RESEARCH ARTICLE

SGVM: Semantic-Guided Variational Model for Sealing Nail Defect Extraction Within Albedo Domain via Photometric Stereo

FANG LIU[1], WEI CAO[2,3], (Member, IEEE), YUPING YE[2,3], FEIFEI GU[2,3], SHIYANG LONG[2,3], AND ZHAN SONG[2,3], (Member, IEEE)

[1]Logistics Academy, Xuzhou 221000, China
[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]Guangdong-Hong Kong-Macao Joint Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen 518055, China

Corresponding author: Zhan Song (zhan.song@siat.ac.cn)

ABSTRACT Automatic 2D vision-based defect detection on sealing nail (SealN) surfaces is challenging due to interference of complex backgrounds with non-homogeneous and low contrast between foreground and background. Inspired by an interesting observation that the albedo domain recovered by the uncalibrated photometric stereo (UPS) shows obvious differences and significant abruptness between defects' and non-defects' regions, we develop a novel semantic-guided variational model (SGVM) to conditional extract structural defects from albedo map. Specifically, SGVM utilizes one developed global regularized label indicator to semantically guide one local regularized relative Gaussian filter (RGF) for achieving large-scale structures (i.e., defects) preservation and small-scale textures (i.e., background) suppression. Furthermore, defects can be efficiently extracted by thresholding the structure map within the label indicator. Additionally, experimental results on numerous challenging defect images reveal that the proposed SGVM outperforms the existing advanced 2D methods in terms of defect extraction.

INDEX TERMS Semantic-guided variational model (SGVM), defect extraction, sealing nail, albedo domain, uncalibrated photometric stereo (UPS).

I. INTRODUCTION

Undoubtedly, surface defect extraction (DE) is crucial for achieving high yield output of massive industrial products such as power battery SealN, bearings, and steel strips. With the demand for high quality- and efficiency- advanced vision-based autonomous defect detection technology has become mainstream and replaced traditional manual inspection methods, which are less efficient. Generally, these inspection technologies can be divided into three categories according to the different defect extraction domains: spatial, spectral, and feature-based methods.

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

Directly measuring the spatial distribution of pixel intensities in the spatial domain [1], [2], [3], [4] is probably the most intuitive and straightforward way to distinguish prominent defects from textures. Tsai et al. [1] designed a weighted covariance matrix with a texture regularity measure, and used its eigenvalues to determine whether a local image block contains defects. Afshar et al. [2] proposed the rotation-invariant variance measure, which was employed for defect extraction from smooth tiles. Additionally, Luo et al. [3] proposed an adaptive dual-threshold technique to detect each non-overlapping pixel block expressed by texture operators (including variance, entropy, and average gradient) on steel surface images. Spatial domain methods mentioned above, despite their good ability to extract local/global texture descriptors representing different classes

from the original intensity, may have limitations in handling small defects and being ineffective in cases where defects and background textures are similar.

The core idea of spectral domain methods is to transform the intensity domain into the frequency domain, on which defects are identified through the differences in response between foreground targets and background. For instance, Aiger et al. [5] proposed a phase-only Fourier transform method, which uses Fourier transform phase information to extract defects. Building upon this method, Zhou et al. [6] further improved its performance by introducing two-step operations involving local estimation and spatiotemporal refinement. Furthermore, Ghorai et al. [7] presented a wavelet decomposition model that utilizes adaptive scaling and directional texture features to achieve better performance on steel surface defect detection. Obviously, such methods are particularly suitable for target surfaces with regular patterns or periodic texture features, such as fabrics, tiles, and steel materials, but the use of inappropriate filters and their radius may lead to the risk of under/over-extracting of defects.

For the feature domain-based methods, saliency defect detection [8], [9], [10], [11] has outstanding performance as the main representative. Saliency feature models simulate the human visual attention mechanism, which captures salient visual features in an image and treats them as defect collections. Achanta et al. [8] proposed a frequency-tuned algorithm that effectively perceives salient object edges and produces high-resolution saliency maps. Subsequently, Peng et al. [9] introduced a saliency defect detection model based on low-rank matrix recovery theory, representing the image matrix as a sparse matrix containing salient objects and a low-rank matrix containing the background through pixel-wise addition composition. In particular, Song et al. [11] regarded defective objects as salient objects in the image and effectively extracted them by proposing multiple constraint terms and improved texture features. Even though those saliency methods can provide promising results, they may encounter under/over-extraction and blurred boundaries on saliency maps when multiple or scattered target defects are simultaneously extracted or when there are similar representations between salient target regions and the background.

To overcome the above issues, inspired by the 2.5D UPS method [12], [13], [14], [15], we propose a novel Semantic-Guided Variational Model (SGVM) on the specified albedo domain estimated by UPS and apply it to the SealN defective extraction task with complex background interference. Unlike the different domain methods mentioned above, the albedo domain estimated by UPS exhibits inherent strong differences in intensity values and significant discontinuity or abruptness in spatial distribution between the defect and surrounding background regions. *In other words, within the albedo domain, the inter-class difference between defects and background is amplified.* On the other hand, the proposed SGVM incorporates one semantic label

indicator as a global regularization to guide local regularized RGF for identifying structures (regarded as defects) and suppressing unwanted complex background textures at any scale, addressing issues such as under/over-smoothing and blurred boundaries. Furthermore, benefiting from the joint prior variational framework with global optimization, the SGVM shows well generalization ability on multiple defect datasets. The overall framework of the proposed method is shown in Fig. 1. The **main contributions** of this paper can be summarized as follows.

1. To our knowledge, we are the first to introduce the albedo map estimated by the UPS method as the processing domain for SealN's defect extraction tasks, rather than the intensity domain that captures original images, and thus obtain a better defect extraction.

2. We propose one globally optimal variational model, namely SGVM, for defect detection, which combines one scale-smoothness prior constrained by the L1 norm and one semantic-guided label prior constrained by the L0 norm. This model extracts structural maps containing defects without under or over-smoothing and boundary artifacts from challenging datasets such as power battery SealN.

3. We provide a numerical approximation term to solve the non-convex problem in the SGVM model, i.e., transforming it into a convex problem that can be iteratively solved by decomposing it into subproblems.

4. We perform comprehensive experiments on challenging examples, such as SealN images with interference from non-homogeneous background textures and other scenes with low contrast between defects and the background ones. Experimental results demonstrate that the proposed model not only outperforms advanced models in defect extraction performance, but also exhibits good generalizability in various complex task scenarios, such as bridge cracks, bearing defects, and steel strip defects, in terms of large-scale structural defect extraction.

