# Dual-Branch Network via Multiple Illumination-Aware Representation Learning for Steel Surface Defect Classification

Yong Seok Oh, Min Geol Kim, Bokyeong Kim, Jun Young Kim, Hyeongseob Jo, Jae Hyeon Park, and Sung In Cho
oys06047@dgu.ac.kr, rlaalsrjf6@sogang.ac.kr, qhruddl51@dgu.ac.kr, junkim1310@dgu.ac.kr, johs@dgu.ac.kr, pjh0011@dongguk.edu, csi2267@sogang.ac.kr

Smart Vision & Media Lab.
BMVC 2025

# I. Introduction

## A. Problem Settings

- Conventional deep learning methods trained under a single illumination direction struggle with highly reflective surfaces, where strong reflections either obscure defects or distort their shapes and boundaries, leading to misclassification.
- Multi-light imaging provides richer visual cues under diverse lighting directions, but existing methods that use independent encoders for each view fail to capture cross-view defect variations. On highly reflective surfaces, they often misinterpret strong reflection regions as defects, leading to misclassification and reduced robustness.

## B. Contributions

- We propose DiAL, a novel Dual-branch architecture for SDC that disentangles illumination-invariant defect features and lighting-dependent features via two dedicated encoders.
- By fusing view-consistent and view-specific features, DiAL enhances reliability and adaptability in defect classification.
- Experiments on a highly reflective surface defect dataset [1] show that DiAL outperforms SOTA methods in classification accuracy, interpretability, and reliability.

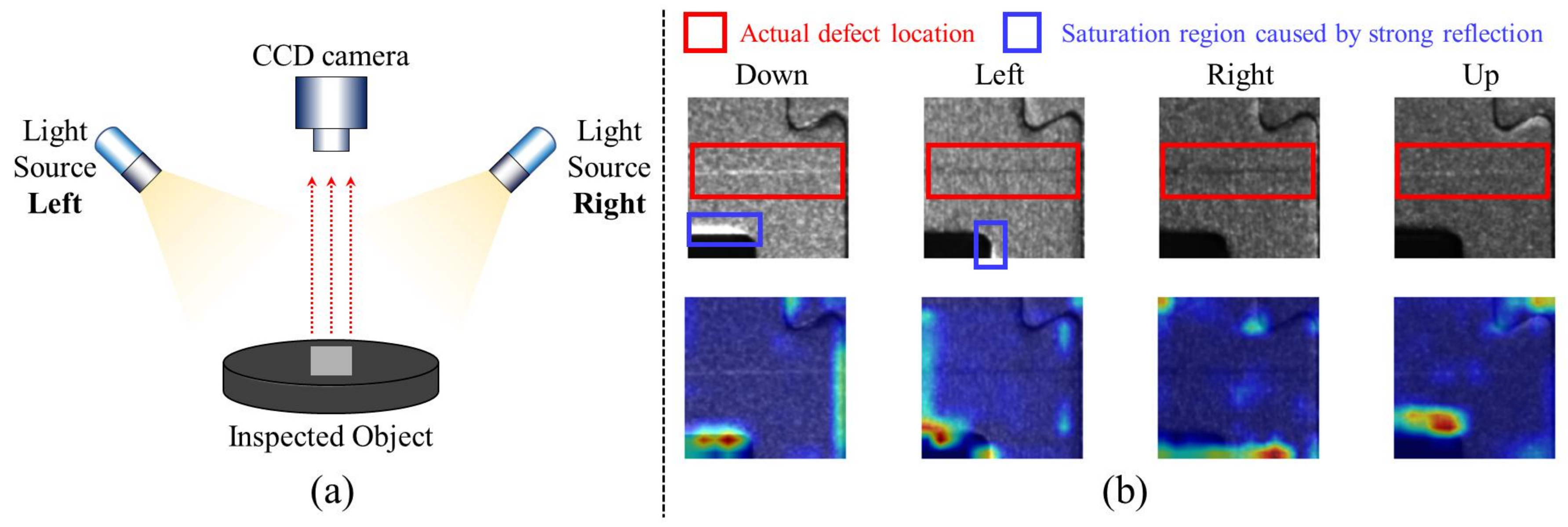


Figure 1: (a) Illustration of the multi-light source illumination and image acquisition system (Down/Up lights are omitted). (b) Grad-CAM visualizations of a conventional multi-light source illumination-based defect classification method under different lighting conditions.

