IIEEJ Transactions on Image Electronics and Visual Computing

Special Issue on Image Electronics Technologies Related to AI

Vol. 14, No. 1 2026

The Institute of Image Electronics Engineers of Japan

Editorial Committee of IIEEJ

IIEEJ Office
Eishu ODAKE
Rieko FUKUSHIMA
Kyoko HONDA

Contact Information

The Institute of Image Electronics Engineers of Japan (IIEEJ)
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel : +81-3-5615-2893　Fax : +81-3-5615-2894
E-mail : hensyu@iieej.org
http://www.iieej.org/ (in Japanese)
http://www.iieej.org/en/ (in English)
http://www.facebook.com/IIEEJ (in Japanese)
http://www.facebook.com/IIEEJ.E (in English)

IIEEJ Transactions on
Image Electronics and Visual Computing
Vol.14 No.1 June 2026

CONTENTS

Special Issue on Image Electronics Technologies Related to AI

Announcements

Guide for Authors

Published two times a year by the Institute of Image Electronics Engineers of Japan(IIEEJ)
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan
Tel:+81-5615-2893 Fax:+81-5615-2894 E-mail:hensyu@iieej.org http://www.iieej.org/

Preface to the Special Issue on
Image Electronics Technologies Related to AI

Editor: Koyo NITTA

The rapid advancement of artificial intelligence (AI) technologies in recent years has brought remarkable progress to a wide range of image electronics and related technologies, including image and video processing, recognition, and generation. These technologies have been widely integrated into various applications, such as autonomous driving, medical image diagnosis, face recognition, industrial anomaly detection, surveillance systems, and drones. At the same time, addressing social issues arising from the misuse of AI technologies, such as the generation of deepfakes and fake images, has become a critical challenge. There is no doubt that AI technologies will continue to play an increasingly vital role in the future of image electronics.

For this special issue, we called for a broad range of contributions, including research papers on foundational AI advances, the evolution of image electronics driven by AI, evaluations of practical systems, case studies on real-world applications, and comprehensive surveys. Following a rigorous review process, four papers — two in Japanese and two in English — were accepted for publication. The two Japanese papers have already been published in the April 2026 issue, and this issue proudly features the two accepted English papers.

The first paper, entitled “Intrinsic-Dimensionality-Based Layer Selection for Training-Free Anomaly Detection,” proposes a method for automatically optimizing the feature-extraction layer in anomaly detection. Utilizing intrinsic dimensionality estimation, the method requires only normal data for training. The proposed approach is compatible with both CNN and ViT backbones, achieving stable and effective anomaly detection performance without the need for abnormal samples.

The second paper, entitled “A Two-Stage Structured-Light Neural Network for Surface Dent Detection and Geometric Measurement of Car Body Shells,” introduces a two-stage neural network-based approach for detecting and measuring dents on automotive panels. By combining structured-light imaging with deep learning, the proposed system enables accurate dent localization and geometric measurement without requiring full 3D reconstruction, demonstrating robust surface inspection performance even on highly reflective surfaces.

AI technologies continue to drive innovative breakthroughs across numerous domains, including image processing and visual computing, with further advances and broader applications anticipated in the future. We are pleased to present some of the latest research in this rapidly evolving area through this special issue. This journal remains highly receptive to AI-related research, system developments, and practical applications as regular submissions, and we look forward to receiving submissions reporting new findings and technological advances.

Finally, I would like to express my sincere gratitude to all the reviewers who generously evaluated the submitted papers despite their busy schedules, as well as to everyone who contributed to the editing and publication of this special issue.

＊Designated Editors to the Special Issue
Koyo NITTA（The University of Aizu）, Tomohiko MUKAI（Tokyo Metropolitan University）

IIEEJ **Paper** *-- Special Issue on Image Electronics Technologies Related to AI --*

Intrinsic-Dimensionality-Based Layer Selection for Training-Free Anomaly Detection

Masahiro TSUCHIYA†††(*Student Member*), Tsubasa HIRAKAWA††, Takayoshi YAMASHITA††, Hironobu FUJIYOSHI††

†OMRON Corporation, ††Chubu University

<**Summary**> In real-world visual anomaly detection, abnormal samples are difficult to collect, making normal-only methods mainstream. Among them, deep feature embedding approaches, which model the distribution of normal features extracted from intermediate layers of pretrained networks and treat deviations as anomalies, are practical, requiring no extra training and enabling lightweight inference. However, most existing methods fix the feature-extraction layer, and the impact of layer selection remains insufficiently understood.

This study addresses this issue by estimating the intrinsic dimensionality of layer-wise output feature maps from normal data and selecting the optimal feature-extraction layer through a one-time, task-wise calibration process. The proposed method, applicable to both CNN and ViT backbones, determines the most suitable layer using only normal samples, without requiring abnormal data or inference-time adaptation.

Experiments on MVTec-AD, VisA, and BTAD demonstrate that the proposed method outperforms the average performance of fixed-layer configurations and achieves accuracy comparable to the best manually selected layer despite using no abnormal samples. These results highlight the practical utility of the proposed approach for stable, training-free anomaly detection across diverse datasets and backbone architectures.

Keywords: anomaly detection, feature selection, intrinsic dimension, few-shot, layer optimization

1. Introduction

Anomaly Detection (AD) aims to determine whether an input image contains anomalies and, if necessary, localize them at the pixel level. In industrial inspection, abnormal samples are rare and costly to collect or annotate, motivating unsupervised approaches that learn exclusively from normal data.

Normal-data-based methods are broadly categorized into three groups. (i) Reconstruction-based methods learn to reproduce normal images and detect anomalies from reconstruction errors (e.g., AutoEncoder[1], VAE[2]). (ii) Self-supervised methods generate pseudo anomalies for training (e.g., CutPaste[3], DRAEM[4]), but may introduce domain gaps from synthetic artifacts. (iii) Deep Feature Embedding (DFE) methods model the distribution of pretrained feature representations without additional training, enabling lightweight inference suitable for industrial deployment.

The proposed workflow is summarized in **Fig. 1**. A critical factor in DFE performance is the choice of feature extraction layer. Although most studies adopt fixed intermediate layers of pretrained backbones such as ResNet or ViT, recent work indicates that the optimal layer varies depending on dataset and anomaly characteristics. This suggests that layer selection strongly affects detection accuracy but remains underexplored compared with scoring or memory-bank design.

This paper addresses the following question: Can an optimal feature-extraction layer be identified automatically using only normal data, without abnormal labels or additional training? Selecting a single appropriate layer can reduce computational cost, improve interpretability, and avoid redundancy often introduced by multi-layer fusion, while maintaining competitive accuracy.

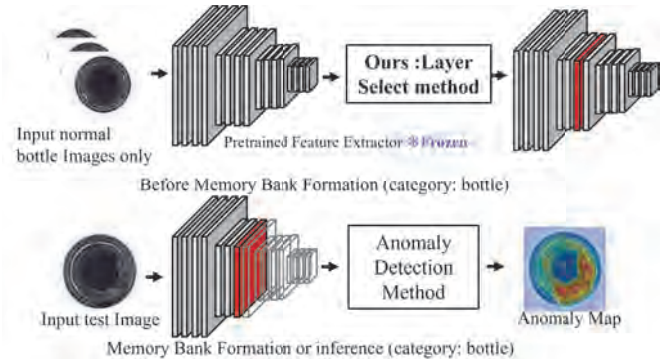

Fig. 1 Overview of the proposed layer selection: ID is estimated from normal data, an optimal layer is calibrated once per category, and reused for efficient, training-free anomaly detection

The main contributions are summarized as follows. Automatic Layer Selection: a training-free framework that identifies optimal feature extraction layers using only normal data. Intrinsic-Dimensionality Criterion: an ID-based metric to filter unstable or overfitted layers and improve generalization. Wide Applicability: compatibility with various anomaly detection algorithms and backbone architectures (CNN and ViT). Empirical Effectiveness:

experiments on MVTec-AD[5], VisA[6], and BTAD[7]show that the proposed method achieves on average 97.7% of the best-layer performance while surpassing layer-average baselines and improving stability without using abnormal data. Robust Generalization: stable performance even in few-shot scenarios with limited normal samples. Overall, this work provides a practical and theoretically grounded approach to layer selection for unsupervised anomaly detection.

2. Related Work

2.1 Feature extractor, layer selection, and representational geometry

Recent studies have shown that the performance of feature-embedding–based anomaly detection depends strongly on the backbone, intermediate layer, and input resolution. Heckler et al.[8]systematically analyzed these factors and reported that the receptive-field size, jointly determined by image resolution and network depth, greatly influences detection accuracy. Although larger images sometimes improve performance, they also increase computational cost. In contrast, adapting the feature-extraction layer can achieve similar accuracy with much lower overhead. Since input resolution is often constrained by acquisition hardware or runtime limits, intrinsic-dimensionality-based layer selection provides a practical and efficient means of task-specific adaptation, especially when abnormal data are unavailable. These layer-dependent behaviors can be interpreted through the manifold hypothesis, which posits that natural images and deep representations lie on low-dimensional manifolds in high-dimensional space[9]. Deep networks gradually reshape these manifolds to enhance separability between classes or conditions. Cohen et al.[10]and Stephenson et al.[11]showed that reductions in manifold radius and intrinsic dimension correlate with improved separability, supporting a geometric view of feature compression. In the context of anomaly detection, anomalies can thus be regarded as deviations from a compact normal manifold.

2.2 Intrinsic dimensionality and our positioning

Intrinsic dimensionality (ID) quantifies the effective degrees of freedom of a feature manifold and serves as a measure of its compactness and stability. Ansuini et al.[12]observed a "hunchback" ID trend—rising in early layers and decreasing toward the output—while Doimo et al.[13]and Kaufman et al.[14]linked inter-layer ID changes to generalization ability. The Two-Nearest-Neighbor (TwoNN) estimator[15]enables efficient, density-insensitive estimation using only neighbor-distance ratios. Compared with alternative metrics such as PCA variance or the Davies–Bouldin (DB)[16]score, ID is non-linear, non-parametric, and robust under few-shot or non-uniform sampling.

Building on these findings, this study employs the ID profile of normal feature manifolds as a quantitative criterion for automatically selecting the most stable and generalizable layer for anomaly detection. Because intrinsic dimensionality reflects both representational compactness and separability, it provides a principled basis for automatically identifying the optimal feature-extraction layer using only normal data.

3. Proposed Method

We propose an intrinsic-dimensionality-based framework to automatically select an optimal feature-extraction layer from normal manifolds. The method consists of three stages: (i) estimating the intrinsic dimensionality (ID) of each layer's feature representations, (ii) excluding layers affected by overfitting or domain mismatch, and (iii) selecting the layer exhibiting the strongest inter-layer ID compression. An overview is shown in **Fig. 2**.

Algorithm 1 details layer-wise ID estimation, and Algorithm 2 summarizes the selection rule.

3.1 Intrinsic dimensionality (ID) estimation

Estimating ID in high-dimensional feature spaces is costly; we adopt the TwoNN estimator[15]for simplicity and robustness. Other estimators include PHdim[17], CRLM[14], GMST[18], and MLE[19], but TwoNN offers a favorable accuracy–efficiency trade-off.

TwoNN assumes locally uniform sampling and estimates intrinsic dimensionality from the ratio between the first and second nearest-neighbor distances:

$$\mu_i = \frac{r_i^{(2)}}{r_i^{(1)}}, \qquad \widehat{d}^{(l)} = \left(\mathrm{median}_i\left(\log \mu_i\right)\right)^{-1}. \quad (1)$$

We adopt median aggregation after outlier clipping for robustness, as detailed in Sec. 4.2.

3.2 Filtering layers affected by overfitting

With pretrained classifiers, ID typically decreases toward deeper layers under domain-aligned settings due to semantic compression.

However, under domain mismatch, this monotonic trend may break down, and ID can increase near the deepest layers, indicating overfitting or domain-specific memorization[12]. To mitigate this effect, we exclude the last 20% of layers from the candidate set $\mathcal{L}_{\text{eff}}$. This relative cutoff normalizes depth across architectures and is validated in Chapter 5.