II. RELATED WORK
A. PRELIMINARIES
The input and output (or say intermediate) signals are denoted as $P_0 \in \mathbb{R}^{m \times n}$ and $P \in \mathbb{R}^{m \times n}$, respectively. Furthermore, the first-order derivative filter $\nabla = [\nabla_h; \nabla_v] \in \mathbb{R}^{2m \times n}$, which includes two directions: $\nabla_h \in \mathbb{R}^{m \times n}$ (horizontal direction) and $\nabla_v \in \mathbb{R}^{m \times n}$ (vertical direction). For terse expression, we let pixel coordinates be $i = (x, y)^T$, i.e., P_i means the pixel at the $(x, y)^T$.

B. RECOVERY OF INITIAL ALBEDO MAP P_0 FROM UPS
In the sequence (with ϖ in total), we denote an image sequence captured by UPS as $I_i = [I_i^1, \ldots, I_i^j, \ldots, I_i^\varpi]$, where I_i^j means the intensity at pixel i in the jth image, with $j \in [1, \varpi]$. Likewise, we use symbols ρ_i and $N_i = [N_x, N_y, N_z](i)$ to represent the albedo and surface normal at pixel i, with $\rho_i \in [0, 1]$, respectively. Based on the Lambertian model assumption [16], one photometric stereo equation for every
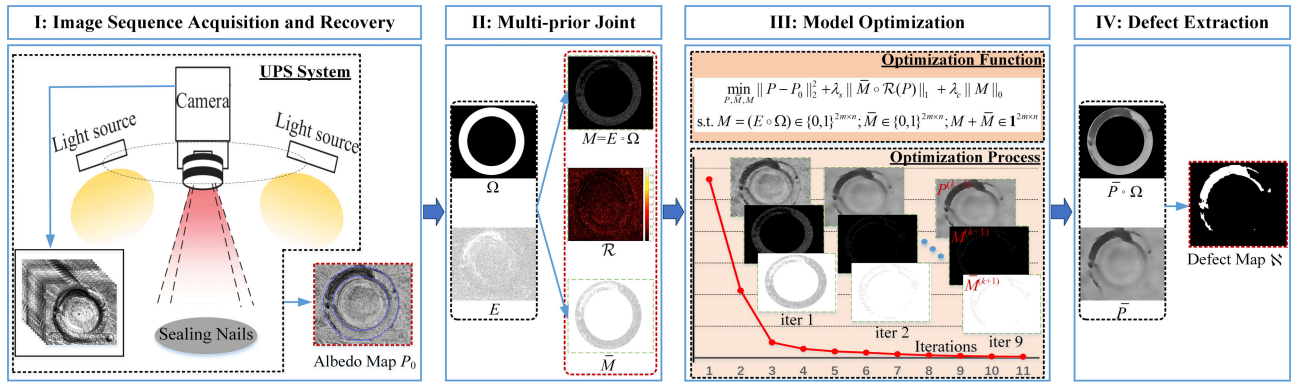
FIGURE 1. Defect extraction framework via the proposed SGVM within albedo domain. The entire framework is implemented by using the albedo map P_0 (as input) recovered by the UPS system in step I and Our SGVM in steps II-IV: firstly, we use UPS to acquire image sequence and recover albedo map P_0; then, we design a novel SGVM including multi-prior ($\mathcal{R}$, $\overline{M}$, and M) regularization to extract defects' structure within albedo domain P_0; finally, having the well-extracted structure map $\overline{P}$, the defect map $\aleph$ can be obtained accordingly by thresholding operation. The blue arrow indicates data flow.

pixel i and every image j can be expressed as:

$$I_i = \rho_i N_i S = \Pi_i S, \tag{1}$$

where $S = [S^1, \ldots, S^j, \ldots, S^\varpi]$ represents the lighting matrix and $S^j = [S_x^j, S_y^j, S_z^j]^T$. In previous works, a problem with $\varpi \geq 3$ and unknown non-coplanar distance light sources is referred to as the UPS problem. Particularly, when S is unknown, Quéau et al. [12], [16] designed the following optimization:

$$\begin{cases} (\widehat{\Pi}, \widehat{S}) = \underset{(\Pi,S)}{\arg\min} \sum_{p \in \Gamma} ||\Pi_i S - I_i||^2 \\ s.t. \quad \overline{curl}(\Pi) = 0 \\ (\widehat{\mu}, \widehat{v}, \widehat{\gamma}) = \underset{(\widehat{\mu},\widehat{v},\widehat{\gamma})}{\arg\min} TV(\widehat{\Pi}\Upsilon(\mu, v, \gamma)) \end{cases} \tag{2}$$

to simultaneously estimate the resulting field $\widehat{\Pi}$ and the final resulting lighting matrix $\widehat{S}$ in the UPS problem. In Eq. (2), the symbol Γ represents a mask within the scene in the image, the operator $\overline{curl}(\cdot)$ indicates integrability constraint, and three regularization parameters μ, v and γ are used for minimization operation. The term Υ represents the generalized bas-relief transformation, and it is addressed using the total variation function $TV(\cdot)$ defined by the work [12].

Naturally, by the results estimated by the optimization above, the corresponding albedo at pixel i can be expressed as

$$\widehat{\rho_i} = ||\widehat{\Pi_i}||, \tag{3}$$

and refer to [12], [16], [17], [18], [19], and [20] for more details. It should be noted that in our work, we employ the above UPS method proposed by Quéeau [12] to obtain the initial albedo map P_0 and replace the symbol $\widehat{\rho}$ with it.

C. PRIOR REGULARIZATION ABOUT RGF
To identify the variance at different scales and achieve image smoothing through the gradient minimization of relative

scales, Cai et al. [21] developed the following cost function in RGF:

$$\underset{P}{\min} \ ||P - P_0||_2^2 + \lambda_s ||\mathcal{R}(P)||_1 \tag{4}$$

for optimizing the smoothing result P, which includes a data fidelity term in terms of L2 norm between P and initial signal P_0 and the prior regularization term

$$\mathcal{R}(P) = \left| \frac{G_{\sigma_1} * \nabla P}{G_{\sigma_2} * \nabla P} \right| \quad s.t. \quad \sigma_1 < \sigma_2 \tag{5}$$

which is constrained by the L1 norm for selective smoothing on ∇P. Here, the defined local Gaussian kernel $G_\sigma(p) = exp\left(-\frac{1}{2\sigma}\left((x - x_0)^2 + (y - y_0)^2\right)\right)$ using the scale parameter σ and the center (x_0, y_0) of the kernel is for scale selection. And the parameter λ_s controls the strength of scale-based smoothing in the gradient domain. For more details, please refer to the work [21].

III. METHODOLOGY
A. MOTIVATION
Similar to RGF, a popular strategy in image processing is to use L1 norm constraints with various priors to achieve specified structural band-pass and texture band-stop. Although this approach makes optimization tractable and produces reasonable output, the scaling issue still exists, indirectly causing elements with large magnitudes to be over-penalized. In other words, when directly applying L1 norm constraints to defect extraction, meaningful structures (or say defects) may risk being indiscriminately over-smoothed or their boundaries blurred. To alleviate this issue, it is necessary to design a high-performance optimizer that can discriminate between large-scale structures (or say emphasized defects) and small-scale structures (or say background to be removed), while preserving sharp boundaries and avoiding excessive smoothing. Inspired by the works of [12] and [21], we formulate this idea as a novel global optimization problem coupling

the L1 norm with the L0 norm in albedo domains, and will be discussed further in the following sections.

