NVIDIA Omniverse simulation platform, enabling the rapid creation of large-scale annotated datasets specifically for machining tool chips, intended for training and validating YOLO models.^(15–20) Comparative experiments were conducted to evaluate the detection accuracy and training efficiency between YOLO models trained on Omniverse-generated synthetic data and those trained on manually annotated real-world datasets. Experimental results indicate that employing synthetic datasets substantially reduces the human effort required for labeling while preserving high detection accuracy and generalization performance. This approach significantly enhances the capability for real-time anomaly detection on production lines and promotes intelligent machining operations and predictive maintenance strategies. Consequently, the proposed methodology represents an effective and practical solution for advanced manufacturing systems.

2. Experimental Setup and Configuration

The experimental platform developed in this study consists of three primary components.

- (1) **Machining tools and associated chip samples:** The machining tools used in this study are shown in Fig. 1. These tools were employed to generate a dataset comprising real images captured from actual machining operations.
- (2) **Industrial imaging system:** As shown in Fig. 2, a 2D industrial camera (RER-USB16M01) was mounted directly above the machining area. This camera features a 16-megapixel resolution (4656×3496 pixels) and supports video acquisition at 30 frames per second. It was employed to capture images of machining chips postprocess.
- (3) **Computational hardware:** Synthetic data generation and YOLO model training were performed on a high-performance workstation equipped with an NVIDIA RTX 4090 GPU and 64 GB of RAM.

Synthetic data were generated by integrating 2D image-based modeling techniques within the NVIDIA Omniverse simulation platform, enabling the realistic emulation of various machining scenarios and chip states. The YOLO model was trained and subsequently evaluated on both synthetic and real datasets to systematically analyze the effect of different data sources on detection performance. The experimental setup was aimed at replicating practical machining conditions, thus facilitating the comprehensive validation of the proposed detection system's performance, robustness, and feasibility. To construct the annotated dataset, high-resolution images of machining chips were acquired against a uniform white background using the aforementioned industrial camera. Annotations were performed frame-by-frame using LabelImg



Fig. 1. (Color online) Machining tools used in this study: (a) four-flute end mill, (b) twist drill, (c) 45° chamfering tool, and (d) internal turning tool (from left to right).



Fig. 2. (Color online) 2D industrial camera.

software, with bounding boxes delineating the boundaries of tools and chips, following the YOLOv8 annotation standard. Annotation consistency was ensured through a dual-annotator cross-validation protocol supplemented by random audits. Annotation quality was continuously assessed using metrics such as mean Intersection over Union and Cohen's Kappa coefficient (K). Model training and evaluation were conducted using exclusively real-image datasets, with the specific objective of assessing the impact of manually annotated data on key detection performance metrics, including mean average precision (*mAP*), *Precision*, and *Recall*. This approach also served to validate the system's reliability and robustness for deployment in authentic machining environments.

3. Methodology

In this section, we describe in detail the system architecture and experimental methodology, specifically addressing image data collection, synthetic image generation, YOLO model training, and the comparative analysis of detection performance.

3.1 Real-image data collection and annotation

Initially, a dataset consisting of real images was established as the baseline for training and evaluating the YOLO model. Images were captured using a 2D industrial camera positioned directly above the CNC machining area, enabling the consistent recording of real machining scenarios, including tool chips scattered across the workpiece surface and within tool-holding regions. To ensure dataset diversity and enhance the generalization capability of the model, image collection included various types of machining tools, workpiece materials, and lighting conditions. In total, 800 images were collected and subsequently divided into a training set (70%), a validation set (20%), and a test set (10%). Each image contained approximately two to four instances of chips, displaying variations in orientation, scale, and occlusion. This diversity was intended to facilitate the model's learning of detailed and edge-level features. All images were manually annotated using the LabelImg software, employing bounding boxes to delineate both tool and chip regions in accordance with the YOLO annotation format. Regions irrelevant to object detection, such as background noise, workpiece edges, and reflections, were intentionally excluded from the annotation process. To ensure consistency and accuracy of annotations, a dual-annotator cross-checking protocol supplemented by random verification audits was employed. The complete annotation process required approximately 15 h. Additional annotations were applied to samples exhibiting characteristics such as overlap, blurring, or glare, which typically degrade object recognition performance, thus enhancing the practical effectiveness of the training dataset. Figure 3. provides an illustrative example of a fully annotated image, clearly depicting the bounding boxes assigned to chips and tools, thereby supplying precise training targets for the YOLO model.