3.3 Layer selection via inter-layer ID compression

Since feature manifolds reorganize across layers[13], we quantify transitions by inter-layer ID change:

$$\Delta \mathrm{ID}_l = \mathrm{ID}_{l-1} - \mathrm{ID}_l. \quad (2)$$

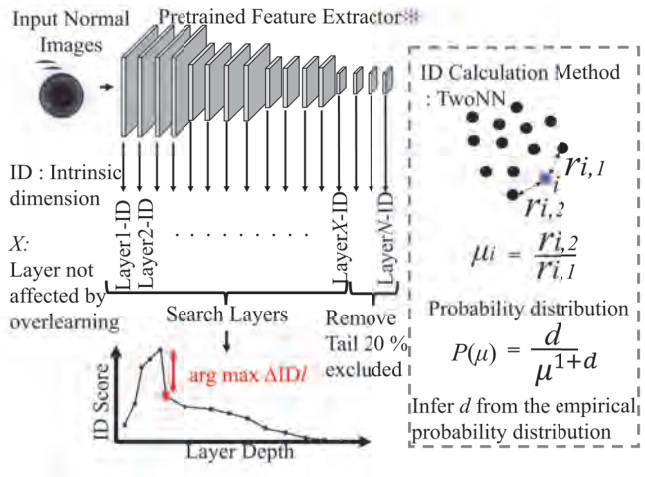

Fig. 2 Overview of the proposed method. TwoNN estimates per-layer ID from normal data; tail layers are excluded; the largest ID drop selects the layer

Algorithm 1 TwoNN-based Intrinsic Dimension per Layer (clip + median version)

Require: Backbone layers (depth-ordered) $\mathcal{L} = \{f^{(l)}\}_{l=1}^{|\mathcal{L}|}$, normal set $\mathcal{D} = \{x_j\}_{j=1}^{N}$, point budget $P_{\max}$, tail ratio ρ=0.2
Ensure: $\{\widehat{d}^{(l)}\}_{l\in\mathcal{L}_{\text{eff}}}$
$k_{\text{tail}} \leftarrow \lceil \rho \cdot |\mathcal{L}| \rceil$
$\mathcal{L}_{\text{eff}} \leftarrow \{\, l \in \mathcal{L} \mid l \le |\mathcal{L}| - k_{\text{tail}} \,\}$ ▷ exclude last 20% layers
for $l \in \mathcal{L}_{\text{eff}}$ **do**
  $S^{(l)} \leftarrow \emptyset$
  for $x \in \mathcal{D}$ **do**
    $Z \leftarrow f^{(l)}(x)$ ▷ (C_l, H_l, W_l) or (tokens, C_l)
    $V \leftarrow \text{VECTORIZER}(Z)$ ▷ $(H_l W_l, C_l)$ or (tokens, C_l)
    $S^{(l)} \leftarrow S^{(l)} \cup \text{SUBSAMPLE}\big(V,\ P_{\max}/|\mathcal{D}|\big)$
  $X \leftarrow \text{STACK\_ROWS}(S^{(l)})$ ▷ (M_l, C_l)
  $X \leftarrow \text{ZSCORE\_CHANNELWISE}(X)$
  $(D_1, D_2) \leftarrow \text{NN\_Dist}(X, 2)$
  $\mu_i \leftarrow \text{Clip}\big(\frac{D_{2,i}}{D_{1,i}+\varepsilon};\ \text{lo} = 1{+}10^{-6},\ \text{hi} = Q_{0.98}\big)$ ▷ ε: numerical stability constant, $Q_{0.98}$: 98th percentile clipping; see Sec. 4.2
  $\widehat{d}^{(l)} \leftarrow \Big(\text{median}_i(\log \mu_i)\Big)^{-1}$ ▷ median-based TwoNN
return $\{\widehat{d}^{(l)}\}_{l\in\mathcal{L}_{\text{eff}}}$

Algorithm 2 Layer Selection by Maximum Inter-layer ID Compression (with tail exclusion)

Require: $\{\widehat{d}^{(l)}\}_{l\in\mathcal{L}_{\text{eff}}}$ from Alg. 1, optional stability $\{\text{CV}^{(l)}\}$, $\lambda \ge 0$
Ensure: Selected layer $l^\star$
for $l \in \mathcal{L}_{\text{eff}}$ with predecessor $(l{-}1) \in \mathcal{L}_{\text{eff}}$ **do**
  $\Delta\widehat{d}^{(l)} \leftarrow \widehat{d}^{(l-1)} - \widehat{d}^{(l)}$ ▷ positive = stronger compression
  if stability is available **then**
    $S^{(l)} \leftarrow \Delta\widehat{d}^{(l)} - \lambda\,\text{CV}^{(l)}$ ▷ prefer large compression with stable ID
  else
    $S^{(l)} \leftarrow \Delta\widehat{d}^{(l)}$
$l^\star \leftarrow \arg\max_{l\in\mathcal{L}_{\text{eff}},\, l>\min(\mathcal{L}_{\text{eff}})} S^{(l)}$
return $l^\star$

We select the layer

$$l^\star = \arg\max_l \Delta\text{ID}_l, \tag{3}$$

so that larger positive values correspond to stronger intrinsic compression (i.e., elimination of redundant degrees of freedom). This yields an interpretable single-layer feature extractor without requiring abnormal samples or additional training.

Once $l^\star$ is determined from normal data, it remains fixed during inference, so ID estimation is performed only once per task.

The procedure is summarized in Algorithm 2.

3.4 Theoretical justification under the manifold assumption

Our compression-based selection is grounded in the manifold hypothesis: normal features concentrate on a low-dimensional manifold in high-dimensional space.

Let $z_l = f_l(x) \in R^D$ denote the feature representation at layer l. Although the ambient dimension is D, normal samples are assumed to concentrate on a smooth manifold $\mathcal{M}_l$ with intrinsic dimensionality $d_l \ll D$.

Under this assumption, the reliability of nonparametric density estimation depends primarily on d_l. For density estimation on a d_l-dimensional manifold, the convergence rate satisfies

$$E\|\hat{p}_l - p_l\| = \mathcal{O}\Big(N^{-4/(4+d_l)}\Big), \tag{4}$$

indicating that smaller intrinsic dimensionality enables more stable modeling.

Similarly, for k-nearest-neighbor distance estimation,

$$r_k(z) \sim \left(\frac{k}{N p_l(z)}\right)^{1/d_l}, \tag{5}$$

suggesting that estimation variance increases with intrinsic dimensionality. Therefore, layers exhibiting strong intrinsic compression are expected to produce more compact and stable normal manifolds.

However, excessive compression may remove anomaly-relevant structure. This trade-off can be interpreted through the Information Bottleneck principle,

$$\min I(X; Z_l) - \beta I(Z_l; S), \tag{6}$$

which balances input compression and preservation of task-relevant information. In practice, over-compression is mitigated by excluding tail layers prone to domain-specific overfitting.

4. Experiments

4.1 Image datasets and evaluation methods

We evaluate the generalization capability of the proposed intrinsic-dimensionality-based layer selection across different industrial domains and anomaly types. Experiments were conducted on three public anomaly detection benchmarks using four representative feature-embedding-based methods. MVTec-AD[5]is a widely used industrial anomaly detection dataset. It comprises 15 industrial categories (textures and objects) with 5,354 images in total, covering more than 70 defect types. VisA[6]

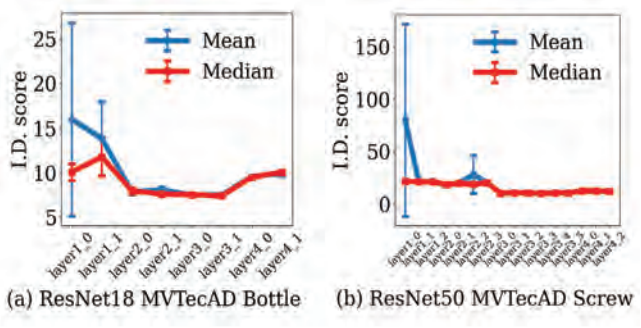

Fig. 3 Stability comparison of mean and median TwoNN aggregation

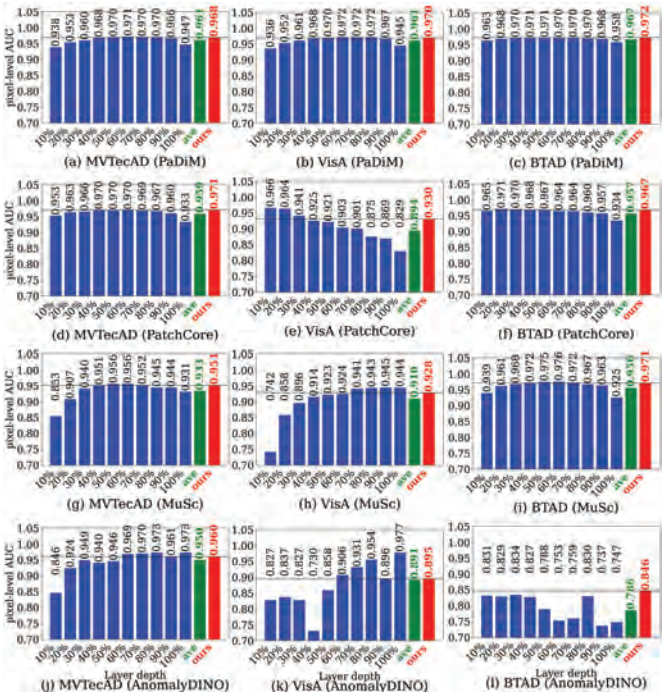

Fig. 4 Dataset-wise pixel-level AUC (MVTecAD/VisA/BTAD). Blue: per-layer, green: average, red: ID-based selection. The proposed selection consistently exceeds the average baseline

includes 12 colorful objects and 10,821 images, offering high intra-class variability. BTAD[7]provides 2,830 real-world images of three objects with both body- and surface-level defects. These datasets differ in scale, texture complexity, and imaging conditions, enabling cross-domain validation of the proposed approach.

For evaluation, we adopt four representative embedding-based anomaly detection methods using pretrained backbones (CNN or ViT). PaDiM[20]models the normal feature distribution as a multivariate Gaussian; PatchCore[21]performs nearest-neighbor search within a memory bank of sampled patch features; MuSc[22]conducts zero-shot mutual scoring between unlabeled images; and AnomalyDINO[23]adapts DINOv2 for few-shot anomaly detection without fine-tuning. All of these methods rely on fixed feature layers, allowing a direct evaluation of how the proposed ID-based layer

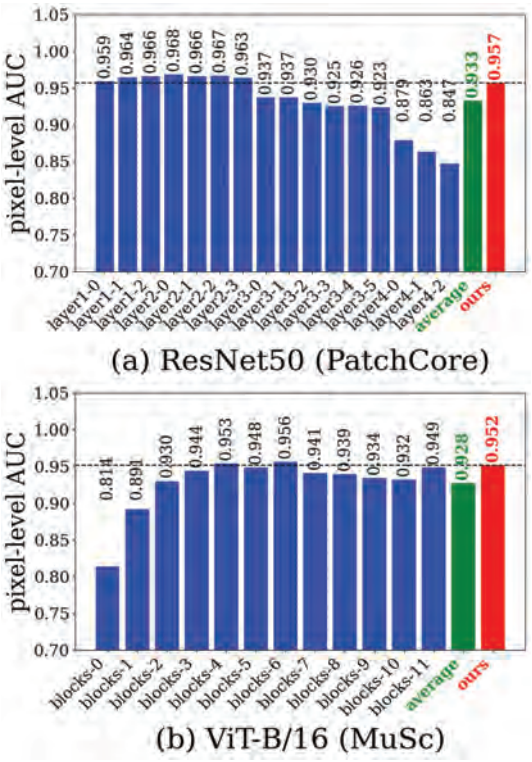

Fig. 5 Layer-wise pixel-level AUC comparison. (a) ResNet50 (PatchCore) and (b) ViT-B/16 (MuSc). The proposed ID-based selection (red) achieves performance comparable to or better than fixed-layer and averaged baselines

selection improves performance across datasets and domains.