B. PROPOSED MODEL

By minimizing a fidelity term $F_f(P)$ in the albedo domain, one local regularized term $F_s(P, \bar{M})$ constrained by the L1 norm, and one semantic-guided term $F_c(P, M)$ constrained by the L0 norm, a novel semantic-guided variational model (SGVM) is proposed, with optimization objective is defined as

$$\min_{P,\bar{M},M} E(P, \bar{M}, M) = \min_{P,\bar{M},M} \{F_f(P) + F_s(P, \bar{M}) + F_c(M)\}$$
$$= \min_{P,\bar{M},M} ||P - P_0||_2^2 + \lambda_s ||\bar{M} \circ \mathcal{R}(P)||_1 + \lambda_c ||M||_0$$

$$s.t. \ M = (E \circ \Omega) \in \{0,1\}^{2m \times n}; \ \bar{M} \in \{0,1\}^{2m \times n}; \ M + \bar{M} \in 1^{2m \times n},$$
(6)

where $M = [M_h; M_v] \in \{0,1\}^{2m \times n}$ is a semantic-guided indicator including edge labeling E and ROI labeling Ω matrices of the same size. The parameter $\lambda_c \in [0, 1]$ controls the sparsity of the regularization term. Specific explanations of each term in the optimization objective are as follows:

- **Data fidelity term**, i.e., $F_f = ||P - P_0||_2^2$, which considers the degradation between the initial albedo value P_0 and the refined value P.

- **Local regularization term**, i.e., $F_s = \lambda_s ||\bar{M} \circ \mathcal{R}(P)||_1$, serves as a scale smoothing filter under the prior $\mathcal{R}(\cdot)$ constraint to realize structures band-pass and textures band-stop. Additionally, it incorporates semantic guidance through the prior weight $\bar{M}$, enabling selective smoothing of textures.

- **Global semantic-guided regularization term**, i.e., $F_c = ||M||_0$, one label indicator includes the edge property labeling matrix $E(P)$ (such as long links and consistent directions within a closed neighborhood) in P obtained by the method [22] and the specified ring ROI position labeling matrix Ω:

$$\Omega_i = \begin{cases} 1, & r_{small} \le \text{dist}\left(\left|i - (\lfloor o_x \rfloor, \lfloor o_y \rfloor)^T\right|\right) \le r_{big}, \\ 0, & otherwise, \end{cases}$$
(7)

where the function dist($\cdot$) measures the spatial Euclidean distance between inner pixel locations. The symbol $(\lfloor o_x \rfloor, \lfloor o_y \rfloor)^T$ is the center of the SealN ring found by the Hough transform. The operator $\lfloor \cdot \rfloor$ stands for rounding down, while r_{small} and r_{big} represent the inner and outer radius of the ring region, respectively. This term aims to identify reliable and meaningful large-scale structural E within the high-attention ROI Ω, such as defect edges, while focusing on smoothing/preserving within semantic-guided regions.

- **Joint term**, i.e., $F_s(\cdot) + F_c(\cdot)$, incorporates a mixed L0-L1 norm. This hybrid norm employs the discrete nature of the L0 norm, that is, includes truncated attributes guided by semantic labeling indicator, and thus mitigates the risk

of over-smoothing caused by solely relying on the L1 norm constraint in $\mathcal{R}(P)$.

C. SOLUTION

Considering the non-convexity of the second term in the objective function (6), we adopt the following numerical approximation

$$\left\| \frac{G_{\sigma_1} * \nabla_d P_i}{G_{\sigma_2} * \nabla_d P_i} \right\|_1 = \left\| \frac{(G_{\sigma_1} * \nabla_d P_i) \circ (G_{\sigma_1} * \nabla_d P_i)}{(G_{\sigma_2} * \nabla_d P_i) \circ (G_{\sigma_1} * \nabla_d P_i)} \right\|_1$$
$$\approx \left\| \frac{G_{\sigma_{1/2}} * (\nabla_d P_i)^2}{(G_{\sigma_2} * \nabla P_i) \circ (G_{\sigma_1} * \nabla_d P_i)} \right\|_1 \approx \|Z_i \circ \nabla_d P_i\|_2^2$$
(8)

to address it. In other words, we utilize $\|Z \circ \nabla P\|_2^2$ as a compact surrogate for the original term, which transforms the non-convex nature induced by the L1 norm into an L2 one. The operator $\circ$ represents element-wise multiplication, and $Z_i \approx \sqrt{G_{\sigma_{1/2}} * (1/(|(G_{\sigma_2} * \nabla P_i) \circ (G_{\sigma_1} * \nabla P_i)| + \varepsilon))}$ with a small positive constant ε (to prevent division by zero). Unlike the approach in [21], we scale down σ by setting the scale parameter $\sigma = \sigma/2$ during each iteration in our experiments, which aids M in discovering and preserving sharp edges during the new iteration. Benefiting from the above operations, we can reformulate the optimization objective (6) as

$$\min_{P,\bar{M},M} ||P - P_0||_2^2 + \lambda_s ||\bar{M} \circ Z \circ \nabla P||_2^2 + \lambda_c ||M||_0. \quad (9)$$

To tackle the remaining non-convexity introduced by the third term in the optimization objective (9), we further propose a numerical solution, namely an alternating subproblem decomposition algorithm, to effectively solve the overall optimization problem in Eq. (9). The details of the entire process are described as follows.

Small Textures Removal (Update $P^{(k+1)}$ while fixing $\bar{M}^{(k)}$ and $M^{(k)}$): Using the $\bar{M}^{(k)}$ from the k-th iteration, the subproblem corresponding to P in problem (9) becomes as follows:

$$\arg\min_{P} ||P - P_0||_2^2 + \lambda_s ||\bar{M}^{(k)} \circ Z^{(k)} \circ \nabla P||_2^2. \quad (10)$$

To solve the problem (10), we reformat it into vector form. Using the vectorization operator $vect(\cdot)$, we reformulate the vectors $\mathbf{p} = vect(P), \mathbf{p}_0 = vect(P_0)$, and $\bar{\mathbf{w}}^{(k)} = vect(\bar{M}^{(k)} \circ Z^{(k)})$, which are of length mn. Additionally, $\overline{T}_d$ contains $\overline{T}_h$ and $\overline{T}_v$, which is Toeplitz matrices with the discrete gradient operator of forward differences and are used to approximate the first-order difference operators ∇_d, where $d \in \{h, v\}$. Denote $D_{\bar{\mathbf{w}}_d^{(k)}} = Diag(\bar{\mathbf{w}}_d^{(k)}) \in \mathbb{R}^{mn \times mn}$ as a diagonal matrix constructed by vector $\bar{\mathbf{w}}_d^{(k)}$. Then, problem (10) can be transformed into a classic least squares regression problem:

$$\arg\min_{\mathbf{p}} ||\mathbf{p} - \mathbf{p}_0||_2^2 + \lambda_s \left\| \sum_{d \in \{h,v\}} D_{\bar{\mathbf{w}}_d^{(k)}} \overline{T}_d \mathbf{p} \right\|_2^2. \quad (11)$$

