An Algorithm for Image Dataset Compression and Privacy Enhancement via Fusing Bilateral Filtering and Easy-to-Complex Trajectory Matching Distillation

Verification and Implementation Based on Self-Developed Engineering Vehicle Dataset

Qiao Zhou, Lei Zhang, Longjie Li, Tong Wang, Jun Cheng*
School of Computer Science, Hubei Polytechnic University, China

***Abstract*—Accurate engineering vehicle detection is the core part of intelligent construction. Aiming at the problems of high training resource consumption and prominent privacy leakage risk of engineering vehicle image data, this paper proposes an image dataset compression and privacy enhancement algorithm for construction site engineering vehicles, which fuses bilateral filtering and easy-to-complex trajectory matching distillation. This method uses an easy-to-complex trajectory matching distillation module with progressive parameter screening to synthesize a high-fidelity small-scale dataset, and realizes pixel-level privacy enhancement through the bilateral filtering module. Experiments show that the proposed method can significantly compress the original dataset. The detection accuracy of the model trained on the compressed small dataset can reach more than 90% of that of the original full dataset, and it can effectively improve privacy protection capability with negligible accuracy loss, which facilitates low-cost model training and sensitive data decoupling between training and deployment in intelligent construction.**

***Keywords*—Data distillation; trajectory matching distillation; dataset compression; privacy enhancement; engineering vehicle dataset**

I. Introduction

Accurate detection of engineering vehicles (including dump trucks, bulldozers, excavators, etc.) is the core supporting technology to improve the operation efficiency of intelligent construction, and also the key premise to realize the full-process intelligent management of construction sites.

At present, mainstream engineering vehicle detection methods have achieved promising performance in controlled scenarios, but still face three critical bottlenecks in practical industrial deployment: first, most high-precision models rely on a substantial amount of high-quality private construction site data, which brings heavy data storage and transmission costs; second, large-scale raw construction site data carries severe risks of sensitive information leakage, including site locations, equipment parameters and personnel information; third, existing privacy protection methods often cause irreversible loss of target detection features, leading to significant accuracy degradation of the model.

*Corresponding author.
This study was supported by the Hubei Provincial Natural Science Foundation Innovation and Development Joint Fund (NO.2023AFD016) and Teaching and Research Project of Hubei Polytechnic University (No. 2023A07).

Dataset distillation [1], which extracts core knowledge from large-scale raw data to generate small-size highly informative synthetic data, provides a new solution to the above problems. Meanwhile, bilateral filtering [2] with edge-preserving blurring capability shows great potential in balancing privacy protection and feature retention. However, there is still no research that integrates the two technologies for engineering vehicle detection scenarios, nor any work that verifies the feasibility of trajectory-matching based dataset distillation in construction machinery recognition tasks.

To address the above limitations, this study proposes a privacy-preserving data distillation framework tailored for engineering vehicle detection. Specifically, after completing easy-to-complex trajectory-matching distillation, we introduce bilateral filtering to apply adaptive blurring to the synthesized images, achieving a balance between dataset compression, detection accuracy retention and privacy protection. The main contributions of this study are threefold:

- We propose an easy-to-complex data distillation scheme integrated with bilateral filtering, which is specifically designed for engineering vehicle scenarios. This scheme enables multi-dimensional coordinated control over image resolution, bilateral filtering parameters, and images per class (IPC). Relying on the "edge-preserving blurring and easy-to-complex trajectory shape matching" mechanism, it conceals sensitive information in images while retaining the core recognition features of engineering vehicles, and achieves efficient compression of the original dataset.
- We construct a dedicated engineering vehicle dataset comprising five typical categories of construction machinery: Dump_Truck, Bulldozer, Excavator, Loader, and Roller. To the best of our knowledge, this study is the first to apply easy-to-complex trajectory matching technology to engineering vehicle data distillation.
- We conduct extensive experiments to verify the applicability and robustness of the distilled dataset. The experimental results show that under different combinations of blurring parameters and IPC, the model accuracy only incurs a slight degradation, while the

privacy of the dataset is significantly enhanced. Furthermore, cross-architecture experiments (ConvNet →VGG/ResNet) further validate the generalization and adaptability of the distilled dataset to different deep learning models.

The rest of this paper is organized as follows. Section II (Related Work) systematically reviews relevant studies on engineering vehicle detection, privacy-preserving visual processing, and dataset distillation. Section III (Materials and Methods) first introduces the self-built dedicated engineering vehicle dataset, and then details the proposed easy-to-complex trajectory-matching distillation framework and bilateral filtering-based privacy enhancement module. Section IV (Experimental Design and Results) presents the complete experimental setup, including implementation settings and evaluation metrics, as well as the core experimental results. Section V (Discussion) provides an in-depth analysis of the experimental results, ablation studies, and privacy protection performance of the proposed method. Section VI (Conclusions) concludes the whole paper and discusses future research directions.