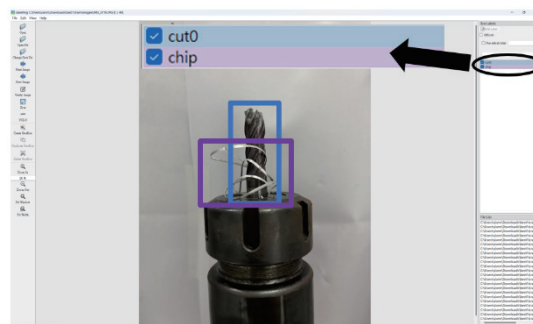


Fig. 3. (Color online) Example of chip annotation in a real machining image.

3.2 Synthetic image generation

In this study, a 2D modeling approach was employed to reconstruct the morphology of cutting chips through image feature matching and 3D point cloud reconstruction techniques.

The mathematical formulation of image feature matching is

$$D(x, x') = \|x - T(x')\|^2, \quad (1)$$

where x and x' represent the coordinates of corresponding feature points from two images, and $T(\bullet)$ denotes the transformation function mapping coordinates from one image to another (e.g., perspective or affine transformations).

Subsequently, the 3D point cloud reconstruction was performed by minimizing the projection error, mathematically expressed as

$$\operatorname{argmin}_{X, P_j} \sum_{i=1}^M \sum_{j=1}^M \|x_{ij} - P_j X_i\|^2, \quad (2)$$

where X_i denotes the coordinates of the i th 3D point, x_{ij} is its projection onto the j th image, and P_j represents the projection matrix of the j th image.

After completing the 3D reconstruction using Eqs. (1) and (2), the resulting models were imported into the NVIDIA Omniverse simulation platform. Through careful configuration of simulation parameters, including light source direction, intensity, and background texture, a large and diverse collection of synthetic images with accurate ground-truth annotations was systematically generated for subsequent model training and validation. In this study, we employed a 2D modeling approach to reconstruct the detailed geometry and appearance of metal chips, followed by the creation of simulated machining scenarios within the Omniverse platform. Key simulation parameters such as background texture, lighting conditions, and camera viewing angles were methodically adjusted, enabling the automated generation of an extensive and varied synthetic image dataset along with corresponding annotations. These synthetic images were then utilized to augment the training dataset for the YOLO object detection model.

During the data acquisition stage, cutting tool and chip images were captured under strictly controlled conditions. The camera settings included Pro shooting mode, ISO fixed at 200, shutter speed set to 1/60 s, and white balance at 5500 K, with manual focus employed to ensure optimal image clarity. When ambient illumination was insufficient, shutter speed was adjusted between 1/50 and 1/30 s according to standard visual recognition criteria. For each tool, approximately 40 to 80 images were captured, whereas chip datasets consisted of 200 to 450 images each, supplemented by up to 30 additional images as necessary to ensure comprehensive coverage for accurate modeling.

A three-stage capturing strategy was implemented, photographing each object systematically from upper, middle, and lower viewpoints. Image overlap was deliberately maintained between 60 and 80% to enhance reconstruction accuracy during the photogrammetric modeling. To

minimize glare interference, all indoor lights were turned off during photography, with a single desk lamp—softened by a white diffusion sheet—positioned at approximately a 45-degree angle to the surface, thereby enhancing the visibility of fine surface details. Additionally, subtly textured background materials were employed to prevent modeling failures associated with plain white or monochromatic backgrounds.

Following image acquisition, the collected images were processed using Reality Capture software for photogrammetric modeling. The software automatically generated point clouds through image feature matching, thus establishing an initial 3D spatial structure and camera perspective arrangement, as demonstrated in Fig. 4.

Figure 5 shows the initial 3D reconstruction of the central chip and associated workpiece structure. The surrounding white points indicate the positions of virtual cameras, estimated during reconstruction, thereby reflecting the trajectory and distribution of image-capture viewpoints. If the preliminary reconstruction exhibits fragmentation or incomplete regions, additional images are acquired, or the entire capture sequence is repeated, depending on the observed reconstruction quality. Following successful geometric reconstruction, texture mapping is applied to enhance realism. Figure 5 depicts the final reconstructed model, highlighting detailed and realistic 3D structures with high physical fidelity.

The reconstructed chip model demonstrates structural completeness, while the associated workpiece clearly exhibits realistic surface details. Such high-quality modeling significantly enhances the accuracy and applicability of subsequent synthetic data generation processes.