By differentiating the corresponding variables $\boldsymbol{p}$, and setting its partial derivatives to 0, we obtain the following solution form:

$$\boldsymbol{p}^{(k+1)} = \left(U + \lambda_s \left(\sum_{d \in \{h,v\}} \overline{T}_d^T D_{\overline{\boldsymbol{w}}_d^{(k)}}^T \overline{T}_d \right) \right)^{-1} \boldsymbol{p}_0 \quad (12)$$

where U is an identity matrix. We can then reformat the obtained $\boldsymbol{p}^{(k+1)}$ into matrix form by inverse vectorization $P^{(k+1)} = vect^{-1}(\boldsymbol{p}^{(k+1)})$, and update $Z^{(k+1)}$ using Eq. (8). It is worth noting that we set the final smooth output $\bar{P} = P_{k+1}$.

Algorithm 1 Exact Solver to Problem (6) in SGVM

Require: Initial albedo map P_0, parameters $\lambda_s = 0.05$ and $\lambda_e = 1$, and the iterations' stop threshold $\beta = 0.01$.
Ensure: Optimal solutions P_{k+1} and the final smooth output $\bar{P} = P_{k+1}$.
  1: **repeat**
  2:   Update P_{k+1} using Eqn. (10);
  3:   Update $M^{(k+1)}$ with P_{k+1} and Z_{k+1} using Eqns. (14) and (8);
  4:   Update $\bar{M}^{(k+1)}$ with $M^{(k+1)}$;
  5:   $k \leftarrow k + 1$;
  6: **until** $(||P^{k+1} - P^k||/P_k \leq \beta$.

Large Structures Preservation (Update $\bar{M}^{(k+1)}$ and $M^{(k+1)}$ while fixing $P^{(k+1)}$): Discarding items unrelated to $\bar{M}^{(k+1)}$ and $M^{(k+1)}$ leads to the following subproblem:

$$\underset{\bar{M},M}{arg\,min} \;\; \lambda_s ||\bar{M} \circ Z^{(k+1)} \circ \nabla P^{(k+1)}||_2^2 \; + \lambda_c ||M||_0 \quad (13)$$

Consistent with [23], the solution of (14) will be derived as

$$M_i^{(k+1)} = \begin{cases} 0, & (Z_i^{(k+1)} \circ \nabla P_i^{(k+1)})^2 < \frac{\lambda_c}{\lambda_s} \\ 1, & otherwise \end{cases} \quad (14)$$

and $\bar{M}_i^{(k+1)}$ can be also obtained accordingly. The whole iterative optimization process is summarized as Algorithm 1.
Defect Discovery: The final defect binarization map:

$$\aleph = \mathcal{T}(\bar{P} \circ \Omega) \quad (15)$$

where the operator $\mathcal{T}$ means thresholding method [24] used to achieve binarization of anisotropy in the region and highlighting of defects.

IV. EXPERIMENTS
A. EXPERIMENT SETTINGS
1) IMPLEMENTATION DETAILS
To fully evaluate the proposed method, we first compare it with other competitive 2D defect detection methods, including Salient Region Detection using Diffusion (SRDD) [25], Detection via Deformed Smoothness Constraint (DDSC) [26], Multiple Constraints and Improved Texture Features (MCITF) [11], Encoder-Decoder Residual Network (EDRNet) [27], and Normalized Attention and Dual-scale Interaction (NaDiNet) [28]. In particular, all defect recognizers are implemented using the

MATLAB2021a framework and run on a laptop with an Intel Core i7-11800H CPU and 16GB memory. Then, we test and evaluate the performance of different methods on challenging datasets primarily obtained from industrial scenarios, including SealN, Original Bridge Crack (OBriC) [29], Steel Strip Surface Defect (SteelD) [30], and Bearing Defect (BearD) [31]. Furthermore, we design a UPS system (see Fig. 2), which consists of 12 non-coplanar LED point light sources with white illumination and a web camera with a 7.4 µm telecentric lens placed 110mm away from the recovered objects. By manually triggering the camera and LEDs while keeping the recovered objects stationary, we were able to capture 12 images within 10 seconds at a resolution of 2448×2048 pixels. With captured image sequences from the UPS system, we can recover the albedo map by employing the model from the article [12]. In our experiments, the initial values for the regularization parameters λ_s and λ_c are empirically set to 0.05 and 1, respectively.
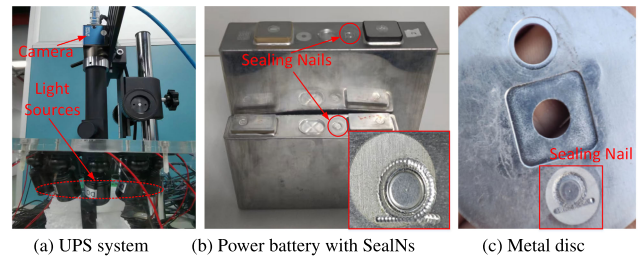

(a) UPS system   (b) Power battery with SealNs   (c) Metal disc

FIGURE 2. Experimental configuration.

(a) 2D Input   (b) DDSC   (c) MCITF   (d) EDRNet

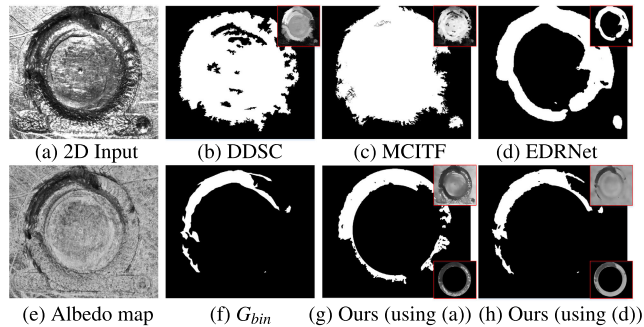(e) Albedo map   (f) G_{bin} (g) Ours (using (a))   (h) Ours (using (d))

FIGURE 3. Visual comparison of defects (after binarization) extraction between other methods within the traditional 2D image (a) and our SGVM within (a) and UPS's albedo map (d). (a) 2D Input (with uniform light and the same exposure setting in this work). (d) Albedo map (used in this work). Note that: output grayscale image in the upper-right corner and output within ROI in the lower-right corner.