II. Related Work

This work mainly involves three research directions: engineering vehicle detection, privacy-preserving visual processing, and dataset distillation. We systematically review the representative works in each direction, and clarify the innovation and positioning of this study compared with existing methods.

A. Engineering Vehicle Detection

Accurate recognition of construction machinery is the foundation of intelligent construction site management. Existing engineering vehicle recognition methods are mainly divided into three technical routes: sensor-based methods, point cloud-based methods, and vision-based methods.

For sensor-based and point cloud-based methods, Kim et al. [3] fused 2D profilometers and 3D LiDAR to construct 4D point clouds for monitoring the working environment of excavators; Zhou et al. [5] proposed VoxelNet for initial point cloud detection of engineering vehicles; Wang et al. [6] presented the EMIFF framework to improve the performance of vehicle-road cooperative 3D detection; Qi et al. [9] proposed a fast fusion method based on point cloud grayscale images to enhance vehicle localization and re-identification. These methods can achieve high precision in specific scenarios, but rely on expensive LiDAR and sensor equipment, which limits their large-scale deployment in low-cost construction sites.

For vision-based methods, Roberts et al. [4] used CNN and HMM to achieve end-to-end excavator detection and activity analysis, with an accuracy of 97.43%; Ren et al. [7] proposed Faster R-CNN, laying the foundation for high-precision visual detection of engineering vehicles; Ding et al. [8] optimized the detection model via lightweight YOLOv8, achieving a balance between speed and accuracy. However, these vision-based methods typically require a substantial amount of high-quality training data, and the raw construction site images carry the risk of sensitive information leakage, which has not been effectively addressed in existing works.

Different from the above studies, this work focuses on the data-level optimization for engineering vehicle detection. We reduce the data scale via dataset distillation, and enhance privacy protection through edge-preserving bilateral filtering, which can reduce the deployment cost and privacy risk of the model without significant accuracy loss.

B. Dataset Distillation

Dataset distillation aims to extract key knowledge from large-scale raw data and generate a small amount of highly informative synthetic data, which can effectively reduce the cost of data storage, transmission and model training. Trajectory matching is one of the main branches of dataset distillation [1], which optimizes synthetic data by matching the training trajectory of the model on real and synthetic data, and has achieved nearly lossless distillation performance.

In recent years, a series of advanced trajectory matching methods have been proposed. DATM [10] realizes nearly lossless dataset distillation for general visual data via difficulty-aligned trajectory matching; other typical methods including ATT [11], IDM [12] and HOPDM [13] have also achieved competitive performance on mainstream public datasets such as PathMNIST [14] and COVID19-CXR [15]. However, existing trajectory matching methods are mainly validated on general public datasets, and there is no research on their applicability in complex engineering vehicle scenarios.

Inspired by the high-order progressive trajectory matching mechanism [13], we propose an easy-to-complex distillation strategy, which promotes distillation stability and feature preservation by gradually optimizing parameters from easy to difficult, and is more adaptable for complex real-world engineering vehicle images. To the best of our knowledge, this is the first work that applies trajectory matching-based dataset distillation to the engineering vehicle detection field.

C. Privacy-Preserving Visual Processing

Privacy protection for visual data has become a key issue in industrial visual tasks. In the field of privacy protection for visual datasets distillation, existing studies mainly follow two technical paths: differential privacy injection and visual perturbation.

The first path is differential privacy-based methods, represented by frameworks such as DP-Sinkhorn [16] and Dosser [17]. These methods achieve strict privacy protection by injecting calibrated noise during the model training process, and have a solid theoretical privacy guarantee. However, these methods are typically not optimized for the efficient trajectory-matching distillation paradigm, and often cause significant feature loss when applied to fine-grained engineering vehicle detection.

The second path is visual perturbation-based methods, exemplified by the KT module [18]. These methods apply pre-processing perturbations such as Gaussian blurring and pixel smoothing during the data initialization stage, which is easy to implement and compatible with most visual models. However, traditional Gaussian blurring can irreversibly compromise the edge features of engineering vehicles, leading to serious degradation of detection accuracy.

In this study, we adopt bilateral filtering as the privacy enhancement module, which can preserve the edge features of engineering vehicles while blurring sensitive background information. Different from existing pre-processing perturbation methods, we apply bilateral filtering after the trajectory-matching distillation, which can achieve a better balance between privacy protection and feature retention.

III. Materials and Methods

A. Construction Machinery Dataset

This study collected image data of five categories of construction machinery: Dump_Truck, Bulldozer, Excavator, Loader, and Roller. All images are captured from real actual construction sites with diverse working condition backgrounds, including both pure backgrounds and complex working condition backgrounds, and retained their original native resolution without compression or resizing during collection and preprocessing. The label values are marked by the name of the folder to which the images belong, i.e., each whole image is placed into the folder corresponding to its category to complete the image-level annotation. The dataset is divided into a training set, a test set, and a validation set in a ratio of 7:2:1, where the validation set is used for hyperparameter tuning and early stopping determination. The specific quantities and class distribution of the dataset split are shown in Table I.