4.2 Experimental settings

All anomaly detection models were evaluated under unified conditions. The input resolution was fixed to 256×256 pixels for fair comparison.

Intrinsic dimensionality was estimated using the TwoNN method[15]. To improve robustness, we adopt median-based aggregation after clipping the top 2% of μ_i values ($Q_{0.98}$), which reduces instability in sparse feature manifolds.

For CNN backbones, ID was computed on post-activation feature maps, and for ViT backbones, after the residual connection in the MLP block. All other hyperparameters followed official implementations. Pixel-level AUC was used as the evaluation metric.

Experiments were conducted under both full-shot and few-shot settings. In few-shot experiments, layer selection was repeated with three random seeds and averaged.

All ID estimation and layer selection were performed exclusively using normal training data.

We evaluated representative CNN-based (PaDiM, PatchCore) and ViT-based (MuSc, AnomalyDINO) methods to verify backbone-agnostic behavior.

Figure 3 shows that median-based aggregation provides more stable ID estimates than mean-based aggregation, particularly in shallow layers. The dataset-wise

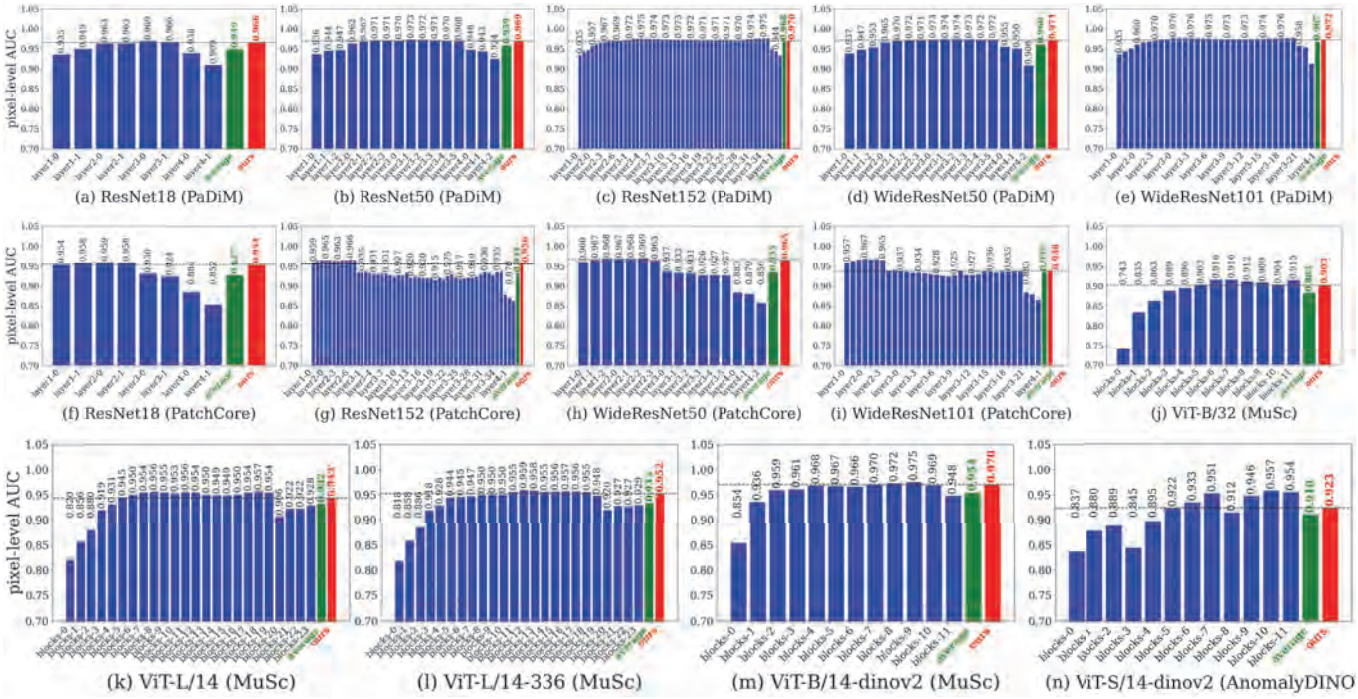

Fig. 6 Pixel-level AUC comparison across anomaly detection methods and backbone architectures. (a)–(e): PaDiM (CNN), (f)–(i): PatchCore, (j)–(m): MuSc, and (n): AnomalyDINO (ViT)

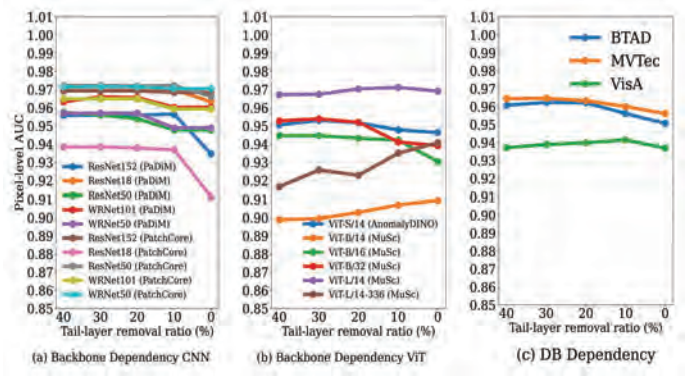

Fig. 7 Pixel-level AUC under different tail-layer removal ratios (0–40%). (a) CNN Backbone vs. removal ratio; (b) ViT Backbone vs. removal ratio; (c) dataset vs. removal ratio. Performance remains stable, indicating low sensitivity of the proposed exclusion strategy

and representative layer-wise comparisons are shown in **Fig. 4** and **Fig. 5**, respectively.

5. Results and Analysis

5.1 Overall performance comparison

As shown in **Fig. 6**, across PaDiM, PatchCore, MuSc, and AnomalyDINO, the proposed ID-based selection consistently outperforms the layer-average baseline and closely approaches the best-performing single layer.

Across 480 experimental combinations spanning multiple methods, backbones, and datasets (MVTec-AD, VisA, BTAD), the proposed method achieves 97.7% ± 3.7% of the ground-truth best-layer performance, compared with 96.1% ± 4.3% for simple layer averaging.

These results demonstrate that the proposed approach reliably approximates the optimal layer without using abnormal samples or ground-truth supervision.

Figures 4 and 5 further confirm that the improvement holds consistently across datasets, detection methods, and backbone architectures.

5.2 Sensitivity analysis of tail-layer filtering

Recent studies on representation geometry suggest that intrinsic dimensionality (ID) generally exhibits compression behavior toward deeper layers, although the detailed evolution differs between CNN and ViT. CNNs often show a clearer monotonic ID reduction[12], whereas ViTs may maintain relatively high ID even near the final layers[24]. Semantically informative representations often emerge around inter-layer ID compression rather than strictly at the deepest layers. Motivated by this observation, we adopt a backbone-agnostic heuristic that excludes the final 20% of layers from the candidate set to mitigate potential domain-specific overfitting.

To examine the robustness of this heuristic, we vary the excluded tail ratio from 0% to 40%. As shown in **Fig. 7**(a,b), when the exclusion ratio changes between 40% and 20%, the variation in pixel-level AUC across different backbones remains within approximately one percentage point (pp). Figure 7(C) further evaluates dataset dependency across BTAD, MVTec, and VisA. Across datasets, performance variation remains within roughly 0.5 pp for exclusion ratios between 40% and 10%.

Although certain ViT backbones occasionally bene-

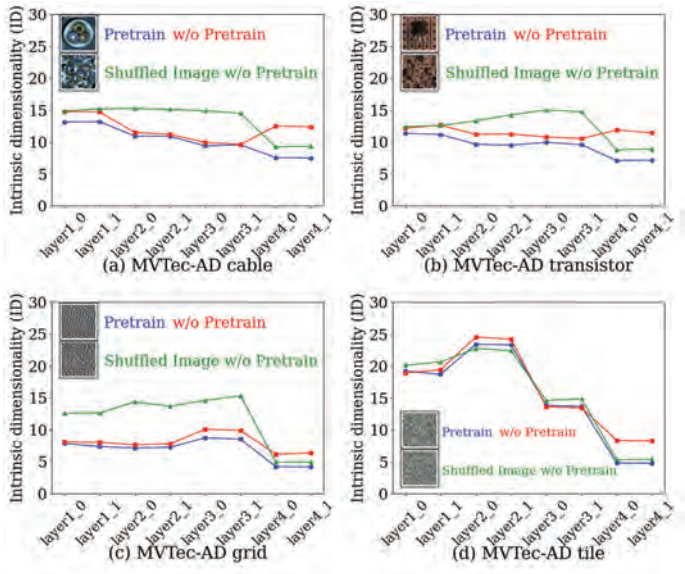

Fig. 8 Layer-wise intrinsic dimensionality (ID) under domain mismatch (ImageNet-only pretraining) versus domain-aligned pretraining. Re-expansion in deeper layers suggests potential domain-specific overfitting

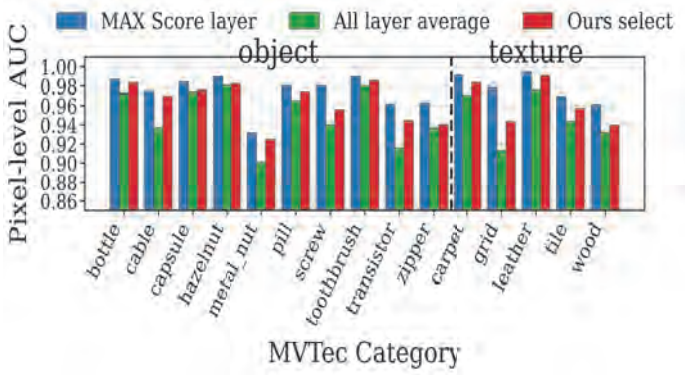

Fig. 9 Pixel-level AUC on object and texture categories of MVTec

fit from including deeper layers, this tendency is not consistent across models or datasets. Overall, selecting candidate layers within roughly the 10–40% tail-exclusion range yields practically stable performance without backbone-specific tuning. Since this variation is smaller than the gain over simple averaging (approximately 1.6 pp), the 20% cutoff is a robust heuristic rather than a tuned parameter.

5.3 Validation of ID as structural indicator

As shown in **Fig. 8**, shuffled images exhibit persistently high intrinsic dimensionality (ID) through intermediate layers, indicating that hierarchical information integration does not fully progress. Near the deepest layers, ID drops abruptly, forming a compensatory compression that suggests an attempt to reconcile disrupted representations without forming a compact feature manifold.

This behavior differs between object-centric and texture-centric categories. For object categories, patch shuffling destroys global shape structure, leading to a pronounced ID plateau. This indicates that hierarchical feature abstraction in pretrained CNN relies strongly on spatial configuration and contour consistency.

In contrast, texture categories show minimal ID change after shuffling. This is consistent with the known texture bias of CNN representations, where recognition depends primarily on local statistical regularities rather than global geometry. Because patch shuffling largely preserves such local statistics, the ID profile remains relatively stable.

Importantly, the proposed layer-selection mechanism does not depend directly on spatial-structure sensitivity. Algorithm 2 selects layers based on maximal inter-layer ID compression, which reflects representational compactness and abstraction rather than explicit spatial coherence. Consequently, even when spatial structure is less informative (e.g., texture-dominated categories), the ID compression dynamics still provide a reliable indicator for stable layer selection.

As shown in **Fig. 9**, the layers selected by the proposed method achieve performance close to the oracle best layer determined post hoc for each dataset and backbone. The remaining performance gap between the proposed layer and this oracle best layer is slightly larger for textures (approximately 1.6 pp) than for objects (approximately 1.1 pp). Nevertheless, the proposed method consistently outperforms simple layer averaging.

These results indicate that intrinsic-dimensionality-based layer selection remains effective even when spatial structural cues are weak, supporting its applicability across diverse anomaly types.