2) EVALUATION METRICS
For a more comprehensive objective evaluation, we employ five commonly used metrics: mean absolute error (MAE), structure measure (SM), receiver operating characteristic (ROC) curve, precision-recall (PR) curve, and the area under the ROC curve (AUC), to assess the effectiveness of defect extraction. MAE is used to measure the degree of difference between the defect binary map $\aleph$ and the binary ground
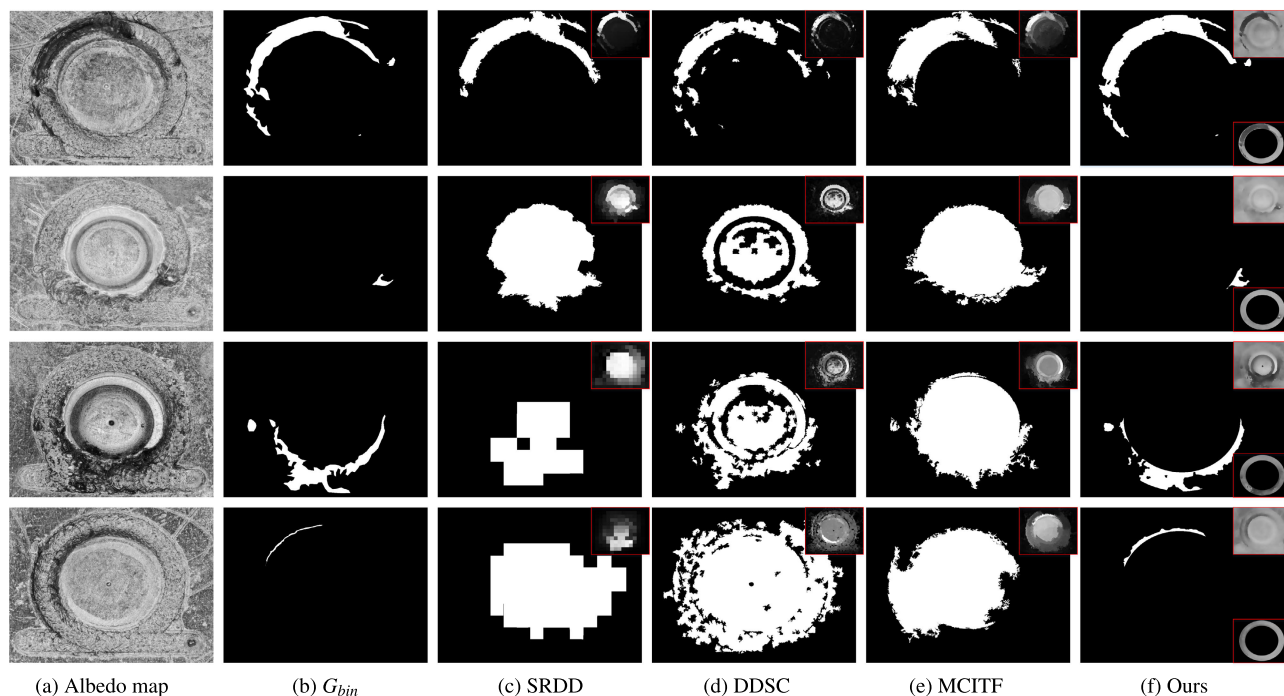
FIGURE 4. Visual comparison of defects (after binarization) extraction between different methods on samples from the SealN dataset. Note that: output grayscale image in the upper-right corner and output within ROI in the lower-right corner.

truth map G_{bin}. SM considers both region-aware and object-aware structural similarity between the predicted and ground truth maps, providing more reliable evaluation results. The ROC curve is a standard for assessing the performance of classifiers, obtained by calculating true positive rates and false positive rates. A model with a ROC curve closer to the upper left corner indicates better performance, and vice versa worse. AUC represents the area under the ROC curve, with a higher value indicating better model performance. Additionally, the PR curve is another standard for evaluating classifier performance, created by using Recall and Precision rates. A curve closer to the upper right corner indicates better performance, and vice versa worse.

B. COMPARISONS

1) SUBJECTIVE COMPARISON OF DIFFERENT DE METHODS ON THE 2D IMAGE OR ALBEDO DOMAINS

For a fair comparison, we adjusted the way of capturing images by our UPS system and the one of traditional 2D imaging used by other methods to achieve a common exposure level. Fig. 3 shows the visual comparison of defect extraction among different competitors. From the output results, we observe that DDSC and MCITF, which process samples obtained through traditional 2D imaging, suffer from the problem of background and defect confusion in both the grayscale and binary levels. EDRNet and our SGVM using Fig. 3 (a) are also in trouble with unsatisfactory defect extraction results, as traditional 2D imaging results ignore material inherent properties (like albedo). By contrast, our proposed method using albedo map, which operates on the

novel UPS-based albedo domain images, not only accurately locates the defect targets but also highlights them effectively, as shown in Fig. 3 (h).

2) SUBJECTIVE COMPARISON OF DIFFERENT DE METHODS ON THE SAME ALBEDO MAP

To fairly compare the performance of different methods, we set the albedo map estimated by UPS as the common input for all methods. Fig. 4 shows the visual comparison of different competitors on the SealN defect data. From the first row, it can be observed that SRDD, DDSC, and our SGVM outperform MCITF in defect localization. Judging from the results in the second row, although the competitors' extraction results include defects, they also confuse redundant background information on it. Essentially, these methods cannot discriminatively distinguish the inter-class differences between defects and backgrounds in the SealN data with complex backgrounds, and then fall into false positives. In contrast, the proposed SGVM can effectively extract clear defect targets with consistency and discriminability from similar or high-intensity background regions, attributed to the proposed model's strong edge discovery capability with both discriminative scale selection and semantic guidance.