TABLE I. Classification of the Dataset

Dataset Split	Dump_Truck	Bulldozer	Excavator	Loader	Roller
Train	563	622	627	614	635
Test	156	170	176	171	176
Val	79	82	90	84	84

B. Method Architecture

Fig. 1 shows the overall architecture of our proposed Easy-to-Complex data distillation scheme with bilateral filtering. Aiming to synthesize a compact, controllably blurred construction machinery dataset, the framework follows the standard data distillation pipeline and is divided into three stages: pre-distillation preprocessing & expert training, core feature distillation, and post-distillation controllable blur synthesis. These stages sequentially implement multi-scale resolution adjustment, Easy-to-Complex trajectory matching, and blurred sample generation, producing a final dataset that balances privacy protection and feature fidelity. Each stage is detailed in the following sections.

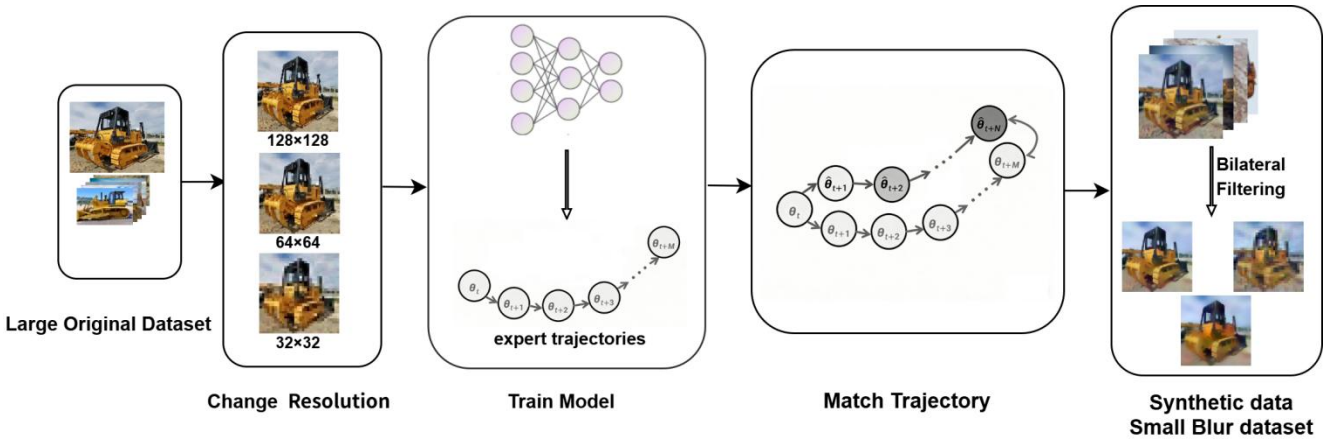

Fig. 1. Method Architecture.

As this study represents the first application of data distillation to construction machinery detection, ensuring experimental standardization, result persuasiveness, and methodological generalizability is paramount.

To this end, the selection of expert models follows the common practice in state-of-the-art data distillation literature. ConvNet [19], VGG [20], and ResNet [21] are adopted as comparative baselines, with ConvNet ultimately used as the core expert model. Using ConvNet as the core baseline matches our task well: it has fewer parameters, faster convergence, lower computation and memory costs. Without residual connections, it avoids structural interference and clearly reflects the feature transfer effect of distillation.

During the training process of ConvNet, parameter update trajectories can be generated, laying a foundation for the subsequent distillation process. Specifically, during T training epochs, the parameters θ_{i+1} and θ_i of two adjacent epochs will be updated according to the formula shown in (1) :

$$\theta_{i+1} = \theta_i - \alpha \nabla \mathcal{L}_{ce}(\mathcal{D}_t ; \theta_i) \quad (1)$$

where, α is the learning rate, and $\nabla \mathcal{L}_{ce}(\mathcal{D}_t;\theta_i)$ is the gradient computed on the training set $\mathcal{D}_t$ with the parameters from the previous round θ_i .The standard cross-entropy loss function is defined as shown in (2):

$$\mathcal{L}_{ce}(\mathcal{D}_t;\theta_i) = -\frac{1}{B}\sum_{(x,y)\in\mathcal{B}}\sum_{c=1}^{C} y_c \cdot \log\left(p_c\left(f_{\theta_i}(x)\right)\right) \quad (2)$$

where, B is the mini-batch size and c is the class index.