5.4 Validation of the manifold assumption via PCA and DB score comparison

The proposed method estimates the intrinsic dimensionality (ID) of feature representations, assuming that these features lie on a low-dimensional manifold. To examine this assumption, we compare the estimated ID with two alternative indicators: (i) the number of principal components explaining 90% of variance via Principal Component Analysis (PCA), and (ii) the Davies–Bouldin (DB)[16]cluster separation score.

We analyze the variation of each indicator with respect to the number of normal training samples and feature-extraction layers using a ResNet18 backbone on the MVTec-AD dataset. As shown in **Fig. 10**, both ID and PCA show low sensitivity to sample count, indicating stable layer-wise trends even in few-shot settings. In contrast, the DB score fluctuates greatly due to the distance concentration phenomenon in high-dimensional spaces, where inter- and intra-cluster distances become unreliable under sparse sampling.

We then investigate whether selected layers differ across

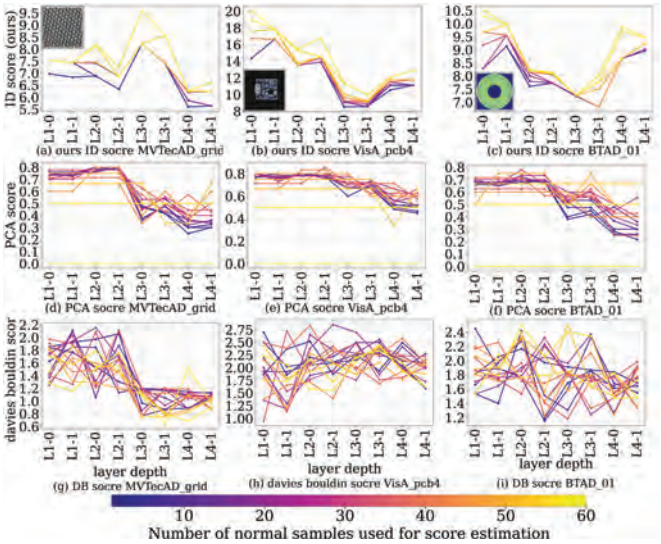

Fig. 10 Stability of ID vs PCA/DB w.r.t. sample size (ResNet18/MVTec-AD). DB is unstable under sparse sampling; ID remains stable

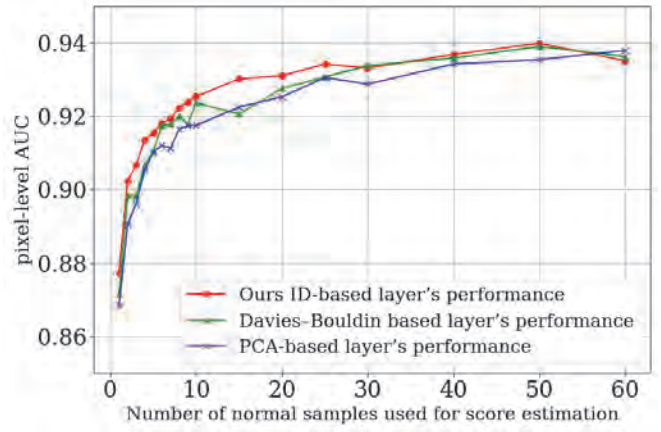

Fig. 11 Few-shot pixel-level AUC under varying normal-sample counts. ID-based selection > PCA/DB across settings

object categories. As illustrated in Fig. 10, ID-based trends vary by task while remaining robust to sample size, whereas PCA-based selection shows similar patterns across categories, reflecting backbone dependency rather than domain adaptation.

Finally, we assess how each indicator contributes to anomaly detection performance. Using each metric, layers were selected for different normal-sample counts, and anomaly detection was evaluated by pixel-level AUC. As shown in **Fig. 11**, when samples are limited, the proposed ID-based selection consistently outperforms PCA and DB, demonstrating that ID provides a more stable and informative criterion for few-shot anomaly detection.

5.5 Computation time analysis

We evaluate the calibration cost of ID-based selection and compare it with representative baselines.

Layer selection is performed only once per category using normal data and does not affect inference-time processing. The reported time therefore corresponds to the one-time calibration stage, including ID estimation over candidate layers (excluding the last 20%) and the final selection step.

Table 1 Computation time and memory-bank construction time (seconds; PaDiM: CNN, MuSc: ViT)

Res.	Model	Ours ID			PatchCore			PaDiM/MuSc		
		1img	10imgs	60imgs	1img	10imgs	60imgs	1img	10imgs	60imgs
256	resnet18	5.9	6.6	9.2	0.9	1.0	1.5	0.3	0.9	3.5
	resnet152	7.2	7.7	12.3	1.0	1.4	3.4	5.3	8.4	14.4
	ViT-B-16	7.3	7.6	9.8	0.9	1.2	2.1	0.3	0.8	3.5
	ViT-L-14	10.4	11.0	14.0	1.0	1.7	5.2	0.3	1.0	5.0
512	resnet18	5.9	7.6	14.7	0.9	1.1	1.7	1.2	2.2	7.1
	resnet152	7.3	9.2	22.0	1.1	1.6	4.5	22.5	43.3	49.0
	ViT-B-16	7.3	7.8	11.2	1.0	1.7	5.2	0.4	1.5	8.3
	ViT-L-14	10.4	11.4	17.1	1.5	4.3	19.6	0.5	2.6	15.9
1024	resnet18	7.4	20.8	96.2	1.0	1.5	6.3	5.6	8.8	21.4
	resnet152	9.6	32.0	162.8	1.4	3.3	23.0	140.4	176.0	462.0
	ViT-B-16	7.3	9.3	21.0	2.1	6.3	31.8	0.9	6.9	41.0
	ViT-L-14	11.0	16.1	45.1	5.3	25.2	137.0	2.0	19.3	119.0

For comparison, we measured the calibration time of PatchCore, PaDiM, and MuSc under their standard settings. All computational cost measurements were performed on the same hardware environment equipped with an NVIDIA RTX A2000 GPU (12GB VRAM), a Xeon-class CPU, and 64GB RAM, and the results are summarized in **Table 1**. All reported times are measured in seconds.

The proposed method introduces only a small additional calibration overhead and remains comparable in cost to memory-bank construction approaches, particularly since the procedure is executed only once per category and does not impact inference efficiency.

6. Limitations

The current framework uses a fixed 20% tail-layer exclusion to mitigate potential overfitting. Sensitivity analysis confirms stability across architectures and datasets, indicating practical effectiveness in real-world use.

Future work includes adaptive tail-depth estimation based on data characteristics and extensions to dynamic multi-layer selection for further performance improvement.

7. Conclusion

We proposed a training-free layer-selection framework based on intrinsic dimensionality (ID) of normal features. The method automatically identifies stable feature layers without abnormal data.

Across three industrial benchmarks and multiple backbones, the proposed approach consistently approximates the optimal layer, achieving 97.7% of best-layer performance while outperforming fixed-layer, random-layer, and PCA/DB-based criteria.

These results indicate that intrinsic compression dynamics provide a practical basis for robust layer selection.

References

1) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel:"Generalization and Network Design

Strategies", Connectionism in Perspective, Vol. 19, pp. 143–155 (1989).
2) D. P. Kingma, M. Welling:"Auto-Encoding Variational Bayes", Proc. of International Conference on Learning Representations (2014).
3) C.-L. Li, K. Sohn, J. Yoon, T. Pfister:"CutPaste: Self-Supervised Learning for Anomaly Detection and Localization", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9664–9674 (2021).
4) V. Zavrtanik, M. Kristan, D. Skočaj:"DRAEM: A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection", Proc. of IEEE/CVF International Conference on Computer Vision, pp.8330–8339 (2021).
5) P. Bergmann, M. Fauser, D. Sattlegger, C. Steger:"MVTec AD: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.9592–9600 (2019).
6) Y. Zou, J. Jeong, L. Pemula, D. Zhang, O. Dabeer:"Spot-the-Difference Self-Supervised Pretraining for Anomaly Detection and Segmentation", Proc. of European Conference on Computer Vision, pp.392–408 (2022).
7) P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, G. L. Foresti:"VTADL: A Vision Transformer Network for Image Anomaly Detection and Localization", Proc. of IEEE International Symposium on Industrial Electronics, pp.1–6 (2021).
8) L. Heckler, R. König, P. Bergmann:"Exploring the Importance of Pretrained Feature Extractors for Unsupervised Anomaly Detection and Localization", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.2917–2926 (2023).
9) L. Cayton:"Algorithms for Manifold Learning", Tech. Rep., Univ. California San Diego (2005).
10) U. Cohen, S. Chung, D. D. Lee, H. Sompolinsky:"Separability and Geometry of Object Manifolds in Deep Neural Networks", Nature Communications, Vol. 11, No.1, pp. 1–13 (2020).
11) C. Stephenson, J. Feather, S. Padhy, O. Elibol, H. Tang, J. McDermott, S. Chung:"Untangling in Invariant Speech Recognition", Advances in Neural Information Processing Systems (2019).
12) A. Ansuini, A. Laio, J. H. Macke, D. Zoccolan:"Intrinsic Dimension of Data Representations in Deep Neural Networks", Advances in Neural Information Processing Systems (2019).
13) D. Doimo, A. Glielmo, A. Ansuini, A. Laio:"Hierarchical Nucleation in Deep Neural Networks", Advances in Neural Information Processing Systems (2020).
14) I. Kaufman, O. Azencot:"Data Representations' Study of Latent Image Manifolds", Proc. of International Conference on Machine Learning, pp.15928–15945 (2023).
15) E. Facco, M. d'Errico, A. Rodriguez, A. Laio:"Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information", Scientific Reports, Vol. 7, No. 1, pp. 12140 (2017).
16) D. L. Davies, D. W. Bouldin:"A Cluster Separation Measure", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, No.2, pp.224–227 (1979).
17) G. Magai, A. Ayzenberg:"Topology and Geometry of Data Manifold in Deep Learning", arXiv preprint arXiv:2204.08624 (2022).
18) J. A. Costa, A. O. Hero:"Geodesic Entropic Graphs for Dimension and Entropy Estimation in Manifold Learning", IEEE Trans. on Signal Processing, Vol. 52, No. 8, pp. 2210–2221 (2004).
19) P. Pope, C. Zhu, A. Abdelkader, M. Goldblum, T. Goldstein:"The Intrinsic Dimension of Images and Its Impact on Learning", Proc. of International Conference on Learning Representations (2020).
20) T. Defard, A. Setkov, A. Loesch, R. Audigier:"PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization", Proc. of International Conference on Pattern Recognition, pp.475–489 (2021).
21) K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, P. V. Gehler:"Towards Total Recall in Industrial Anomaly Detection", Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.14298–14308 (2022).
22) X. Li, Z. Huang, F. Xue, Y. Zhou:"MuSc: Zero-Shot Industrial Anomaly Classification and Segmentation with Mutual Scoring of the Unlabeled Images", arXiv preprint arXiv:2401.16753 (2024).
23) S. Damm, M. Laszkiewicz, J. Lederer, A. Fischer:"AnomalyDINO: Boosting Patch-based Few-shot Anomaly Detection with DINOv2", Proc. of IEEE/CVF Winter Conference on Applications of Computer Vision (2025).
24) A. Valeriani, M. A. Musallam, S. Papyan, M. Elad:"Intrinsic Dimensionality Dynamics in Vision Transformers", arXiv preprint arXiv:2302.00294 (2023).

(Received October 31, 2025)
(Revised February 26, 2026)

Masahiro TSUCHIYA
(Student Member)
He received his M.S. degree in Engineering from Tokyo University of Agriculture and Technology, Japan, in 2018. Since 2024, he has been pursuing his Ph.D. degree in Computer Science at Chubu University, Japan. He has been working at OMRON Corporation as a researcher since 2023. His research interests are related to computer vision and machine learning.