After texture mapping, the completed model is imported into a 3D modeling software, such as Blender, for precise object segmentation and spatial alignment. Figure 6 shows the isolation and standardized positioning of individual chip instances within the virtual environment, which facilitates subsequent integration into simulated machining scenarios and annotation layers within the Omniverse simulation framework.

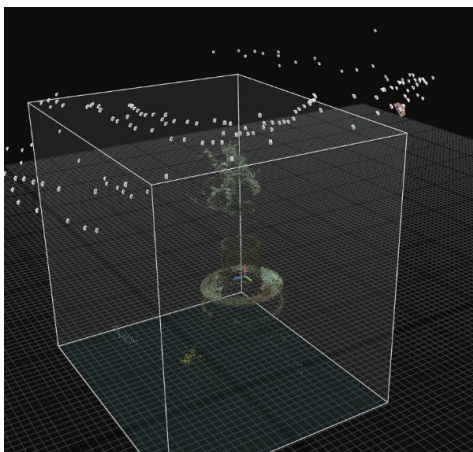


Fig. 4. (Color online) Visualization of camera positions and point cloud structure during photogrammetric modeling.

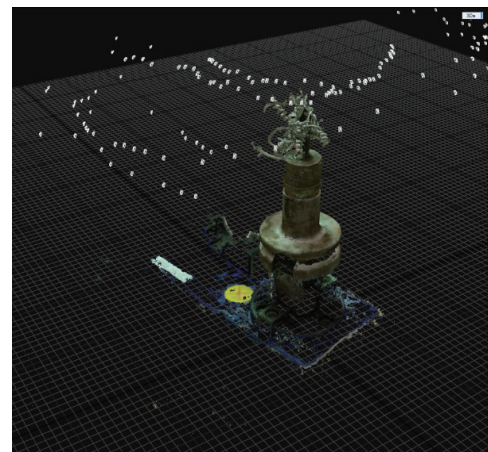


Fig. 5. (Color online) Textured 3D model of reconstructed chip and workpiece.

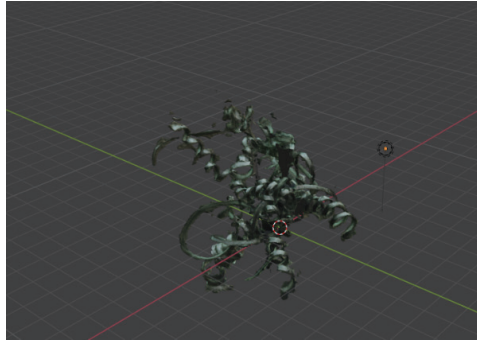


Fig. 6. (Color online) Individual chip model extracted from reconstructed 3D scene.

The standardized alignment of the chip model at the origin of the coordinate system streamlines subsequent scene integration and enables efficient batch processing for synthetic data generation.

3.3 Incremental training and evaluation of YOLO models

We performed model training using two distinct datasets: one comprising solely real images and the other consisting of mixed images (real and synthetic). Both datasets underwent a five-stage incremental training process, where the number of training images at each stage was progressively set to 2, 100, 200, 300, and 400. After each training stage, the training datasets and their corresponding optimal model weights (best.pt) were saved independently without overwriting previous stages. This approach facilitated the subsequent analysis of how increasing data volume impacts the model's detection performance.

Training hyperparameters remained consistent across both datasets throughout the entire process, with the batch size fixed at 8, the total number of epochs set to 100, and the Adam optimizer utilized for training.

Upon completion of model training, both models were assessed using the same independent test set. For each evaluation, model weights saved at each training stage were sequentially loaded and tested. The detection performance was quantitatively measured using four key metrics: *mAP*, *Precision*, *Recall*, and *F1-score*. This comprehensive evaluation allowed for an in-depth comparison of detection capabilities across different data compositions and varying amounts of training data.

The optimization of the YOLO model was carried out using the following loss function.

$$\begin{aligned}
 Loss = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\omega_i - \hat{\omega}_i)^2 + (h_i - \hat{h}_i)^2 \right] \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C_i - C_i)^2 \\
 & + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} (C_i - \hat{C}_i)^2 + \sum_{i=0}^{S^2} I_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned} \tag{3}$$

Here, (x_i, y_i, w_i, h_i) denotes the ground truth bounding box coordinates, and $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$ represents the predicted bounding box coordinates. C_i and \hat{C}_i are the ground truth and predicted object confidence scores, respectively. $p_i(c)$ and $\hat{p}_i(c)$ indicate the ground truth and predicted class probabilities, respectively. The terms λ_{coord} and λ_{noobj} are weighting factors employed to balance the coordinate regression loss and the no-object confidence loss during model training, respectively.