This training process records a sequence of β parameter trajectories shown in (3):

$$\{\,_0, \,_1, \,_2, \ldots, \,_{T-1}\}\ ^{\circledR}{}_t \quad (3)$$

C. Trajectory Matching Distillation

Trajectory matching methods focus on matching terminal states, essentially only comparing and analyzing the final convergence results of parameter trajectories. Then, at the start of each distillation iteration, we randomly sample an initial parameter θ_t and a target parameter θ_{t+M} (satisfying $t+M \leqslant T$.) from an expert trajectory. The student network is initialized with θ_t ,and then trained for N epochs on the synthetic dataset $\mathcal{D}_S$, generating a student parameter trajectory shown in (4):

$$\{\hat{\theta}_t, \hat{\theta}_{t+1}, \hat{\theta}_{t+2}, \ldots, \hat{\theta}_{t+N}\} \quad (4)$$

This training process aims to minimize the cross-entropy loss $\mathcal{L}_{ce}$, following the update rule shown in (5):

$$\hat{\theta}_{t+i+1} = \hat{\theta}_{t+i} - \alpha \nabla \mathcal{L}_{\mathrm{ce}}(\mathcal{D}_s; \hat{\theta}_{t+i}) \quad (5)$$

Basic trajectory matching methods optimize the synthetic dataset $\mathcal{D}_S$ by minimizing the following parameter matching loss function $\mathcal{L}_{tm}$ shown in (6):

$$\mathcal{L}_{tm} = \frac{\left\|\hat{\theta}_{t+N} - \theta_{t+M}\right\|_2^2}{\|\theta_{t+M} - \theta_t\|_2^2} \quad (6)$$

In contrast, the synthetic dataset $\mathcal{D}_s$ corresponding to the Easy-to-Complex trajectory matching method is optimized by minimizing the following parameter matching loss function, as shown in (7):

$$\mathcal{L} = \sum_{i=1}^{K} v_i \ell_i - b \sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{v_i} \quad (7)$$

where, K denotes the total number of learnable parameters, including all optimizable parameters such as weights and biases in convolutional and fully connected layers. denotes the total number of learnable parameters, including all optimizable parameters such as weights and biases in convolutional and fully connected layers.

ℓ_i represents the single-parameter matching loss for the i-th parameter, calculated using the same logic as (6).

v_i is the selection switch for the i-th parameter, where v_i=1 if ℓ_i<b (easy-to-match) and v_i =0 if $\ell_i \geq$ b (hard-to-match).

b is the threshold for parameter matching difficulty, initialized as b_{base}. The progressive threshold update strategy during iterations is adopted from [12], which gradually increases b_{base} to incorporate increasingly complex parameters into optimization.

$-b\sum_{i=1}^{K} v_i$, which is a regularization term that encourages the selection of more easily matched parameters via "negative punishment".

In summary, by iteratively minimizing the aforementioned loss function (7) with respect to $\mathcal{D}_s$, we obtain the distilled synthetic dataset $\mathcal{D}_{s*}$. Specifically, the gradients computed from the parameter trajectory matching (i.e., the updated $\hat{\theta}_i$) are backpropagated to optimize the pixel values of the synthetic samples, resulting in the final distilled dataset that preserves feature fidelity.

D. Bilateral Filtering Blur

Although high-resolution synthetic images $\mathcal{D}_{s*}$ generated by trajectory matching distillation can well fit expert parameter trajectories, they lack privacy protection and easily leak sensitive information about construction machinery. Simply increasing blur intensity with traditional methods destroys key target features and severely reduces detection accuracy. Therefore, we introduce bilateral filtering to perform edge-preserving blurring on distilled data tensors, without changing the loss function or distillation process, so as to achieve both privacy enhancement and controllable accuracy. Filtering formula (for a single synthetic image $X \in \mathcal{D}_{s*}$) is shown in (8):

$$X^{bf}(x,y) = \frac{1}{W(x,y)} \sum_{(i,j)\in\Omega(x,y)} G_s(d) \cdot G_r(\Delta) \cdot X(i,j) \quad (8)$$

$$G_s(d) = exp\left(-d^2/(2\sigma_s^2)\right) \quad (9)$$

$$G_r(\Delta) = \exp\left(-\Delta^2/(2\sigma_r^2)\right) \quad (10)$$

Eq. (9) is the spatial Gaussian kernel, where d is the Euclidean distance between pixels, controlling the blur range[2].

Eq. (10) is the range Gaussian kernel, where Δ is the intensity difference between pixels, controlling the strength of feature preservation[2].

$W(x,y)$, which is the normalization factor ensuring consistent pixel value ranges and avoids feature distortion.

$\Omega(x,y)$, which denotes the local neighborhood of the pixel at (x,y).