Tsubasa HIRAKAWA
He received his Ph.D. in Computer Science from Hiroshima University, Japan, in 2017. From 2017 to 2019, he was a research fellow at Chubu University and has been a specially appointed associate professor at the Chubu Institute for Advanced Studies since 2019. He was a fellow of the Japan Society for the Promotion of Science from 2014 to 2017 and a visiting researcher at ESIEE Paris, France, in 2014 and 2015.

Takayoshi YAMASHITA
He received his Ph.D. in Computer Science from Chubu University, Japan, in 2011. He worked at OMRON Corporation from 2002 to 2014 and was a lecturer in the Department of Computer Science, Chubu University, from 2014 to 2017, and an associate professor from 2017 to 2021. He has been a professor there since 2021. His research interests include object detection, object tracking, pattern recognition, and machine learning. He is a member of IEEE, IEICE, and IPSJ.

Hironobu FUJIYOSHI
He received his Ph.D. in Electrical Engineering from Chubu University, Japan, in 1997. From 1997 to 2000, he was a post-doctoral fellow at the Robotics Institute of Carnegie Mellon University, Pittsburgh, PA, USA, working on the DARPA Video Surveillance and Monitoring project and the humanoid vision project for the HONDA Humanoid Robot. He is now a professor in the Department of Robotics, Chubu University. His research interests include computer vision and video understanding. He is a member of the IEEE, IEICE, and IPSJ.

IIEEJ Paper

A Two-Stage Structured-Light Neural Network for Surface Dent Detection and Geometric Measurement of Car Body Shells

Sikun WANG† (*Member*), Cunwei LU†

†Fukuoka Institute of Technology

<Summary> Surface defect detection of automotive panels, especially scratches and dents, is difficult because of strong reflections on metallic surfaces. This study presents a two-stage structured-light neural network for automatic dent detection and geometric measurement. The system combines high-frequency structured-light imaging with deep learning to reduce reflection interference. In the first stage, a heatmap network (HmNet) detects dent centers from a single structured-light image. In the second stage, a regression network (DentRegressor) estimates each dent's depth and diameter, achieving end-to-end detection and measurement without 3D reconstruction. Experiments on 875 structured-light images achieved 98.54% precision, 90.11% recall, and a 94.13% F1-score for dent localization. The regression results yielded RMSEs of 0.267 mm for depth and 13.96 mm for diameter. Compared with the Mask R-CNN method, the proposed system provides higher accuracy, better robustness to reflections, and faster inference, showing strong potential for industrial surface inspection applications.

Keywords: structured light, surface defect detection, dent measurement, heatmap regression

1. Introduction

1.1 Background

With the growing emphasis on energy conservation and environmental protection, the remanufacturing and reuse of automotive components have become an essential part of the circular economy[1]. According to Japan's End-of-Life Vehicle Recycling Law (2002 promulgated), used parts must undergo strict visual and structural inspections before reentering the market to ensure safety and reliability. Minor dents and scratches on the vehicle body surface not only affect appearance but may also weaken mechanical strength, reducing reuse value. Therefore, high-precision and automated surface defect detection is crucial for improving remanufacturing efficiency and quality control.

For highly reflective metal surfaces such as car body panels, conventional inspection methods are mainly divided into contact and non-contact types[2]. Contact methods, including coordinate measuring machines and robotic probes, offer high accuracy but may deform thin (0.6–1 mm) elastic panels or even cause surface damage. Non-contact methods, such as vision-based thresholding, edge detection, and structured-light measurement, are faster and non-destructive. However, on reflective surfaces, these approaches often fail due to specular reflection, shadows, and noise[5]. As shown in **Fig. 1**, the left image illustrates the strong reflectivity of a vehicle body surface, where ambient light and surrounding objects cause stray reflections. The right image shows the result of a conventional structured-light measurement on a reflective metal surface, exhibiting uneven illumination, overexposure in the center, and insufficient brightness at the edges, all of which degrade defect detection accuracy.

Structured-light measurement infers fine surface deformations from projected stripe or dot pattern distortions[3,5], effectively amplifying micro-shape variations. Yet, geometric analysis alone is insufficient for automatic defect recognition, particularly under reflective or noisy conditions. With the rise of deep learning, convolutional neural networks and their derivatives (e.g., Mask R-CNN, YOLO, U-Net) have shown strong feature extraction and classification capabilities in industrial vision[4]. Combining structured-light measurement with intelligent recognition algorithms greatly enhances detection sensitivity, robustness, and automation[7], offering significant theoretical and practical value for intelligent inspection of automotive components.

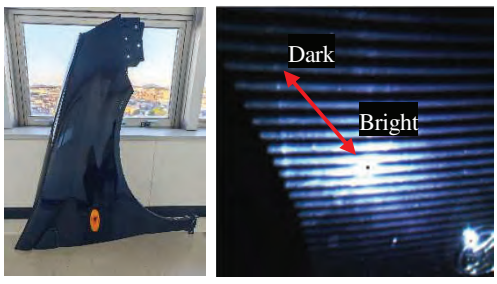

Fig. 1 Examples of reflective surface imaging (adapted from[7])

1.2 Related works

In recent years, numerous image-processing and deep-learning-based methods have been proposed by researchers worldwide for detecting dents and scratches on highly reflective metal surfaces such as automotive body panels. Overall, existing studies have mainly focused on two directions: improving optical structures and optimizing intelligent algorithms.

In terms of optical structure improvement, our research group previously proposed the Indirect Illumination Method[6], which forms indirect structured light by passing LED illumination through a semi-transparent diffusion film. This approach effectively reduces interference from reflective surfaces during image acquisition and enables accurate detection of dents and scratches, demonstrating the feasibility of indirect illumination under complex reflective conditions.

However, the method requires multiple projections of structured light with different initial phases to obtain phase-shifted images, resulting in longer detection time and a more complex system configuration, which limits its real-time performance and industrial applicability. After capturing four structured-light images with different phases, the resulting image clearly reveals the dent positions.

To further improve detection speed and accuracy, subsequent research developed a hybrid system that integrates high-frequency structured light with neural network analysis[7]. As shown in **Fig. 2**, the system projects high-frequency structured light using a multi-array LED lattice to enhance the visualization of subtle surface deformations. It then employs a Mask R-CNN segmentation network for automatic dent recognition and quantitative analysis, successfully identifying three dent locations.

In summary, significant progress has been made in optical imaging, image enhancement, and deep-learning-based detection. However, achieving high detection accuracy while further improving system robustness and real-time performance remains a major challenge in the surface defect detection of automotive body panels. Building upon our previous research, this study explores a high-precision, non-contact detection approach suitable for complex reflective surfaces by combining high-frequency structured light, image enhancement, and optimized deep neural network techniques.

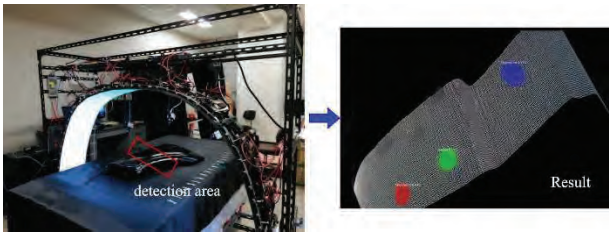

Fig. 2 Detection results using the neural-network-based detection system (reproduced from our previous work[7])

2. Method

2.1 System overview

The proposed system is an improved version of the previous detection framework that combines high-frequency structured light with neural networks. In the original system, high-frequency structured light was projected onto the car body surface, and Mask R-CNN was used as the main detection algorithm for dent segmentation. However, on highly reflective metallic surfaces, Mask R-CNN suffered from blurred mask boundaries, slow inference, and low accuracy for small dents[8].

To overcome these issues, the software algorithm was optimized and reconstructed while keeping the hardware unchanged. The improved system adopts a two-stage deep learning structure for dent localization and geometric parameter regression. This design maintains the high sensitivity and sub-millimeter precision of structured-light imaging, while improving speed, stability, and interpretability. With a single structured-light reflection image as input, the system outputs each dent's center, maximum depth, and characteristic diameter.

As shown in **Fig. 3**, high-frequency lattice structured light is projected onto the car body, and the reflected image is captured by an industrial camera. The deformed light pattern indicates potential dents. The heatmap detection network (HmNet) generates a probability heatmap of dent centers, from which peak points are extracted using non-maximum suppression (NMS), and the threshold for NMS was set to 0.3. These centers define regions of interest (ROIs) that are passed to the regression network (DentRegressor) to predict dent depth and diameter. The system finally outputs dent positions, geometric parameters, confidence values, and visualization results for automatic and quantitative dent detection. The structured-light projector generates a black–white lattice pattern using an LED array with a spacing of approximately 2 mm, and the projector is arranged parallel to the car body surface to ensure consistent imaging.

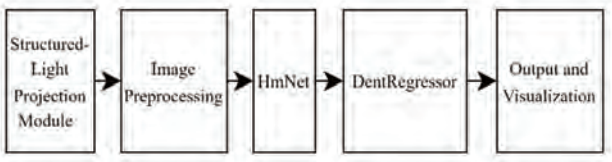

Fig. 3 Workflow of the proposed two-stage dent detection system

2.2 Definition of dent geometry

In car body surface inspection, the geometry of a dent can generally be approximated as a locally smooth, rotationally symmetric surface, whose shape can be expressed by a depth function:

$$z = f(r) \tag{1}$$

where r is the radial distance from the dent center, and z is the vertical displacement of the surface relative to the reference plane.

In the practical measurement process, a high-precision surface roughness profiler was used to measure the depth distribution of multiple dent samples. The results show that the variation of dent depth with respect to the radius can generally be divided into two stages, as illustrated in **Fig. 4**. The upper part of the figure shows a schematic diagram of the dent structure, while the lower part presents an actual structured-light image of a dent. It can be observed that when r is within approximately half of the dent radius, the depth changes significantly with radius and exhibits a large curvature. In the structured-light image, this region appears as a strongly stretched and distorted lattice pattern. When r exceeds half of the dent radius, the depth variation becomes gradual, and the deformation of the reflected structured light weakens, making the boundary difficult to distinguish. From a physical perspective, this behavior corresponds to the typical deformation characteristics of a thin metal sheet under localized impact loading: the central region undergoes plastic deformation and cannot recover its original shape after impact, while the peripheral region remains within the elastic range.

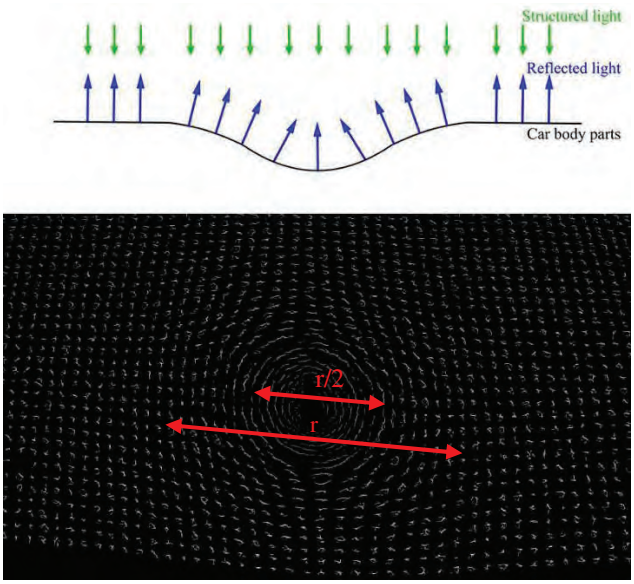

Fig. 4 Schematic and structured-light images of a typical dent

Therefore, under different impact forces and contact areas, the dent depth profile does not strictly conform to a single mathematical model; its shape typically lies between a parabolic and a Gaussian model. The geometric characteristics of the dent, denoted as S, are defined as follows and are used as both input and output parameters in the subsequent network model:

$$S = (d_{max}, D_{50}) \tag{2}$$

where

$d_{\max}$: maximum dent depth, representing the magnitude of vertical deformation.

D_{50}: diameter corresponding to half of the maximum depth, representing the lateral deformation extent.