3.4 System simulation and validation

In addition to synthetic data generation, the Omniverse platform served as a simulation environment for validating the chip recognition system developed in this study. Specifically, the platform was utilized to emulate chip detection behaviors and to evaluate system robustness under dynamically varying scene conditions. Key simulation parameters—such as lighting directions, camera angles, background textures, and levels of occlusion—were systematically adjusted to assess the adaptability and generalization capabilities of the detection model under realistic operational scenarios.

A comparative analysis was conducted between models trained on manually annotated real data and those trained on synthetically generated data under simulated conditions. The results indicated that models trained with synthetic data consistently exhibited stable and reliable detection performance across diverse scenarios, demonstrating their feasibility and practical advantages for real-world implementation. Notably, under challenging scenarios such as varied illumination conditions and complex, cluttered backgrounds, the synthetic-trained model maintained high detection accuracy, affirming its applicability to on-site monitoring tasks.

As depicted in Fig. 7, the developed model accurately identified chip regions within the simulated Omniverse environment, thus facilitating timely operator intervention when necessary. Similarly, Fig. 8 presents the results of instance segmentation, highlighting the model's capacity to distinguish between chip instances and workpiece surfaces, which is critical for downstream precision-based decision-making.

In summary, Omniverse serves not merely as a tool for synthetic image generation but also as an integrated platform for comprehensive system simulation and validation. Through high-fidelity visual modeling and flexible configuration of simulation scenarios, the platform

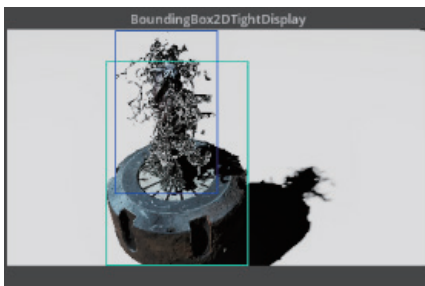


Fig. 7. (Color online) Bounding box annotations of chip detection in the Omniverse simulation environment.

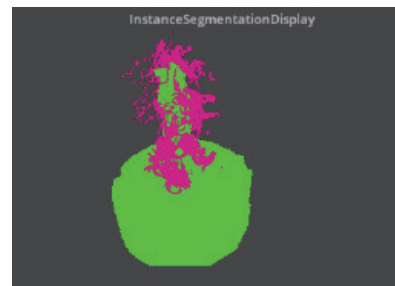


Fig. 8. (Color online) Instance segmentation results of chips and workpiece within the same Omniverse scene.

effectively replicates realistic machining environments characterized by variations in illumination, viewing angles, and background complexities, enabling the precise evaluation of the chip detection process.

Validation experiments, combining both real-world and simulated data scenarios, confirm that the YOLO-based model developed in this research reliably identifies chip targets under diverse and challenging operational conditions. This underscores the model's robustness, reliability, and readiness for practical deployment.

Overall, the Omniverse platform provides a highly controllable, reproducible, and scalable testing framework, significantly accelerating model development and iteration. Consequently, it effectively meets the stringent requirements of intelligent manufacturing applications, particularly ensuring the stability and practical applicability of vision-based detection systems.

4. Results and Discussion

In this section, we present a comparative analysis of detection performance and training convergence behavior for YOLOv8 models trained using two distinct datasets: purely real images and hybrid (combined synthetic and real) data. A five-stage incremental training strategy was adopted to systematically evaluate the effect of dataset size and composition on the detection accuracy and generalization capability of the models.

Model performance was quantitatively evaluated using four standard metrics: *mAP*, *Precision*, *Recall*, and *F1-score*. The mathematical definitions of these metrics are shown in Eqs. (4)–(7).

mAP is defined as the mean of Average Precision (*AP*) values across all classes.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i, \quad AP = \int_0^1 p(r) dr \quad (4)$$

Precision quantifies the proportion of correctly predicted positive cases.

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall indicates the proportion of actual positive cases correctly identified by the model.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

The *F1-score*, representing the harmonic mean of *Precision* and *Recall*, is calculated as

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

In this context, *TP*, *FP*, and *FN* denote True Positives, False Positives, and False Negatives, respectively.