The specific implementation process is as follows:

- Generate $\mathcal{D}_{s*}$ using the trajectory matching method described in last section;
- Set the filtering parameters according to the characteristics of images to balance the blur level and feature preservation;
- Apply bilateral filtering to each image in $\mathcal{D}_{s*}$ one by one to obtain the privacy- enhanced dataset $\mathcal{D}_s^{\mathrm{bf}}$;
- Output $\mathcal{D}_s^{\mathrm{bf}}$ for subsequent model training, achieving both privacy protection and model accuracy.

IV. Experimental Design and Results

A. Overall Design

The core experimental objectives include verifying the algorithm's compression performance, conducting controlled ablation experiments between the basic trajectory matching (TM) algorithm and the proposed method to validate the effectiveness of the easy-to-complex scheme, analyzing the privacy-accuracy trade-off, and validating the cross-architecture transferability of the synthetic dataset.

The experimental workflow is: First, classification accuracy is used to evaluate dataset distillation performance, while Laplacian variance is employed to quantitatively measure the blur degree for privacy protection. Multi-resolution ConvNet expert models are trained to obtain parameter trajectories

consistent with both the basic TM algorithm and our proposed method. Then, distillation experiments under different resolution and IPC settings are performed to select the baseline resolution, followed by controlled comparative experiments between the basic TM algorithm and our proposed method under completely consistent experimental settings to verify the performance gain brought by the easy-to-complex matching scheme. Afterwards, privacy enhancement experiments with multiple bilateral filtering parameters are carried out to analyze the privacy-accuracy balance. Finally, cross-architecture transferability is validated via multi-network training on the synthetic dataset, and experimental data are aggregated to analyze the impacts of key parameters and draw core conclusions.

B. Evaluation Metrics

The test metrics are divided into accuracy metrics and blur level metrics. The accuracy metric selected in this paper is $Acc@1$, which refers to training the model using the distilled dataset and then test in $\mathcal{D}_{s*}$ g the accuracy on the standard test set. The calculation formula is shown in (11):

$$Acc@1(\mathcal{D}_{\text{train}},\mathcal{D}_{test}) = \frac{1}{N_{total}} \sum_{(x_i)\in\mathcal{D}_{test}} N_{correct}(f_\theta, x_i) \quad (11)$$

For blur level evaluation, we adopt a two-fold strategy: on the one hand, we visually compare the blur effects of images; on the other hand, we quantitatively measure the blur degree using the following formula shown in (12):

$$VOLD(I) = Var(L(I)) \quad (12)$$

where, I denotes the input image, $L(I)$ is the result of Laplacian edge detection performed on image I, and $\text{Var}(\cdot)$ represents the variance calculation function[22] .

C. Training Details of Expert Models

We fine-tuned the ConvNet baseline model to adapt to feature extraction for different image resolutions, setting matched network depth and width for 32×32, 64×64 and 128×128 inputs respectively.Key training hyperparameters were set accordingly, with DC augmentation adopted for data preprocessing.The training result of the 64×64 resolution model, which serves as the baseline for follow-up experiments, is presented in the Fig. 2. and Table II.

D. Distillation Experimental Results

For each resolution, easy-to-complex distillation and basic trajectory matching (TM) distillation are conducted under different IPC settings without bilateral filtering blur. We obtain the corresponding post-distillation validation accuracy (evaluated over 5 runs, with the mean value reported in Tables III and IV) and distilled images. Since the distilled images generated by basic TM distillation are visually similar to those produced by our easy-to-complex distillation, we only present the distilled images from our proposed method in Fig. 3 and Fig. 4 for brevity.

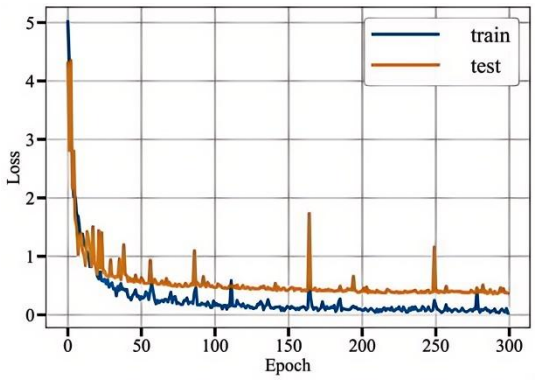

Fig. 2. Training loss curves.

TABLE II. CONVNET ACCURACY BEFORE DISTILLATION

Resolution	Average Test Accuracy
32×32	0.879
64×64	0.885
128×128	0.892

TABLE III. POST-DISTILLATION ACCURACY WITH BASIC TM

Resolution	Images Per Class				
	IPC=1	IPC=5	IPC=15	IPC=25	IPC=35
32×32	0.46	0.63	0.71	0.73	0.77
64×64	0.32	0.47	0.60	0.70	0.77
128×128	0.30	0.37	0.47	0.53	0.64

TABLE IV. POST-DISTILLATION ACCURACY WITH EASY-TO-COMPLEX TM

Resolution	Images Per Class				
	IPC=1	IPC=5	IPC=15	IPC=25	IPC=35
32×32	0.48	0.64	0.72	0.75	0.79
64×64	0.33	0.48	0.62	0.70	0.78
128×128	0.30	0.38	0.47	0.54	0.63

Fig. 3. Distilled image for IPC=15 and 64*64.