2.3 Two-stage neural network

(a) Heatmap detection network

HmNet serves as the first stage of the proposed two-stage detection system. It generates a pixel-level confidence heatmap of dent centers from structured-light reflection images. Based on the existing high-frequency structured-light hardware setup, the original Mask R-CNN model was redesigned and lightweighted to improve efficiency and robustness. As shown in **Table 1**, compared with the original Mask R-CNN model, the proposed lightweight design reduces the number of parameters by 80% and improves inference speed from 0.08 FPS to 1.8 FPS. The number of trainable parameters was reduced from 63 million to 13 million, corresponding to a reduction of approximately 79%.

The overall architecture adopts a Fully Convolutional Network (FCN) structure with ResNet-18 as the backbone for feature extraction, followed by multiple up-sampling layers to restore spatial resolution. The network takes a single-channel structured-light grayscale image as input and outputs a proportionally sized heatmap, where the intensity represents the probability of a dent center.

As shown in **Fig. 5**, the network consists of three main components: Backbone, Neck, and Head. The Backbone extracts multi-scale features, the Neck performs channel compression and semantic fusion, and the Head restores spatial resolution through two stages of transposed convolution. The final normalized probability map is obtained via a Sigmoid activation function. This design significantly reduces the number of parameters and improves both training

Table 1 Comparison of model size and inference efficiency

Method	Params	FPS	Memory
Mask R-CNN	63M	0.08	4.8GB
HmNet	13M	1.8	1.2GB

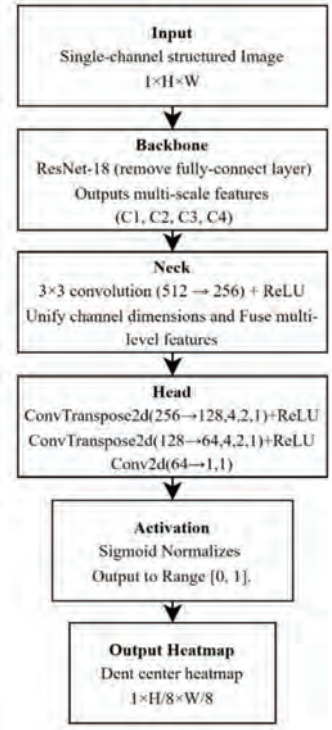

Fig. 5 Architecture of the Heatmap Network

and inference efficiency. Moreover, it demonstrates enhanced robustness on reflective metal surfaces and higher sensitivity to small dents.

In the training dataset, each dent is annotated as a circular region whose center corresponds to the dent center and whose radius is half of the measured diameter $D_{50}/2$. This region is then converted into a two-dimensional Gaussian distribution:

$$H_{(x,y)} = exp\left(-\frac{(x-x_c)^2+(y-y_c)^2}{2\sigma^2}\right) \quad (3)$$

where the standard deviation $\sigma$is proportional to the radius and defined as $\sigma = \max\left(6, \frac{r}{2}\right)$.

(b) DentRegressor network

The DentRegressor network is the proposed regression model for estimating dent geometric parameters. It directly predicts the maximum depth $d_{\max}$and characteristic diameter $D_{50}$from local structured-light images. The input is a single-channel grayscale image. By analyzing the deformation patterns of the local structured-light lattice, the network learns the nonlinear mapping between image brightness distribution, curvature variation, and dent morphology, thereby achieving direct geometric parameter estimation without the need for 3D reconstruction.

As shown in **Fig. 6**, the network consists of two main modules: a Backbone and a Regression Head. The backbone adopts a ResNet-18 architecture with fully connected layers

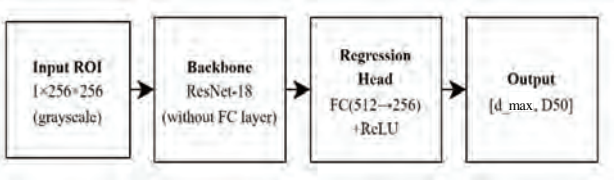

Fig. 6 Architecture of the DentRegressor network

removed, which is used to extract local geometric and texture features. After four residual convolutional down-sampling blocks, the output feature map has a size of $512 \times 8 \times 8$, capturing the deformation and grayscale patterns within the dent region. A global average pooling layer then compresses the 2D feature map into a 512-dimensional vector, preserving overall shape information and eliminating spatial offset effects. On this basis, the Regression Head consists of two fully connected layers: the first maps the 512-dimensional feature vector to 256 dimensions with ReLU activation to enhance nonlinear representation, while the second outputs two continuous values corresponding to the maximum depth $d_{\max}$and characteristic diameter D_{50}.

Compared with traditional pixel-level regression networks based on Mask R-CNN[8], DentRegressor no longer performs pixel-wise segmentation. Instead, it adopts a "local perception–global regression" strategy to directly predict geometric parameters. This design reduces model complexity, avoids segmentation errors caused by blurred mask boundaries on reflective surfaces, and improves computational efficiency and measurement stability.

(c) Loss function design

In the first-stage training, dent centers occupy only a very small fraction of pixels in the structured-light image, while the background area is overwhelmingly larger. This leads to a severe imbalance between positive and negative samples. If conventional loss functions such as Mean Squared Error (MSE) is applied, the network tends to produce near-zero outputs to minimize the overall error, which results in divergent heatmap responses and degraded detection accuracy for dent centers.

To address this issue, Gaussian Focal Loss is employed to better constrain the network output and adapt it to sparse center-point prediction tasks. This loss function originates from the CenterNet model[9] and is specifically designed for heatmap-based keypoint regression. Unlike the standard Focal Loss, it renders each ground-truth center point as a two-dimensional Gaussian distribution rather than a single pixel. This formulation provides spatially continuous gradients, allowing smoother convergence during training. Its mathematical expression is given as follows:

$$L = -\frac{1}{N}\sum_{x,y}\begin{cases}(1-\hat{H}_{xy})^{\alpha}\log(\hat{H}_{xy}), if\ H_{xy}=1\\(1-H_{xy})^{\beta}(\hat{H}_{xy})^{\alpha}\log(1-\hat{H}_{xy}), if\ H_{xy}<1\end{cases} \quad (4)$$

where

H_{xy} denotes the value of the ground-truth Gaussian heatmap at pixel (x, y), and $\hat{H}_{xy}$ represents the predicted value from the network.

α and β are scaling factors that balance the gradient contributions of positive and negative samples, while N is a normalization factor corresponding to the number of target objects in the image. In this study, the parameters are set to $\alpha = 2.0$ and $\beta = 4.0$.

In the dent regression stage, the DentRegressor network aims to predict the geometric parameters of each dent, the maximum depth $d_{\max}$ and characteristic diameter D_{50} — based on the local structured-light image (ROI) centered at the detected dent. This process essentially involves learning the parameterized relationship of the surface morphology function $f(r)$ from the pixel features of the structured-light image. Therefore, the loss function must not only accurately measure the discrepancy between the predicted and physically measured values but also ensure regression stability and physical interpretability.

Several loss functions including MAE, Smooth L1, and Mean Squared Error (MSE) were evaluated during preliminary experiments. MAE is generally more robust to outliers but may lead to slower convergence in regression tasks. Since MSE penalizes larger errors more strongly and encourages the reduction of geometric deviations, it demonstrated more stable convergence and lower validation error in our experiments. Therefore, MSE was selected as the optimization objective for the regression network.

The MSE loss measures regression accuracy by computing the squared difference between the predicted and ground-truth values, expressed as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[\left(\hat{d}_{max}^{(i)} - d_{max}^{(i)}\right)^2 + \left(\hat{D}_{50}^{(i)} - D_{50}^{(i)}\right)^2\right] \quad (5)$$

where $\hat{d}_{\max}^{(i)}$ and $\hat{D}_{50}^{(i)}$ denote the predicted dent depth and diameter of the i-th sample, while $d_{\max}^{(i)}$ and $D_{50}^{(i)}$ represent the corresponding ground-truth measurements.

3. Experiments and Results

3.1 Experimental environment

All experiments in this study were conducted in a deep learning environment built on the Windows 11 operating system, using the PyTorch framework for network training and inference. The hardware platform was equipped with an NVIDIA RTX 3080Ti GPU (16 GB), an Intel Core i9-12950HX CPU, and 32 GB of RAM.

The input data consisted of structured-light reflection images with an original resolution of 2560 × 1440. During preprocessing, all images were converted to grayscale, normalized, and resized to match the input dimensions required by each network: 512 × 288 for the HmNet stage and 256 × 256 for the ROI images in the DentRegressor stage.

The dataset used in the experiments was a custom-built dataset containing 875 images of car body surfaces in eight different colors (red, yellow, blue, green, black, white, silver, and beige), as shown in **Fig. 7**. These samples exhibited significant variations in color, reflectivity, and material properties, designed to simulate the optical reflection characteristics of real vehicle shells under various coating conditions. Each image contained 1–5 dents formed under different impact forces. The dent depth ranged from 0.074 mm to 1.178 mm, and the diameter ranged from 14 mm to 67 mm.

All sample depths were measured using a BRK BC-800 surface roughness tester and annotated with LabelMe 4.5.1. Each dent was marked as a circular region, where the circle center represents the dent center, and the radius and label fields record the measured diameter and depth values.

The dataset was divided into a training set (70%) and a validation set (30%). The training set was used for model optimization, and the validation set for performance evaluation and generalization assessment. The Adam optimizer was adopted with a learning rate of 1×10^{-3}, batch size of 4, and 30 epochs of training. Data augmentation

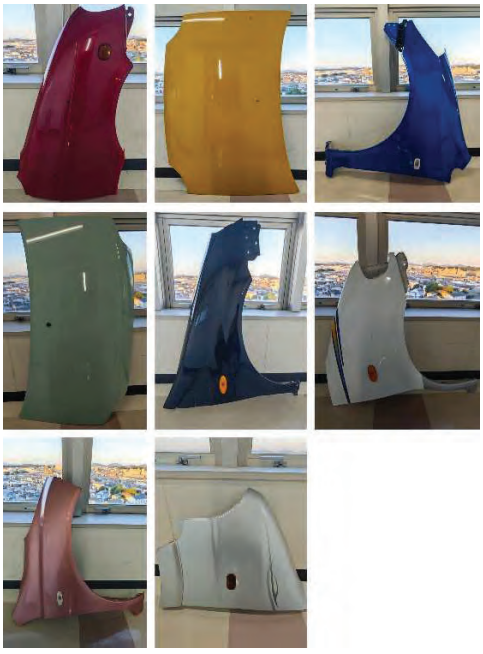

Fig. 7 Examples of car body surface samples

techniques including rotation, brightness perturbation, and slight affine transformations were applied to enhance the model's robustness to reflective noise and viewing-angle variations.

3.2 Dent center detection results

In the first-stage detection, the HmNet network performs localization by generating a heatmap of dent centers. To evaluate detection accuracy, the peak positions of the predicted heatmap were matched with the manually annotated dent centers. A detection was considered correct when the Euclidean distance between the predicted and true centers was less than a threshold of 100 pixels.

To verify the improvement over our previous detection systems, the proposed method was compared with the Mask R-CNN-based dent detection system reported in our earlier work[7]. The previous method achieved a Precision of 0.91 and Recall of 0.86 on the same dataset, while the proposed HmNet improved these values to 0.9854 and 0.9011, respectively, demonstrating better detection accuracy and robustness, and an overall F1-score of 0.9413 on the validation set. Robustness was evaluated by testing the model on car body panels of eight different colors with varying reflective properties. The detection accuracy remained above 94% across these conditions, indicating stable dent localization performance under different surface colors and reflection levels.

As shown in **Fig. 8**, the figure presents the dent center detection results for car body surfaces of eight different colors. Each row corresponds to one color sample. The left column shows the original images captured by the camera, and the right column displays the detection results overlaid on the grayscale version of the original image. In the result images, pixels with stronger red intensity indicate a higher probability of being the dent center. It can be observed that the bright red regions closely overlap with the ground-truth dent centers. HmNet accurately captures local deformations in the structured-light lattice across various car body colors and remains robust even under reflective conditions. Notably, it maintains stable probability responses on light-colored surfaces such as white, silver, and beige, where surface reflectivity and ambient light color are highly similar.