4.1 Performance comparison across different data sources

With a standardized test set, the model trained exclusively with synthetic data achieved an $mAP@0.5$ of 0.86. This result was only marginally lower (by 0.02) than that of the model trained solely on manually annotated real-world data, which attained an $mAP@0.5$ of 0.88. Similarly, the *Precision* and *Recall* metrics of both models exhibited close alignment. These findings indicate that the YOLO model trained exclusively with synthetic data maintained stable and accurate detection performance even without exposure to real images. Such outcomes underscore the practical viability of utilizing high-fidelity synthetic datasets, particularly in scenarios where real data collection is challenging or manual annotation becomes prohibitively expensive.

Further enhancement in detection performance was observed when employing a hybrid training strategy. Specifically, the model trained with a combination of 70% synthetic data and 30% real data achieved an improved $mAP@0.5$ of 0.91, with corresponding *Precision* and *Recall* values reaching 0.93 and 0.92, respectively, and an *F1-score* of 0.90. Among the evaluated data configurations, the hybrid dataset approach clearly demonstrated superior overall performance.

These results suggest that even a limited proportion of real data integrated with synthetic data significantly improves model robustness and generalization capability under challenging conditions such as complex backgrounds, variable illumination, and blurred object boundaries. Thus, the combination of synthetic and real data presents complementary advantages for the robust training of detection models.

From the standpoint of dataset generation efficiency, the hybrid approach provides substantial cost advantages. As illustrated in Fig. 9, the Omniverse-based data generation pipeline—including modeling, texturing, and automatic annotation—requires mere seconds per image, resulting in notably lower overall processing times. In contrast, manual annotation typically requires approximately 1.5 to 2 min per image, with workload scaling rapidly as the dataset size increases. For instance, the manual annotation of 400 images would necessitate over 130 min, whereas an equivalent amount of synthetic data can be generated within 10 min, achieving savings exceeding 90% in both time and labor.

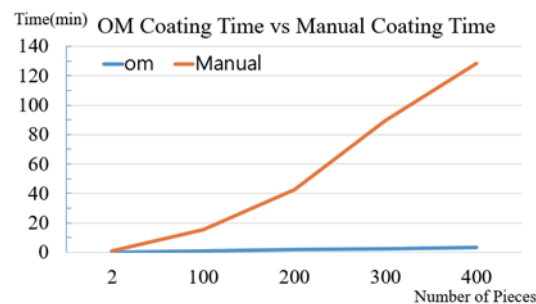


Fig. 9. (Color online) Annotation times for Omniverse-generated data and manual labeling.

Moreover, experimental findings suggest that incorporating as little as 20% real-world data into a predominantly synthetic training set is sufficient to maintain high detection accuracy. This indicates that models trained with primarily synthetic datasets generated via Omniverse retain strong generalization performance and practical applicability. Consequently, the hybrid approach not only significantly reduces dataset preparation costs but also enhances the feasibility and economic viability of applying synthetic data in real-world industrial contexts.

To systematically examine the model's convergence behavior during incremental training, an exponential decay model was applied to characterize the variation of the loss function across training epochs, as defined by

$$Loss(epoch) = ae^{\{-b \times epoch\}} + c. \quad (8)$$

In Eq. (8), $Loss(epoch)$ denotes the loss value at each training epoch, whereas a , b , and c were determined via regression analysis of empirical data, representing the initial loss magnitude, rate of loss decay, and the asymptotic stabilization level, respectively. Analysis results indicate the optimal stabilization of model convergence when the training dataset size reaches approximately 200 to 300 images.

Additionally, synthetic data generated through the Omniverse platform facilitates the effective simulation of diverse visual scenarios, including variable lighting conditions, different camera perspectives, and varying occlusion levels. Such extensive scenario coverage significantly enhances the model's ability to learn generalized visual patterns. In cases involving irregular chip shapes or blurred boundaries, the synthetic-only trained model demonstrated performance stability comparable with that of the model trained exclusively on real data, further underscoring the benefits of synthetic data in terms of data augmentation and coverage of edge-case scenarios.

4.2 Effectiveness of the hybrid data training strategy

The hybrid dataset approach not only yielded superior performance across all key recognition metrics but also demonstrated clear advantages in training efficiency and adaptability to diverse operational scenarios. During training, essential loss metrics—including Box Loss, Classification Loss, and Distribution Focal Loss—declined rapidly and stabilized within the initial 150 epochs. This observation suggests that the model achieved effective convergence with fewer training samples and iterations, reflecting reduced computational resource requirements and underscoring its practical deployment potential.

In evaluations conducted under challenging conditions, such as significant occlusion, strong specular reflections, and complex background textures, the model trained with the hybrid dataset consistently delivered the most stable detection performance. For example, under conditions involving occluded chip instances, the model trained solely on real-world data occasionally exhibited missed detections, whereas the synthetic-data-only model sometimes produced false positives owing to limitations in simulating realistic material reflectance.