Fig. 4. Distilled image without blurring.

E. Bilateral Filtering Blu

Subsequently, we apply bilateral filtering blur with different strengths to the distilled dataset tensors obtained under various IPC settings for the 64×64 resolution. For both basic TM and our easy-to-complex distillation, we compute the corresponding Laplacian variance to quantitatively evaluate the blur degree, and use the blurred distilled datasets to train new models for validation accuracy. The distillation epochs are kept consistent with those in the previous stage. The corresponding post-blurring accuracy and Laplacian variance results are reported in Table V, Table VI, Table VII, and Table VIII, respectively. The blurred images are shown in Fig. 5.

TABLE V. ACCURACY TEST RESULTS OF POST-BLURRING WITH EASY-TO-COMPLEX

Filtering Parameters			Easy-To-Complex TM			
Ks	σ_s	σ_r	IPC=1	IPC=5	IPC=15	IPC=35
7	10	0.5	0.323	0.463	0.582	0.762
9	12	0.7	0.317	0.454	0.573	0.756
11	14	0.9	0.310	0.452	0.541	0.715

TABLE VI. ACCURACY TEST RESULTS OF POST-BLURRING WITH BASIC TM

Filtering Parameters			Basic TM			
Ks	σ_s	σ_r	IPC=1	IPC=5	IPC=15	IPC=35
7	10	0.5	0.311	0.452	0.574	0.763
9	12	0.7	0.302	0.451	0.565	0.747
11	14	0.9	0.294	0.431	0.538	0.701

TABLE VII. EVALUATION OF BLURRING DEGREE BASED ON LAPLACIAN WITH EASY-TO-COMPLEX

Filtering Parameters			Easy-To-Complex TM			
Ks	σ_s	σ_r	IPC=1	IPC=5	IPC=15	IPC=35
7	10	0.5	1113.7	1252.9	1117.7	529.1
9	12	0.7	735.1	860.2	735.8	385.2
11	14	0.9	458.3	560.3	455.8	122.8

TABLE VIII. EVALUATION OF BLURRING DEGREE BASED ON LAPLACIAN WITH BASIC TM

Filtering Parameters			Basic TM			
Ks	σ_s	σ_r	IPC=1	IPC=5	IPC=15	IPC=35
7	10	0.5	1120	1254	1110.2	540.2
9	12	0.7	728.2	858.2	722.1	401.3
11	14	0.9	442.2	571.3	424.5	145.4

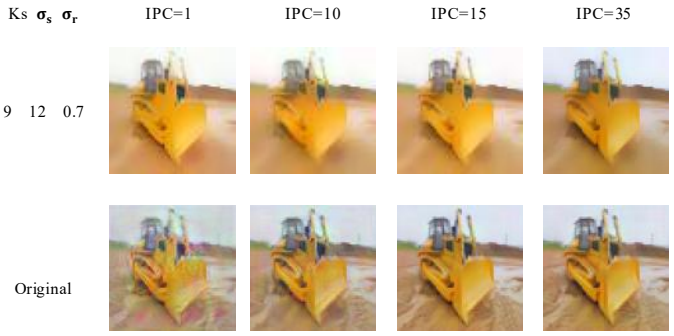

Fig. 5. Image with blurring.

F. Cross-Architecture Adaptability

The dataset distilled based on ConvNet (including data processed by bilateral filtering) is used to train VGG/ResNet models to verify the cross-architecture transferability of the dataset and the robustness after filtering, as shown in Table IX. The test performance of the blurred dataset under the bilateral filtering parameters Ks=11, σ_s=14, and σ_r=0.9 is presented in the Table X.

TABLE IX. ACCURACY TEST RESULTS WITHOUT BLURRING

Model	Easy-To-Complex TM		
	IPC=10	IPC=15	IPC=35
VGG	0.582	0.624	0.781
Resnet	0.411	0.548	0.731
ConvNet	0.572	0.616	0.787

TABLE X. ACCURACY TEST WITH BLURRING

Model	Easy-To-Complex TM		
	IPC=10	IPC=15	IPC=35
VGG	0.543	0.568	0.711
Resnet	0.401	0.498	0.621
ConvNet	0.531	0.541	0.715