Table2 summarizes the dent center detection performance of the proposed HmNet network across car body surfaces of eight different colors: black, white, red, green, blue, yellow, beige, and silver. The table lists the number of ground-truth dents (GT), true positives (TP), false positives (FP), false negatives (FN), and the corresponding evaluation metrics

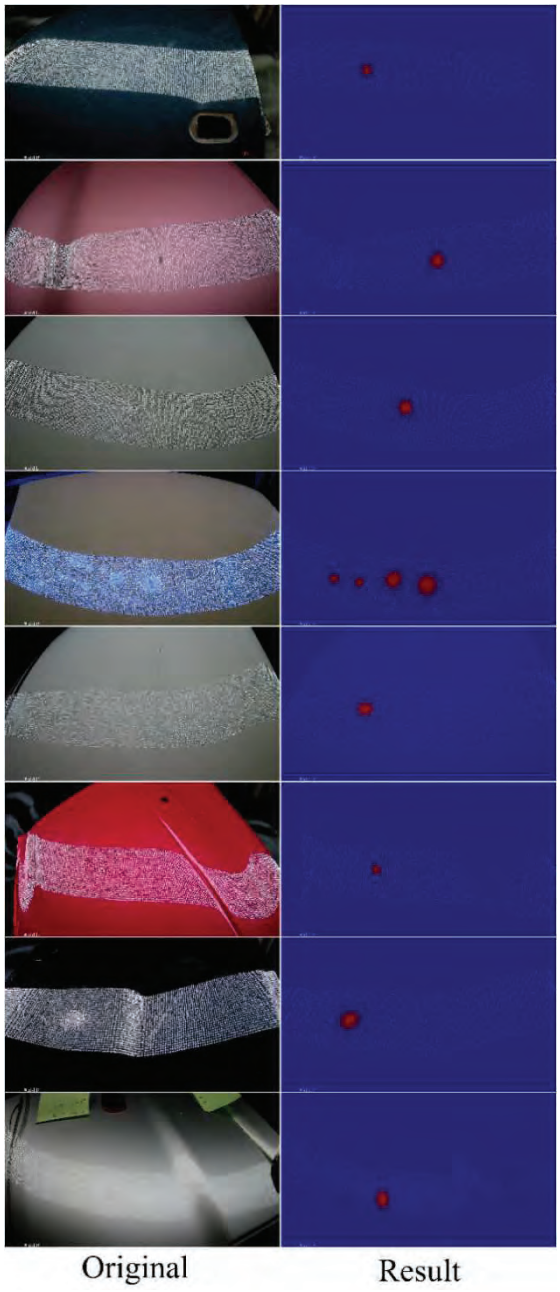

Fig. 8 Dent center detection results for eight colors

Table 2 Dent center detection performance of HmNet on car body surfaces with different colors.

Color	GT	TP	FP	FN	P	R	F1
black	33	33	0	0	1	1	1
white	33	33	0	0	1	1	1
red	59	57	0	2	1	0.9661	0.9828
green	32	29	3	3	0.9063	0.9063	0.9063
blue	32	32	0	0	1	1	1
yellow	112	82	2	30	0.9762	0.7321	0.8367
beige	40	39	0	2	1	0.9512	0.9750
silver	32	32	0	0	1	1	1

including Precision (P), Recall (R), and F1-score (F1) for each color category. The results show that the network achieves perfect detection on most color surfaces, while slight performance degradation is observed for the green and yellow samples. The lower accuracy in green surfaces is mainly due to weaker structured-light reflections, which make the lattice deformation less distinguishable in the images. For yellow surfaces, the dense distribution of dents leads to overlapping deformation regions, making lattice recognition more difficult. Overall, the average detection accuracy remains above 94%, demonstrating the robustness of the proposed method under different color and reflective conditions.

3.3 Dent parameter regression results

In the second-stage geometric parameter regression, the DentRegressor network predicts each dent's maximum depth $d_{\max}$ and characteristic diameter D_{50} using the ROI image cropped around the detected dent center. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were employed as evaluation metrics to quantify regression accuracy.

Experimental results on the validation set show that the depth prediction achieved RMSE of 0.267 mm and MAE of 0.198 mm, while the diameter prediction achieved RMSE of 13.959 mm and MAE of 8.587 mm.

In automotive surface inspection applications, dents with depths on the order of 0.1–1.0 mm are generally considered critical for quality evaluation in remanufacturing processes. Therefore, measurement systems are typically required to achieve sub-millimeter accuracy to reliably capture such geometric variations.

In this study, the proposed method achieved a depth prediction accuracy of RMSE to be 0.267 mm and MAE to be 0.198 mm. Considering that the measurement error is significantly smaller than the characteristic scale of typical dents, the system is capable of effectively distinguishing meaningful geometric differences. These results indicate that the proposed method satisfies practical accuracy requirements for micro-dent inspection on automotive body panels.

As shown in **Fig. 9**, the scatter plots of predicted versus measured values show that most points lie close to the diagonal, indicating a high degree of consistency between predictions and ground truth. The vertical axis represents the dent diameter (mm), and the horizontal axis represents the dent depth (mm). However, both the predicted diameters D_{50} and depths $d_{\max}$ tend to be slightly smaller than the measured values. The mean difference between predicted and measured depth values is −0.103 mm while the mean

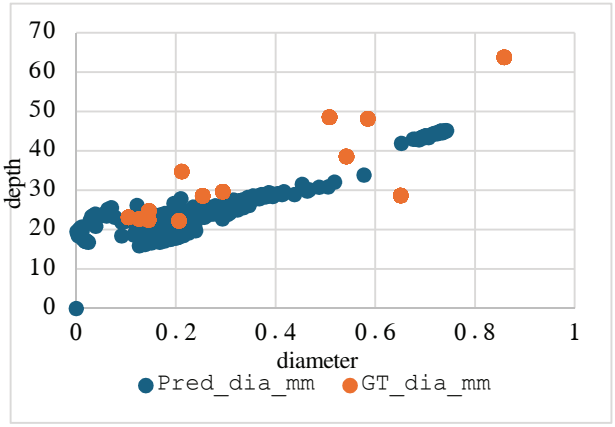

Fig. 9 Scatter plots of predicted versus measured dent parameters

difference for diameter is −8.16 mm, indicating that the predictions are slightly smaller than the measured values. This discrepancy may arise from two factors: (1) measurement errors during data acquisition, particularly for dents with blurred boundaries or strong reflections, where diameters are often overestimated; and (2) subjective variations in manual annotations, as the network tends to focus on the high-response regions near the dent center, leading to slightly smaller overall predictions.

4. Conclusion

This paper presented a novel two-stage detection and regression framework for surface dent inspection of automotive body panels using structured-light imaging and deep neural networks. The proposed system effectively addresses the limitations of traditional Mask R-CNN and YOLO methods, which struggle with blurred boundaries, low precision for small dents, and heavy computation on reflective surfaces.

In the first stage, the HmNet achieves sub-pixel localization of dent centers through anchor-free keypoint detection, enabling stable performance even under complex lighting and reflective conditions. In the second stage, the DentRegressor network directly predicts dent depth and diameter from local structured-light deformations, replacing pixel-level segmentation with efficient global regression. This approach allows the system to perform both detection and quantitative measurement within a single structured-light image.

Experimental results demonstrate that the proposed method achieved an average Precision of 0.9854, Recall of 0.9011, and an overall F1-score of 0.9413 on the validation set. The overall detection accuracy remained above 94%, indicating that the model exhibits strong robustness and stability under different surface colors and reflective

conditions. For geometric parameter estimation, the proposed DentRegressor network achieved RMSE of 0.267 mm and MAE of 0.198 mm in depth prediction, and RMSE of 13.959 mm and MAE of 8.587 mm in diameter prediction. These results satisfy the accuracy requirements for micro-dent measurement on automotive body panels and further confirm the feasibility of the proposed system for industrial inspection applications.

Overall, the proposed method combines the optical sensitivity of structured light with the adaptive learning capability of neural networks, realizing a compact, accurate, and real-time solution for non-contact surface inspection. Future research will focus on integrating real-time video analysis, multi-view fusion, and adaptive illumination control to further improve the generalization and automation of the detection system.

References

1) A. Patel, S. Singh: “Implementing Circular Economy Strategies in the Automobile Industry – A Step toward Creating Sustainable Automobiles,” Benchmarking: An International Journal, Vol.30, No.7, pp.2225–2233 (2023).
2) Z. Ren, F. Fang, N. Yan, Y. Wu: “State of the Art in Defect Detection Based on Machine Vision,” International Journal of Precision Engineering and Manufacturing-Green Technology, Vol.9, pp.661–691 (2022).
3) J. Xu, J. Douet, J. Zhao, L. Song, K. Chen: “A Simple Calibration Method for Structured Light-Based 3D Profile Measurement,” Optics & Laser Technology, Vol.48, pp.187–193 (2013).
4) G. Wang, Z. Li, G. Weng, Y. Chen: “An Overview of Industrial Image Segmentation Using Deep Learning Models,” Intelligence & Robotics, Vol.5, No.1, pp.143–180 (2025).
5) L. Yang, Y. Liu, J. Peng: “Advances Techniques of the Structured Light Sensing in Intelligent Welding Robots: A Review,” International Journal of Advanced Manufacturing Technology, Vol.110, pp.1027–1046 (2020).
6) S. Wang, Y. Ying, C. Lu: “An Image Measurement System for Detecting Dents and Scratches on the Surface of Used Car Body Parts,” Proc. of 8th International Conference on Business and Industrial Research (ICBIR), pp.543–548 (2023).
7) S. Wang, J.-W. Wang, C. Lu: “Optimization of Mask R-CNN Based Auto Parts Defect Detection System with Image Enhancement,” Proc. of 4th International Joint Conference on Robotics and Artificial Intelligence (JCRAI '24), pp.111–114 (2025).
8) K. He, G. Gkioxari, P. Dollár, R. Girshick: “Mask R-CNN,” Proc. of IEEE International Conference on Computer Vision (ICCV), pp.2961–2969 (2017).
9) K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, Q. Tian: “CenterNet: Keypoint Triplets for Object Detection,” Proc. of IEEE/CVF International Conference on Computer Vision (ICCV), pp.6569–6578 (2019).

(Received November 12, 2025)
(Revised May 25, 2026)

Sikun WANG (*Member*)

He received the master’s degree in engineering from Fukuoka Institute of Technology, Japan. He is currently pursuing the Ph.D. degree at Fukuoka Institute of Technology. His research interests include computer vision, deep learning, image processing, industrial visual inspection, surface defect detection, and practical applications of artificial intelligence. His recent research focuses on AI-based visual measurement methods for automotive body surface dent detection and fish counting in complex aquaculture environments.

Cunwei LU

He received the bachelor's degree in 1983 and the master's degree in 1988 from Shandong University of Science and Technology, China, and the Ph.D. degree from Osaka University in 1999. From 1983, he served as an Assistant Professor, Lecturer, and Associate Professor at Shandong University of Science and Technology. In 1999, he joined Fukuoka Institute of Technology as an Associate Professor and was promoted to Professor in 2005, a position held to the present. His research interests include 3D image measurement, pattern recognition, and practical applications of artificial intelligence.

Call for Papers

Special Issue on “Extended Papers Presented in IEVC2026”

IIEEJ Editorial Committee

The 9th IIEEJ International Conference on Image Electronics and Visual Computing (IEVC2026) was successfully held in Hiroshima, Japan, from March 16 to 19, 2026. The conference brought together researchers, engineers, developers, and students from academia and industry to present and discuss the latest advances in image electronics and visual computing across a wide range of fields.

Following the success of IEVC2026, the IIEEJ Transactions on Image Electronics and Visual Computing is pleased to announce a special issue entitled “Extended Papers Presented in IEVC2026.” This special issue aims to publish high-quality, extended versions of the papers presented at IEVC2026. Authors are invited to submit significantly enhanced manuscripts that expand upon their conference papers to include substantial new contents, such as additional experiments, deeper analysis, improved methodologies, extended discussions, and so on. Accepted papers will be published in either Vol.14, No.2 (December 2026 issue) or Vol.15, No.1 (June 2027 issue), depending on the submission timing, review process, and scheduling issues.