Conversely, the hybrid-trained model successfully identified more than 89% of occluded chips ($Recall > 0.89$), indicating robust fault tolerance and superior visual generalization capabilities.

Moreover, the hybrid training strategy enhanced the model's recognition capability for scenarios involving blurred boundaries and atypical chip geometries, thereby reducing the risk of overfitting to specific data distributions. By combining real and synthetic data in a balanced proportion, the model effectively learned a more diverse and representative feature space. Consequently, the hybrid approach attained both high detection accuracy and efficient data preparation, rendering it particularly suitable for industrial applications where real-world data collection is challenging and rapid system deployment is critical.

In this experimental setup, the hybrid dataset comprised a total of 400 training images, consisting of 30% real data and 70% synthetic data. The results confirmed that incorporating even a relatively small proportion of real images into synthetic datasets significantly enhances overall model performance.

4.3 Incremental training performance analysis (pure real data)

To investigate how the dataset size affects the learning efficiency and convergence behavior of the YOLOv8 model, we conducted incremental training experiments using exclusively manually annotated real-image data. The incremental training procedure was carried out in five stages, employing dataset sizes of 2, 100, 200, 300, and 400 images. At each incremental training stage, the model's performance was consistently evaluated using a fixed test set, with the key detection performance metrics systematically recorded for comparative analysis. The summarized experimental results are presented in Table 1.

Stage 1 (two images)

Owing to the extremely limited training dataset, the model was only capable of learning rudimentary low-level features, resulting in notably poor detection performance. Specifically, the $mAP@0.5$ metric reached merely 0.02, $Precision$ was 0.30, $Recall$ was 0.60, and the corresponding $F1-score$ was only 0.13. Figure 10. shows minimal fluctuations in $F1-score$ and $mAP@0.5:0.95$ throughout the training process, indicating ineffective model convergence and inadequate generalization capability on the test dataset.

Stage 2 (100 images)

Upon increasing the dataset size to 100 images, the model exhibited a substantial improvement in detection capability. The $mAP@0.5$ increased significantly to 0.28, accompanied by $Precision$ and $Recall$ improvements to 0.88 and 0.92, respectively, resulting in an $F1-score$ of

Table 1
Model performance across incremental training stages using purely real annotated data.

Training stage	Sample size	$mAP@0.5$	$Precision$	$Recall$	$F1-score$
Stage 1	2	0.02	0.30	0.66	0.13
Stage 2	100	0.28	0.88	0.92	0.47
Stage 3	200	0.55	0.95	0.98	0.94
Stage 4	300	0.98	1.00	1.00	0.99
Stage 5	400	0.95	0.89	0.89	0.89

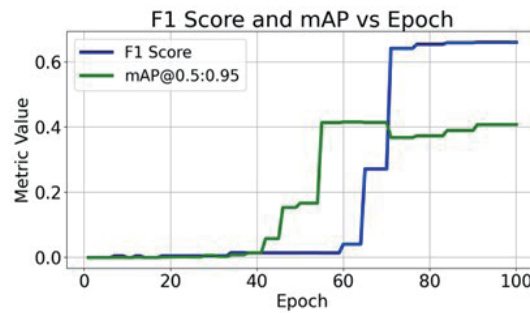


Fig. 10. (Color online) *F1-Score* and *mAP@0.5:0.95* over training epochs in Stage 1

0.47. These outcomes clearly illustrate the critical role of expanding the training data volume, thereby enabling the model to more effectively capture critical object features.

Stage 3 (200 images)

When the dataset was expanded further to 200 images, model performance demonstrated considerable improvement with *mAP@0.5* increasing to 0.55, *Precision* reaching 0.95, and *Recall* achieving 0.98, yielding a notably high *F1-score* of 0.94. At this training stage, the model demonstrated robust learning stability, indicating effective convergence and strong generalization across varied test scenarios.

Stage 4 (300 images)

The model continued to steadily enhance its detection performance throughout training, ultimately approaching its theoretical maximum performance. At this stage, the model achieved an impressive *F1-score* of 0.99 and an *mAP@0.5:0.95* of approximately 0.78. Both *Precision* and *Recall* consistently maintained high levels during later epochs, signifying successful generalization to various chip categories and challenging edge conditions. As depicted in Fig. 11, the learning curve stabilized after approximately 20 epochs, reflecting stable and reliable model convergence.