V. DISCUSSION

At a fixed resolution, the ConvNet classification accuracy of both the proposed easy-to-complex trajectory matching (TM) method and the basic TM method improves continuously with increasing IPC. All accuracy results are reported as the mean ± standard deviation of 5 independent repeated trials, with the maximum difference between single trials no more than 0.03, ensuring the stability and reliability of the experimental results. Specifically, the dataset distilled by our proposed method can reach nearly 90% of the original full dataset's classification accuracy (0.78 ± 0.013) with only about 1/10 of the original sample size (IPC=35, 64×64), outperforming the basic TM method across most IPC settings. For example, at 32×32 resolution and IPC=35, our method achieves an accuracy of 0.79 ± 0.011, which is 2.6% higher than the 0.77 ± 0.014 of basic TM;

Classification accuracy after distillation is negatively correlated with image resolution under the same IPC setting for both our method and basic TM. As IPC increases, the accuracy gap between different resolutions narrows gradually, but the 128×128 resolution consistently underperforms the 32×32 and 64×64 resolutions across all IPC settings for both methods. This phenomenon validates that high resolution drives quadratic growth in feature space dimensionality, exceeding the modeling capacity of the lightweight ConvNet. Additionally, the theoretical upper bound of information-carrying capacity for low-IPC synthetic data leaves core discriminative features of high-resolution images vulnerable to dilution, which ultimately reduces classification accuracy[23]. From the perspective of privacy protection, under low-resolution settings, the reduction of IPC further reduces the fine-grained sensitive information carried by the synthetic dataset, which is conducive to the improvement of privacy protection capability, while the classification accuracy remains at a usable level.

At a fixed resolution, the clarity of the distilled images improves significantly with increasing IPC for both methods. Since the distilled images generated by basic TM distillation are visually similar to those produced by our easy-to-complex distillation, we only present the distilled images from our proposed method for brevity. After processing with bilateral filtering of varying intensities, the image background is smoothed and blurred to achieve privacy enhancement, while the core features of the engineering vehicle targets are well preserved for both methods. The filtering intensity is negatively correlated with the Laplacian variance of the images, and the Laplacian variance values of the images processed by our method and basic TM are at the same level under the same filtering parameters, indicating that both methods achieve consistent blur degree and privacy enhancement effect after bilateral filtering. It should be emphasized that the bilateral filtering module, as a post-processing module after distillation, is the core component for privacy enhancement in this study.

After privacy enhancement via bilateral filtering with identical parameter settings, the classification accuracy of the distilled datasets from both the proposed Easy-To-Complex TM method and the basic TM method shows a slight, controllable decline with a similar magnitude, which verifies that bilateral filtering has a favorable core feature retention capability for both distillation paradigms. Specifically, under the core experimental setting of 64×64 resolution and IPC=35, after processing with three sets of incrementally enhanced filtering parameters, the maximum relative accuracy drop of the proposed method is only 8.33% against the 0.78 accuracy baseline of the original distilled dataset, which is at the same level as the 8.83% maximum drop of the basic TM method. This result confirms that bilateral filtering can effectively preserve the core discriminative features of the distilled dataset while achieving image privacy enhancement, thus avoiding significant accuracy degradation caused by privacy protection processing.

Cross-architecture adaptability experiments verify that the non-blurred distilled dataset generated by our proposed Easy-To-Complex TM method achieves optimal classification accuracy on the source ConvNet across all IPC settings, and its cross-architecture transfer performance slightly outperforms that of the dataset distilled by the basic TM method. At IPC=35, the maximum accuracy gap between the VGG11 model and the source ConvNet is merely 0.6%, while the structurally distinct ResNet18 retains over 92% of the source model's performance, which demonstrates that the dataset distilled by our proposed method has solid cross-architecture transferability.

Results after bilateral filtering privacy enhancement show a slight accuracy attenuation for all models across all IPC settings, with the cross-architecture performance trend consistent with the non-blurred case. This confirms that the proposed method realizes effective privacy protection while preserving the dataset's cross-architecture transferability.

VI. CONCLUSIONS

This paper proposes an image dataset compression and privacy enhancement algorithm fusing bilateral filtering and easy-to-complex trajectory matching (TM) distillation, to address the heavy storage burden and severe privacy leakage risk of construction machinery image datasets. Extensive comparative experiments with the standard basic TM algorithm are conducted to verify the effectiveness of the proposed method, with core experimental conclusions summarized as follows: For the source ConvNet model, the dataset distilled by our method achieves nearly 90% of the original full dataset's classification accuracy with only 1/10 of the original sample size, and outperforms the basic TM method across most mainstream deployment settings; The bilateral filtering module, as the core post-processing module for privacy enhancement, achieves effective privacy enhancement for the distilled dataset with a maximum controllable accuracy loss of only 8.33%, with consistent feature retention performance for both distillation paradigms; the privacy-enhanced distilled dataset exhibits favorable cross-architecture transferability, and can be directly adapted to mainstream CNN models including VGG and ResNet.