1. Submission Guidelines:

- Submissions must be extended versions of the papers accepted and presented at IEVC2026, for either in the General Papers (oral and poster presentations) category or in Late Breaking Papers (LBP) category. In particular, for papers accepted in the General Papers category, the submitted manuscript must significantly extend the version included in the proceedings to be published in the IEEE Xplore Digital Library (generally at least 30% new material is recommended).
- Manuscripts must follow the official formatting and submission guidelines of the journal.
- Detailed guidelines regarding copyright and how to cite the IEVC2026 proceedings will be prepared before the submission deadline, and provided individually. Therefore, authors are requested to contact the IIEEJ editorial office via the email address listed in Section 5 prior to submission.

2. Review Process:

Submitted manuscripts will undergo the standard double-blind peer review process of the journal.

3. Publication of Special Issue:

IIEEJ Transactions on Image Electronics and Visual Computing
Vol. 14, No.2 (December 2026 issue) or Vol. 15, No. 1 (June 2027 issue)

4. Submission Deadline:

- For publication in Vol. 14, No. 2 (December 2026 issue): Friday, June 12, 2026
- For publication in Vol. 15, No. 1 (June 2027 issue): Friday, October 30, 2026

5. Contact details for Inquiries: IIEEJ editorial office E-mail: hensyu@iieej.org

6. Online Submission URL: http://www.editorialmanager.com/iieej/

Call for Papers
Special Issue on
Image-Related Technology for Social Contribution

IIEEJ Editorial Committee

In recent years, global environmental issues, declining birthrates and increasing aging populations, disaster prevention and mitigation, and the correction of regional disparities have become increasingly serious, diverse, and complex. Amid these challenges, advances in imaging and video technologies, sensing technologies, AI, robotics, drone utilization, and data analytics have led to expanded applications of imaging and electronic technologies. These technologies are now being utilized not only to solve problems and improve efficiency in various fields—such as disaster prevention and mitigation, medical and welfare support, education and cultural promotion, environmental conservation, and regional revitalization—but also to create new forms of social value. At the same time, when implementing such technologies in modern society, it is necessary to consider aspects such as fairness, ethics, and privacy. Efforts are therefore required to rebuild the relationship between technology and society in a desirable and sustainable manner.

In this special issue, we invite various categories (general paper, short paper, system development paper, data paper, practice-oriented paper) of papers under the theme of "Social Contribution." We would be grateful if you could submit your papers on this special issue.

1. Topics covered include but not limited to

   Regional Revitalization, Regional Problem Solving, Disaster Prevention, Disaster Recovery, Education, Medicine, Healthcare, Home Life, Welfare, Safety and Security Technologies, Crisis Management, Wearable Devices, VR/AR, Smartphone Applications, Universal Design, Affective design, Usability, UI/UX, Ergonomics, Accessibility, AI Utilization, Automation, Robotics, Drone Utilization, Data Analysis, Digital Archiving

2. Treatment of papers

   Submission paper style format and double-blind peer review process are the same as the regular paper. If the number of accepted papers is less than the minimum number for the special issue, the acceptance paper will be published as the regular contributed paper. We ask for your understanding and cooperation.

3. Publication of Special Issue:

   IEEJ Transactions on Image Electronics and Visual Computing Vol.15, No.1 (June 2027)

4. Submission Deadline:

   Friday, October 30, 2026

5. Contact details for Inquires:

   IIEEJ Office E-mail: hensyu@iieej.org

6. Online Submission

   URL: http://www.editorialmanager.com/iieej/

Revised: January 6, 2017

Revised: July 6, 2018

Revised: Dec. 10, 2024

Guidance for Paper Submission

1. Submission of Papers

(1) Preparation before submission

- The authors should download "Guidance for Paper Submission" and "Style Format" from the "Academic Journals", "English Journals" section of the Society website and prepare the paper for submission.
- Two versions of "Style Format" are available, TeX and MS Word. To reduce publishing costs and effort, use of TeX version is recommended.
- There are four categories of manuscripts as follows:
  - Ordinary paper: It should be a scholarly thesis on a unique study, development or investigation concerning image electronics engineering. This is an ordinary paper to propose new ideas and will be evaluated for novelty, utility, reliability and comprehensibility. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Short paper: It is not yet a completed full paper, but instead a quick report of the partial result obtained at the preliminary stage as well as the knowledge obtained from the said result. As a general rule, the authors are requested to summarize a paper within four pages.
  - System development paper: It is a paper that is a combination of existing technology or it has its own novelty in addition to the novelty and utility of an ordinary paper, and the development results are superior to conventional methods or can be applied to other systems and demonstrates new knowledge. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Data Paper: A summary of data obtained in the process of a survey, product development, test, application, and so on, which are the beneficial information for readers even though its novelty is not high. As a general rule, the authors are requested to summarize a paper within eight pages.
  - Survey Paper: A summary of existing Research and Developments, organized under some viewpoint, compared for the sake of positioning purpose, observed as the changes in generations. Comprehensive references, overall perspective, objective evaluation, are needed without advertising specific organizations. It is also appreciated that the status and problems of the field, and the effect of them to the researchers and concerned people are undersood by the author, and the resulant paper encourages the new entry into the field, accelerates further development of related technologies, and prompts the development in even other fields or brand new researches. As a general rule, the authors are requested to summarize a paper within eight pages.

- To submit the manuscript for ordinary paper, short paper, system development paper, or data paper, at least one of the authors must be a member or a student member of the society.
- We prohibit the duplicate submission of a paper. If a full paper, short paper, system development paper, or data paper with the same content has been published or submitted to other open publishing forums by the same author, or at least one of the co-authors, it shall not be accepted as a rule. Open publishing forum implies internal or external books, magazines, bulletins and newsletters from government offices, schools, company organizations, etc. This regulation does not apply to a preliminary draft to be used at an

annual meeting, seminar, symposium, conference, and lecture meeting of our society or other societies (including overseas societies). A paper that was once approved as a short paper and being submitted again as the full paper after completion is not regarded as a duplicate submission.

(2) Submission stage of a paper

- Delete all author information at the time of submission. However, deletion of reference information is the author's discretion.
- At first, please register your name on the paper submission page of the following URL, and then log in again and fill in the necessary information. Use the "Style Format" to upload your manuscript. An applicant should use PDF format (converted from dvi of TeX or MS Word format) for the manuscript. As a rule, charts (figures and tables) shall be inserted into the manuscript to use the "Style Format". (a different type of data file, such as audio and video, can be uploaded at the same time for reference.)

  http://www.editorialmanager.com/iieej/
- If you have any questions regarding the submission, please consult the editor at our office.

Contact:
Person in charge of editing
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-Ku, Tokyo 116-0002, Japan
E-mail: hensyu@iieej.org
Tel: +81-3-5615-2893, Fax: +81-3-5615-2894

2. Review of Papers and Procedures

(1) Review of a paper

- A manuscript is reviewed by professional reviewers of the relevant field. The reviewer will deem the paper "acceptance", "conditionally acceptance" or "returned". The applicant is notified of the result of the review by E-mail.
- Evaluation method

  Ordinary papers are usually evaluated on the following criteria:
  - ✓ Novelty: The contents of the paper are novel.
  - ✓ Utility: The contents are useful for academic and industrial development.
  - ✓ Reliability: The contents are considered trustworthy by the reviewer.
  - ✓ Comprehensibility: The contents of the paper are clearly described and understood by the reviewer without misunderstanding.

A short paper can be evaluated by having a quickness on the research content and evaluated to have new knowledge with results even if that is partial or for specific use, apart from the novelty and utility of an ordinary paper.

A system development paper is evaluated based on the following criteria, apart from the novelty and utility of an ordinary paper.

- ✓ Novelty of system development: Even when integrated with existing technologies, the novelty of the combination, novelty of the system, novelty of knowledge obtained from the developed system, etc. are recognized as the novelty of the system.
- ✓ Utility of system development: It is comprehensively or partially superior compared to similar systems. Demonstrates a pioneering new application concept as a system. The combination has appropriate optimality for practical use. Demonstrates performance

limitations and examples of performance of the system when put to practical use.

A data paper is considered novel if new deliverables of test, application and manufacturing, the introduction of new technology and proposals in the worksite have any priority, even though they are not necessarily original, apart from the novelty and utility of an ordinary paper. Also, if the new deliverables are superior compared to the existing technology and are useful for academic and industrial development, they should be evaluated.

A survey paper is evaluated by comprehensiveness, overviewing point, and objectiveness apart from the novelty of an ordinary paper. Reliability, comprehensibility, completeness of reference papers are common to those in an ordinary paper. Utility is evaluated how the paper will enlighten the readers in the target fields.

(2) Procedure after a review

- In case of acceptance, the author prepares a final manuscript (as mentioned in 3.).
- In the case of acceptance with comments by the reviewer, the author may revise the paper in consideration of the reviewer's opinion and proceed to prepare the final manuscript (as mentioned in 3.).
- In case of conditional acceptance, the author shall modify a paper based on the reviewer's requirements by a specified date (within 60 days), and submit the modified paper for approval. The corrected parts must be colored or underlined. A reply letter must be attached that carefully explains the corrections, assertions and future issues, etc., for all of the acceptance conditions.
- In case a paper is returned, the author cannot proceed to the next step. Please look at the reasons the reviewer lists for the return. We expect an applicant to try again after reviewing the content of the paper.

(3) Review request for a revised manuscript

- If you want to submit your paper after conditional acceptance, please submit the reply letter to the comments of the reviewers, and the revised manuscript with revision history to the submission site. Please note the designated date for submission. Revised manuscripts delayed more than the designated date be treated as new applications.
- In principle, a revised manuscript willl be reviewed by the same reviewer. It is judged either acceptance or returned.
- After the judgment, please follow the same procedure as (2).

3. Submission of final manuscript for publication

(1) Submission of a final manuscript

- An author, who has received the notice of "Acceptance", will receive an email regarding the creation of the final manuscript. The author shall prepare a complete set of the final manuscript (electronic data) following the instructions given and send it to the office by the designated date.
- The final manuscript shall contain a source file (TeX edition or MS Word version) and a PDF file, eps files for all drawings (including bmp, jpg, png), an eps file for author's photograph (eps or jpg file of more than 300 dpi with length and breadth ratio 3:2, upper part of the body) for authors' introduction. Please submit these in a compressed format, such as a zip file.
- In the final manuscript, write the name of the authors, name of an organizations, introduction of authors, and if necessary, an appreciation acknowledgment. (cancel macros in the Style file)

- An author whose paper is accepted shall pay a page charge before publishing. It is the author's decision to purchase offprints. (ref. page charge and offprint price information)

(2) Galley print proof

- The author is requested to check the galley (hard copy) a couple of weeks before the paper is published in the journal. Please check the galley by the designated date (within one week). After making any corrections, scan the data and prepare a PDF file, and send it to our office by email. At that time, fill in the Offprint Purchase Slip and Copyright Form and return the scanned data to our office in PDF file form.
- In principle, the copyrights of all articles published in our journal, including electronic form, belong to our society.
- You can download the Offprint Purchase Slip and the Copyright Form from the journal on our homepage. (ref. Attachment 2: Offprint Purchase Slip, Attachment 3: Copyright Form)

(3) Publication

- After final proofreading, a paper is published in the Academic journal or English transaction (both in electronic format) and will also be posted on our homepage.

IIEEJ Trans. on Image Electronics and Visual Computing

Vol. 14, No. 1 2026

Editor in Chief: Osamu Uchida
The Institute of Image Electronics Engineers of Japan
3-35-4-101, Arakawa, Arakawa-ku, Tokyo 116-0002, Japan

Print: ISSN 2188-1898
Online: ISSN 2188-1901
CD-ROM: ISSN 2188-191x