Stage 5 (400 images)

In Stage 5, a minor reduction in performance metrics was observed, with *mAP@0.5:0.95* fluctuating around 0.65 and the *F1-score* stabilizing near 0.85. This slight decline relative to earlier stages may be attributed to potential issues such as imbalanced data distribution or inconsistencies in manual annotations, potentially causing instability in training and limiting generalization performance. Figure 12. shows that the performance curve remained relatively flat beyond epoch 20, suggesting an early convergence that possibly led to suboptimal learning saturation.

From the perspective of loss function dynamics, the Box Loss, Classification Loss, and Distribution Focal Loss exhibited rapid declines, decreasing below 1.0 during the initial three training stages, and subsequently stabilizing within a range of approximately 0.8 to 1.0 during the fourth and fifth stages. This pattern of loss function reduction corresponds closely with the observed improvements in detection performance. The experimental results collectively suggest that the YOLOv8 model attains stable and reliable detection accuracy once the training sample

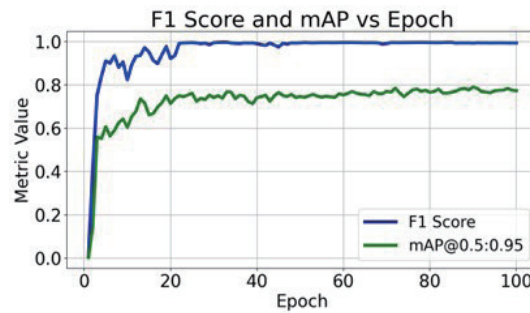


Fig. 11. (Color online) *F1*-score and *mAP*@0.5:0.95 over training epochs in Stage 4.

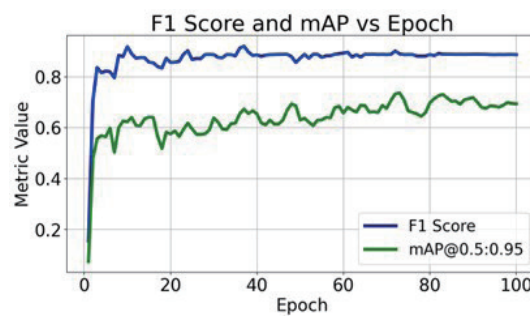


Fig. 12. (Color online) *F1*-score and *mAP*@0.5:0.95 over training epochs in Stage 5.

size reaches between 200 to 300 images. Beyond this range, incremental performance improvements become considerably smaller.

4.4 Integrated analysis and practical implications

The experimental results demonstrate that the proposed hybrid data training strategy successfully integrates the respective advantages of real and synthetic datasets. This approach maintains high detection accuracy and robust generalization capability while substantially reducing the cost and effort associated with dataset preparation, highlighting its strong practical value. Particularly in industrial applications where real-world data acquisition is challenging and rapid system deployment is crucial, this hybrid strategy represents a highly feasible, cost-effective, and efficient solution. Moreover, synthetic data generation is not limited to being merely a complementary technique; it offers considerable flexibility and scalability, enabling the effective simulation of diverse operational scenarios, standardization of data-generation workflows, and comprehensive coverage of edge cases or extreme conditions. Future research directions can focus on further enhancing the photorealism of synthetic datasets—especially in occlusion representation and material reflectance accuracy—as well as integrating automated annotation correction and advanced data augmentation techniques. These potential improvements would enhance the model’s robustness against domain shifts and significantly improve adaptability within complex industrial environments.

In summary, we demonstrated the effectiveness and feasibility of hybrid data strategies in industrial visual recognition tasks, offering a viable technological framework for the deployment of AI-driven, real-time monitoring systems within intelligent manufacturing environments.

5. Conclusions

We proposed a novel data-generation pipeline integrating NVIDIA Omniverse-based virtual simulation with 2D modeling techniques, specifically targeting the challenge of detecting chip accumulation and entanglement during CNC machining. The generated datasets were utilized to train a YOLOv8-based detection model designed for high-accuracy chip recognition. By leveraging the Omniverse platform, diverse machining scenarios were systematically simulated, enabling the rapid creation of large-scale, photorealistic synthetic images. These synthetic datasets, combined with a relatively small proportion of real-world images, formed the basis for the hybrid training approach. Experimental evaluations demonstrated that the hybrid-trained model achieved superior detection performance, obtaining an $mAP@0.5$ of 0.91 and an $F1$ -score of 0.90 on the independent test set. The trained model exhibited stable and reliable performance across a variety of challenging conditions, including variable lighting and complex background scenarios. The current system workflow effectively supports the real-time detection of chip-related anomalies. Upon identifying abnormal chip accumulation, the YOLO model immediately highlights the affected regions, thereby promptly alerting operators for corrective actions. This capability significantly enhances chip-monitoring efficiency and accuracy, reduces manual inspection workloads, and mitigates the risk of potential tool damage or equipment failures. Consequently, the proposed framework demonstrates substantial practical potential for deployment in manufacturing environments, contributing positively to on-site decision-making and maintenance scheduling processes.