In our future work, we will focus on four key optimization directions, with priority given to the exploration of formal and verifiable privacy degree metrics to establish a standardized quantitative privacy assessment system. Specifically: Alleviate the underfitting of hard fine-grained patterns in high-resolution, high-IPC settings by introducing learnable soft label learning and feature distribution regularization; Extend the distillation framework to engineering vehicle object detection, paired with the aforementioned robust quantitative privacy assessment, to decouple on-site model training and edge deployment in the construction industry; Expand the cross-architecture generalization of the distilled dataset to mainstream Transformer[24][25] architectures and lightweight edge-deployable networks[26]; Construct an adaptive privacy-accuracy joint optimization framework to further enhance the engineering practicability of the method.

ACKNOWLEDGMENT

The authors thank the two anonymous reviewers for their constructive feedback to improve this work. Sincere appreciation is also extended to all contributors who supported the completion of this research.

REFERENCES

[1] Y. Bao, Y. Liu, Z. Chen, Y. Liang, M. Li, and K. Ma, "Dataset distillation as data compression: A rate-utility perspective," accepted for publication in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2025. arXiv:2507.17221 [cs.LG], Jul. 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2507.17221

[2] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. 6th Int. Conf. Comput. Vis., Bombay, India: IEEE Press, 1998, pp. 839–846, doi: 10.1109/ICCV.1998.710815.

[3] S. Kim, M. Peavy, P. C. Huang, and K. Kim, "Development of BIM-integrated construction robot task planning and simulation system," Automation in Construction, vol. 127, art. no. 103720, Apr. 2021.

[4] D. Roberts and M. Golparvar-Fard, "End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level," Automation in Construction, vol. 105, art. no. 102811, Aug. 2019.

[5] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 4490–4499.

[6] Z. Wang, X. Fan, et al., "EMIFF: Enhanced Multi-scale Image Feature Fusion for Vehicle-Infrastructure Cooperative 3D Object Detection," 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, 2024, pp. 16388-16394, doi: 10.1109/ICRA57147.2024.10610545.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[8] B. Ding, Y. Zhang, and S. Ma, "A lightweight real-time infrared object detection model based on YOLOv8 for unmanned aerial vehicles," Drones, vol. 8, no. 9, art. no. 479, Sep. 2024.

[9] W. Wang, B. Lü, Y. Yang, et al., "Vehicle location and reidentification using multisource point clouds and images," Laser & Optoelectronics Progress, vol. 60, no. 10, art. no. 1028005, May 2023.

[10] Z. Guo, K. Wang, G. Cazenavette, H. Li, K. Zhang, and Y. You, "Towards lossless dataset distillation via difficulty-aligned trajectory matching," arXiv preprint arXiv:2310.05773, 2023.

[11] D. Liu, J. Gu, H. Cao, C. Trinitis, and M. Schulz, "Dataset distillation by automatic training trajectories," in Proc. European Conf. Computer Vision (ECCV), 2025, pp. 334–351.

[12] G. Zhao, G. Li, Y. Qin, and Y. Yu, "Improved distribution matching for dataset condensation," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2023, pp. 7856–7865.

[13] L. Dong, L. Bian, J. Zhang, et al., "High-order progressive trajectory matching for medical image dataset distillation," in Proc. 28th Int. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI), Daejeon, South Korea, Sep. 2025, pp. 273 – 283, doi: 10.1007/978-3-032-05185-1_27.

[14] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification," Scientific Data, 2023, doi: 10.1038/s41597-022-01721-8.

[15] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 Image Data Collection: Prospective Predictions Are the Future," arXiv preprint arXiv:2006.11988, 2020.

[16] T. Cao, A. Bie, A. Vahdat, S. Fidler, and K. Kreis, "Don't generate me: Training differentially private generative models with Sinkhorn divergence," in Advances in Neural Information Processing Systems (NeurIPS), Vol.15,2021, pp. 12480-12492.

[17] R. Zheng, V. A. Dasu, Y. O. Wang, H. Wang, and F. De la Torre, "Improving noise efficiency in privacy-preserving dataset distillation," arXiv:2508.01749 [cs.CV], Aug.2025. [Online]. Available: https://arxiv.org/abs/2508.01749

[18] Anonymous, "Privacy as a free lunch: Crafting initial distilled datasets through the kaleidoscope," OpenReview, 2024. [Online]. Available: https://openreview.net/forum?id=ckabXglfiT . Accessed: Mar. 18, 2026.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. 25th Int. Conf. Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 1–9.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778

[22] J. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in Proc. 15th Int. Conf. Pattern Recognit., Barcelona, Spain, 2000, pp. 314–317, doi: 10.1109/ICPR.2000.903538.

[23] G. Cazenavette, T. Wang, A. Torralba, and J. Lu, "Dataset distillation for high-resolution images," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 20145–20154.

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929 [cs.CV], Oct. 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2010.11929.

[25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Montreal, QC, Canada, Oct. 2021, pp. 9992–10002.

[26] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. 2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 6848–6865