3) SUBJECTIVE COMPARISON OF DIFFERENT DE METHODS ON OTHER PUBLIC DATASETS

In addition, our SGVM can be applied to other types of defect datasets such as BearD, SteelD, and OBriC, demonstrating equally promising performance. Here, we provide challenging samples from these three datasets, as shown in
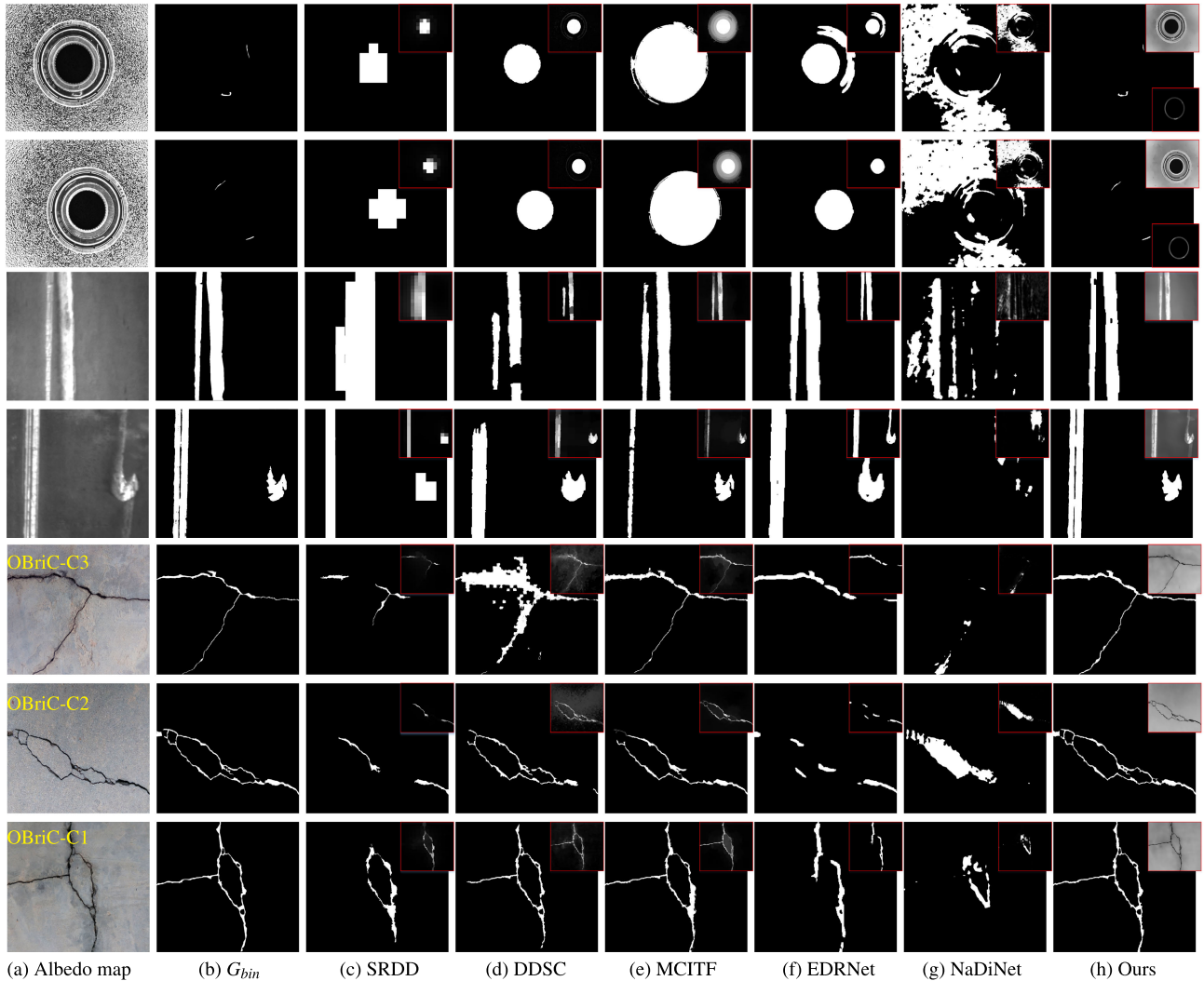
FIGURE 5. Visual comparison of defects (after binarization) extraction on samples in three datasets including BearD, SteelD, and OBriC, respectively. Note that: output grayscale image in the upper-right corner and output within ROI in the lower-right corner.
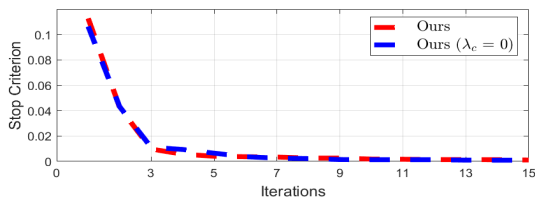
(a) Albedo map (b) G_{bin} (c) SRDD (d) DDSC (e) MCITF (f) EDRNet (g) NaDiNet (h) Ours

FIGURE 6. Convergence behavior.

Fig. 5. From these cases, an obvious conclusion can be drawn that the proposed SGVM successfully achieves separation between large-scale foreground and small-scale background, leading to visually pleasing defect extraction results.

C. PARAMETER STUDY AND QUANTITATIVE EVALUATION

1) HOW DO PARAMETERS λ_S AND λ_C INFLUENCE THE PROPOSED SGVM

The regularization parameters λ_s and λ_c are crucial for the selective removal/preservation of structures in our

proposed model. Table 1 shows the performance of our model with varying regularization parameters λ_s and λ_c, using the three widely used metrics for defect detection evaluation: MAE, SM, and AUC. It can be observed that our SGVM with $\lambda_s = 0.05$ achieves the lowest MAE values and both the highest SM and AUC values on the SealN test dataset, indicating its defect extraction performance has close intensity, structural consistency compared to the ground truth. The difference is that the parameter $\lambda_s = 0.005$ shows the best average MAE, SM, and AUC values on the BearD, SteelD, and OBriC datasets, demonstrating close defect extraction results to the ground truth. Besides, fixing the optimal λ_s value for different datasets, we observe that the SM and AUC values increase with larger λ_c values, while the MAE values are vice versa. It can be considered that the semantic-guided regularization term introduced in the optimization function allows for better preservation of desired large-scale structures, resulting in improved defect extraction performance. To this end,

TABLE 1. The performance of the proposed models with varying λ_s and λ_c values is evaluated on the SealN, BearD, SteelD, and OBriC datasets using MAE, SM, and AUC metrics.

	SealN									BearD								
			λ_s				λ_c					λ_s				λ_c		
	0.0005	0.005	0.05	0.5	1	0	0.5	0.7	1	0.0005	0.005	0.05	0.5	1	0	0.5	0.7	1
MAE ↓	0.0547	0.0372	**0.02698**	0.0337	0.0462	0.03352	0.030242	0.02859	**0.02698**	0.0010	**0.0009**	0.0194	0.0206	0.0174	0.0193	0.0143	0.0106	**0.0009**
SM ↑	0.6439	0.7559	**0.8482**	0.6493	0.6182	0.6519	0.7785	0.8141	**0.8482**	0.6912	**0.8175**	0.5225	0.5036	0.4698	0.5318	0.6425	0.6833	**0.8175**
AUC(%) ↑	91.26	93.02	**94.834**	87.28	87.46	88.36	91.27	92.88	**94.83**	86.20	**89.00**	75.68	71.62	67.94	75.93	81.11	85.54	**89.00**