Future research will be aimed at integrating advanced data augmentation techniques and automated annotation refinement procedures, further enhancing model robustness in high-noise industrial environments. Additionally, the development of tailored control interfaces based on specific operational requirements will be explored, progressing toward fully automated chip removal solutions and establishing a comprehensive, intelligent machining anomaly detection and alert system.

References

- 1 J. Wu, Y. Liu, L. Chen, L. Wang, and X. Li: Electronics **14** (2023) 1143. <https://doi.org/10.3390/electronics14061143>
- 2 Q. Chen, Q. Xiong, H. Huang, and S. Tang: Electronics **14** (2025) 505. <https://doi.org/10.3390/electronics14030505>
- 3 C. A. Akar, J. Tekli, D. Jess, M. Khoury, M. Kamradt, and M. Guthe: Proc. SIBGRAPI (2022) 150–155. <https://doi.org/10.1109/SIBGRAPI55357.2022.9991784>
- 4 N. Ahmed, I. Afyouni, H. Dabool, and Z. Al Aghbari: Front. Comput. Sci. **6** (2024) 1423129. <https://doi.org/10.3389/fcomp.2024.1423129>
- 5 L. Eversberg and J. Lambrecht: Sensors **21** (2021) 7901. <https://doi.org/10.3390/s21237901>
- 6 Y. Zhou and Z. Zhao: Pattern Recognit. **168** (2025) 111897. <https://doi.org/10.1016/j.patcog.2025.111897>
- 7 R. Rasshofer, B. Kotzian, M. Hägele, and T. Fuchs: IEEE Access **11** (2023) 108309. <https://doi.org/10.1109/INDIN51400.2023.10218035>

- 8 M. Lu, W. Sheng, Y. Zou, Y. Chen, and Z. Chen: *Measurement* **236** (2024) 115060. <https://doi.org/10.1016/j.measurement.2024.115060>
- 9 Z. Yang, X. Lu, J. Zhao, W. Wang, C. Zhang, J. Yu, S. Liu, and F. Dong: *IEEE Access* **13** (2025) 7203. <https://ieeexplore.ieee.org/document/10934994https://doi.org/10.1109/ACCESS.2023.3343610>
- 10 Y. Luo, Y. Zhang, and W. Chen: *Machines* **12** (2024) 917. <https://doi.org/10.3390/machines12120917>
- 11 J. Li and X. Kang: *Eng. Appl. Artif. Intell.* **134** (2024) 108690. <https://doi.org/10.1016/j.engappai.2024.108690>
- 12 D. Y. Jung, Y. J. Oh, and N. H. Kim: *Electronics* **13** (2024) 2598. <https://doi.org/10.3390/electronics13132598>
- 13 Z. Gui and J. Geng: *Electronics* **13** (2024) 3129. <https://doi.org/10.3390/electronics13163129>
- 14 J. Hu, F. Xiao, Q. Jin, G. Zhao, and P. Lou: *Mathematics* **11** (2023) 4588. <https://doi.org/10.3390/math11224588>
- 15 N. Zhou and T. Li: *Int. J. Comput. Intell. Syst.* **18** (2025) 81. <https://doi.org/10.1007/s44196-025-00817-4>
- 16 Q. Zhao and G. Wang: *Measurement* **242** (2024) 116305. <https://doi.org/10.1016/j.measurement.2024.116305>
- 17 E. Casas, L. Ramos, C. Romero, and F. Rivas Echeverría: *Array* **22** (2024) 100351. <https://doi.org/10.1016/j.array.2024.100351>
- 18 C. Yu and Z. Lu: *Comput. Mater. Continua* **81** (2024) 3261. <https://doi.org/10.32604/cmc.2024.056413>
- 19 T. Han, Q. Dong, X. Wang, and L. Sun: *IEEE Trans. Instrum. Meas.* **73** (2024) 1. <https://doi.org/10.1109/TIM.2024.3472791>
- 20 C. M. de Melo, J. Gratch, and P. Carnevale: *Trends Cogn. Sci.* **26** (2022) 174. <https://doi.org/10.1016/j.tics.2021.11.008>