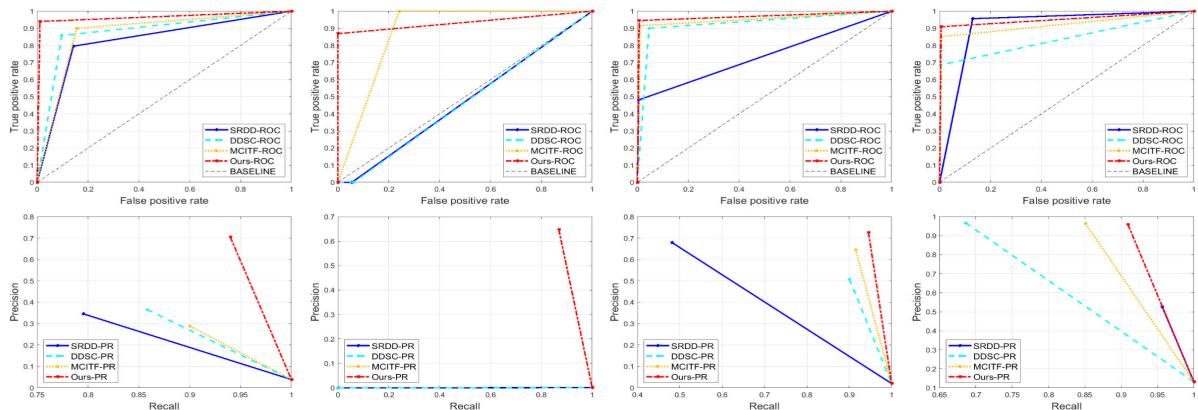|	SteelD									OBriC								
			λ_s				λ_c					λ_s				λ_c		
	0.0005	0.005	0.05	0.5	1	0	0.5	0.7	1	0.0005	0.005	0.05	0.5	1	0	0.5	0.7	1
MAE ↓	0.0188	**0.0181**	0.0534	0.4089	0.4704	0.0510	0.0320	0.0243	**0.0181**	0.4162	**0.0151**	0.5082	0.5193	0.5394	0.0522	0.0356	0.0193	**0.0151**
SM ↑	0.9271	**0.9298**	0.7618	0.5157	0.4195	0.7986	0.8416	0.9153	**0.9298**	0.3258	**0.7611**	0.2871	0.2810	0.2729	0.3262	0.6519	0.7037	**0.7611**
AUC(%) ↑	94.34	**94.82**	80.17	76.39	72.92	83.64	89.84	93.01	**94.82**	78.66	**94.98**	72.23	69.50	63.76	70.70	83.84	89.10	**94.98**

FIGURE 7. Comparison of defect extraction performance between our method and other competitors on four datasets including SealN, BearD, OBriC, and SteelD, respectively. Top Row: ROC curves. Bottom Row: PR curves.

we pre-set different λ_s and λ_c values for our SGVM on different datasets.

2) CONVERGENCE PERFORMANCE

By using Eqns. (8) and (14), the non-convex objective function (6) can be solved iteratively by decomposing it into subproblems. From Fig. 6, it can be observed that the proposed SGVM exhibits fast convergence and stable behavior on the SealN dataset with complex backgrounds. To better understand the impact of different prior constraints on convergence, we consider two cases: Ours and Ours w/o λ_c. Figure 6 shows that the latter has a smaller initial stop criterion value (defined as $||P^{(k+1)} - P^{(k)}||_F / ||P^{(0)}||_F$) compared to the former at the initial iteration. By combining the advantages of multiple priors, the proposed model demonstrates faster convergence compared to the model w/o λ_c, as shown in iterations 5 to 9. In other words, the proposed SGVM, with the assistance of multiple prior constraints, can converge more effectively to a global optimal solution for problem 9, and the same holds true for the original problem 6.

3) DEFECT EXTRACTION PERFORMANCE BETWEEN DIFFERENT METHODS

Fig. 7 shows the performances of ROC and PC curves for the proposed SGVM compared to other advanced methods with defect extraction capabilities. On the one hand, the proposed method exhibits the highest true positive rate on all tested datasets when the false positive rate is extremely low,

TABLE 2. Mean AUC values (%) between the proposed method and other competitors. Note that OBriC Scenes are collected from Fig. 5.

	SRDD	DDSC	MCITF	Ours
SealN	82.66	88.07	87.22	96.46
BearD	47.24	46.91	87.97	93.42
SteelD	91.31	84.12	92.3	95.16
OBriC Scenes				
OBriC-C1	68.56	93.66	95.92	96.63
OBriC-C2	79.06	91.82	94.60	97.20
OBriC-C3	67.87	85.90	87.33	95.52

indicating superior defect detection performance. On the other hand, when the recall value is large, the proposed method also achieves the highest precision value compared to other competing methods on all tested datasets, that is, a higher precision rate in defect detection is obtained. Therefore, in terms of ROC and PR curves, it is obvious that the proposed SGVM consistently outperforms other competitive methods with larger margins in terms of classification performance, demonstrating its superior defect extraction capabilities.

Table 2 further lists the AUC performances of the proposed SGVM and three other competitive defect extractors on multiple datasets. It should be noted that the AUC scores in Table 2 represent the average performance across multiple sets of data in different datasets. Also, for each image, the best score is highlighted in bold. From Table 2, one can conclude that the proposed SGVM achieves the highest scores on

TABLE 3. Running performance (in second). All methods are run within MATLAB.

Images (size)	Methods			
	SRDD	DDSC	MCITF	Ours
SealN (2448 × 2048)	7.92	71.36	1185.55	52.22
BearD (1228 × 1029)	5.47	73.16	1127.61	23.14
OBriC (1024 × 1024)	2.49	20.47	372.03	9.66
SteelD (200 × 200)	0.87	0.47	7.94	2.93

most of the defect sets, which is highly correlated with its inclusion of stronger small/large structure discriminators and the ability to preserve structures within the ROI-labeled regions.

Table 3 reports the average runtime of different competitive methods for defect extraction on samples with different sizes from various datasets, which is used to demonstrate their corresponding computational efficiency. All defect extraction methods were implemented on MATLAB and tested on a mobile computer with four Intel Core i7-11800H processors at 2.3GHz, following the original setting of all released codes. From Table 3, we observe that the proposed SGVM is only slower than the SRD method. This is because the SRD method utilizes highly resource-intensive superpixel segmentation implemented through external executable C code. However, the saliency defect effects produced by this method are not ideal, as shown in Fig. 4, which is attributed to the fact that it does not consider the scale characteristics of defects (foreground) and small structures (background). Although our SGVM is not the fastest one, it delivers satisfactory defect extraction results, as demonstrated in Figs. 4 and 5, and its speed can be easily improved further through C++ programming or GPU-accelerated implementations using image pyramid schemes in the future.

V. CONCLUSION
In this paper, we propose a semantic-guided variational model for extracting large-scale structures (or say defects) from defect scenarios with complex textured backgrounds, such as SealN. The core idea is to combine scale-adaptive band selection and semantic-guided prior regularization on the albedo domain estimated by UPS, achieving effective preservation of defect targets with unblurred boundaries and suppression of complex backgrounds. Experimental comparisons on four challenging defect datasets demonstrate that the proposed SGVM outperforms other competitive methods in both subjective and objective evaluations for defect extraction. It is worth noting that our SGVM technique can be applied to various visual-based scenarios, such as 3D defect detection, image segmentation, and feature extraction, and therefore holds potential for further development.

REFERENCES
[1] Z.-D. Tsai and M.-H. Perng, "Defect detection in periodic patterns using a multi-band-pass filter," *Mach. Vis. Appl.*, vol. 24, no. 3, pp. 551–565, Apr. 2013.

[2] S. H. Hanzaei, A. Afshar, and F. Barazandeh, "Automatic detection and classification of the ceramic tiles' surface defects," *Pattern Recognit.*, vol. 66, pp. 174–189, Jun. 2017.

[3] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 626–644, Mar. 2020.

[4] B. V. Lad, M. F. Hashmi, and A. G. Keskar, "Boundary preserved salient object detection using guided filter based hybridization approach of transformation and spatial domain analysis," *IEEE Access*, vol. 10, pp. 67230–67246, 2022.

[5] X. Bai, Y. Fang, W. Lin, L. Wang, and B.-F. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2135–2145, Nov. 2014.

[6] X. Zhou, Z. Liu, C. Gong, and W. Liu, "Improving video saliency detection via localized estimation and spatiotemporal refinement," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 2993–3007, Nov. 2018.

[7] Q. Jin and L. Chen, "A survey of surface defect detection of industrial products based on a small number of labeled data," 2022, *arXiv:2203.05733*.

[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.

[9] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 818–832, Apr. 2017.

[10] Q. Huangpeng, H. Zhang, X. Zeng, and W. Huang, "Automatic visual defect detection using texture prior and low-rank representation," *IEEE Access*, vol. 6, pp. 37965–37976, 2018.

[11] G. Song, K. Song, and Y. Yan, "Saliency detection for strip steel surface defects using multiple constraints and improved texture features," *Opt. Lasers Eng.*, vol. 128, May 2020, Art. no. 106000.

[12] Y. Quéau, F. Lauze, and J.-D. Durou, "Solving uncalibrated photometric stereo using total variation," *J. Math. Imag. Vis.*, vol. 52, no. 1, pp. 87–107, May 2015.

[13] B. Haefner, Z. Ye, M. Gao, T. Wu, Y. Queau, and D. Cremers, "Variational uncalibrated photometric stereo under general lighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8538–8547.

[14] G. Sikander and S. Anwar, "A novel machine vision-based 3D facial action unit identification for fatigue detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2730–2740, May 2021.

[15] T. Yang and F. Gu, "Overview of modulation techniques for spatially structured-light 3D imaging," *Opt. Laser Technol.*, vol. 169, Feb. 2024, Art. no. 110037.

[16] Y. Quéau, F. Lauze, and J.-D. Durou, "Solving the uncalibrated photometric stereo problem using total variation," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis. (SSVM)*. Leibnitz, Austria: Springer, Jun. 2013, pp. 270–281.

[17] G. Chen, K. Han, B. Shi, Y. Matsushita, and K. K. K. Wong, "Self-calibrating deep photometric stereo networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8731–8739.

[18] G. Chen, K. Han, B. Shi, Y. Matsushita, and K. K. Wong, "Deep photometric stereo for non-Lambertian surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 129–142, Jan. 2022.

[19] Y. Ju, B. Shi, M. Jian, L. Qi, J. Dong, and K.-M. Lam, "NormAttention-PSN: A high-frequency region enhanced photometric stereo network with normalized attention," *Int. J. Comput. Vis.*, vol. 130, no. 12, pp. 3014–3034, Dec. 2022.

[20] Y. Ju, M. Jian, C. Wang, C. Zhang, J. Dong, and K.-M. Lam, "Estimating high-resolution surface normals via low-resolution photometric stereo images," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Aug. 4, 2023, doi: 10.1109/TCSVT.2023.3301930.

[21] B. Cai, X. Xing, and X. Xu, "Edge/structure preserving smoothing via relativity-of-Gaussian," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 250–254.

[22] P. Dollár and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1558–1570, Aug. 2015.

[23] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," in *Proc. SIGGRAPH Asia Conf.*, Dec. 2011, pp. 1–12.

[24] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," *J. Graph. Tools*, vol. 12, no. 2, pp. 13–21, Jan. 2007.

[25] L. Zhou, Z. Yang, Z. Zhou, and D. Hu, "Salient region detection using diffusion process on a two-layer sparse graph," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5882–5894, Dec. 2017.

[26] X. Wu, X. Ma, J. Zhang, A. Wang, and Z. Jin, "Salient object detection via deformed smoothness constraint," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2815–2819.

[27] G. Song, K. Song, and Y. Yan, "EDRNet: Encoder–decoder residual network for salient object detection of strip steel surface defects," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9709–9719, Dec. 2020.

[28] G. Li, C. Han, and Z. Liu, "No-service rail surface defect segmentation via normalized attention and dual-scale interaction," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023.

[29] L. Li. (2017). *The Original Bridge Crack Dataset*. [Online]. Available: https://github.com/Iskysir/Bridge_Crack_Image_Data

[30] K. Song. (2020). *NEU Surface Defect Database*. [Online]. Available: http://faculty.neu.edu.cn/songkc/en/zdylm/263265/list/index.htm

[31] Y. Peng. (2023). *Bearing Defect Dataset*. [Online]. Available: https://ieee-dataport.org/documents/bearing-defect-dataset

YUPING YE received the Ph.D. degree in pattern recognition and intelligent system from the University of Chinese Academy of Sciences (CAS), Beijing, China, in 2016. He is currently a Staff Member with the Shenzhen Institute of Advanced Technology (SIAT), CAS, Shenzhen, China. His research interests include computer vision and computer graphics.

FEIFEI GU received the Ph.D. degree in mechanical and automation engineering from Xi'an Jiaotong University, in 2017. She is currently an Associate Research Fellow with the Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. Her research interests include stereo vision-based 3D sensing and image processing.

FANG LIU received the M.S. degree in mechanical engineering and automation from the China University of Mining and Technology, in 2004. She is currently an Associate Professor with the Logistics Academy, Xuzhou, China. Her research interests include automation testing and fault diagnosis.

SHIYANG LONG received the B.S. degree in agriculture and engineering from Zhongkai University, in 2020. He is currently pursuing the M.E. degree in electronic information with the Guangdong University of Technology, and interning state with the Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. His research interests include computer vision and embedded machine vision.

WEI CAO (Member, IEEE) received the Ph.D. degree in control science and engineering from the Wuhan University of Science and Technology, in 2022. He is currently a Postdoctoral Researcher with the Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. His research interests include computer vision and 2D/3D signal processing.

ZHAN SONG (Member, IEEE) received the Ph.D. degree in mechanical and automation engineering from the Chinese University of Hong Kong, Hong Kong, in 2008. He is currently a Professor with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China. His research interests include structured light-based sensing and vision-based human–computer interaction.

• • •