# II. Proposed Method

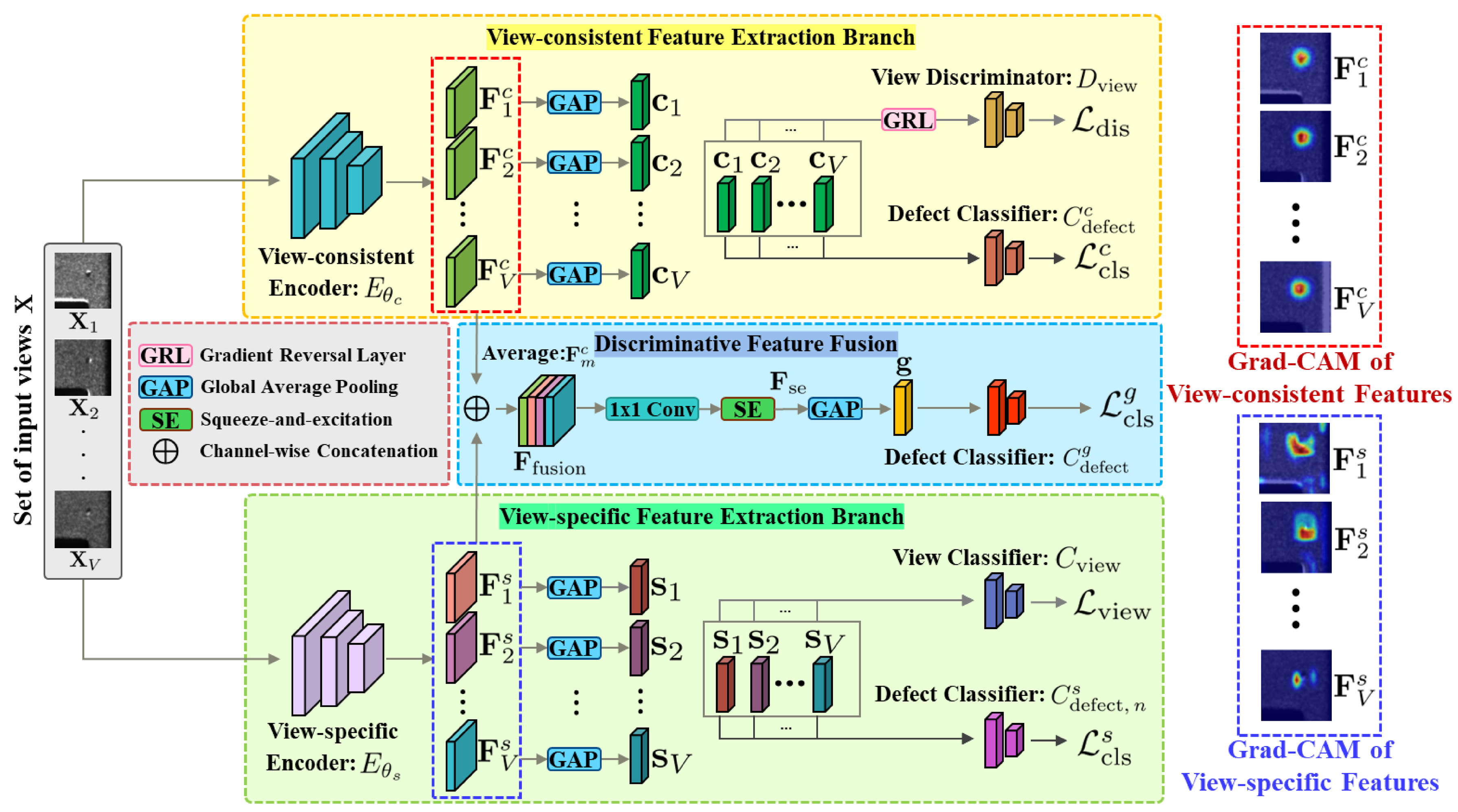


Figure 2: The overall architecture of the proposed DiAL

## A. View-consistent Feature Extraction (VCE) Branch

- To be robust against misclassification caused by saturation regions and to reliably extract consistent defect representations across varying lighting conditions, we design the View-consistent Feature Extraction (VCE) branch.

$$\mathcal{L}^c_{cls} = \sum_{n=1}^{V} \mathrm{CE}\left(C^c_{\text{defect}}(\mathbf{c}_n), y\right), \quad \mathcal{L}_{dis} = \sum_{n=1}^{V} \mathrm{CE}\left(D_{\text{view}}(\mathrm{GRL}(\mathbf{c}_n)), y^{\text{view}}_n\right). \quad (1)$$

## B. View-specific Feature Extraction (VSE) Branch

- While the VCE branch effectively suppresses misclassification caused by saturation, it cannot capture defect cues that emerge only under specific lighting conditions. To address this, we design the View-specific Feature Extraction (VSE) branch, which emphasizes illumination-dependent defect features.

$$\mathcal{L}_{\text{view}} = \sum_{n=1}^{V} \mathrm{CE}(C_{\text{view}}(\mathbf{s}_n), y^{\text{view}}_n), \quad \mathcal{L}^s_{cls} = \sum_{n=1}^{V} \mathrm{CE}(C^s_{\text{defect},n}(\mathbf{s}_n), y). \quad (2)$$

## C. Discriminative Feature Fusion

- By combining the features from VSE and VCE for the final classification, the model reduces misclassification caused by reflections and effectively leverages illumination-specific defect cues.

$$\mathbf{F}_{se} = \mathrm{SE}(\mathrm{Conv}_{1\times 1}(\mathbf{F}_{fusion})). \quad (3) \qquad \mathcal{L}^g_{cls} = \mathrm{CE}(C^g_{\text{defect}}(\mathrm{GAP}(\mathbf{F}_{se})), y). \quad (4)$$

# III. Experiments

## A. Quantitative Results

| Method | Normal | Bright Line | Deformation | Dent | Scratch | Overall accuracy |
|---|---|---|---|---|---|---|
| ETE [2] | 90.18 | 55.40 | 73.13 | 80.00 | 86.40 | 81.38 |
| SN-ELF+MRF [3] | 92.77 | 64.79 | 77.11 | 90.37 | **92.15** | 86.71 |
| FOMI [1] | 94.03 | 63.03 | **86.03** | 95.07 | 91.03 | 88.40 |
| SICN [4] | **95.00** | **66.19** | 85.82 | 94.44 | 91.18 | **88.83** |
| MHAF [5] | 91.48 | 60.56 | 71.14 | **96.30** | 89.47 | 84.61 |
| Proposed | **96.00** | **82.33** | **86.67** | **96.33** | **91.33** | **91.44** |

Table 1: Quantitative comparison of class-wise and overall classification accuracy across surface defect types under multi-light source illumination. The results compare our proposed method against existing approaches. The highest classification performance is indicated in **bold red**, while the second-highest performance is shown in **bold**.

| Lighting condition | Down | Left | Right | Upper | Overall accuracy |
|---|---|---|---|---|---|
| Accuracy | 99.00 | 96.85 | 95.35 | 97.74 | 97.23 |

Table 2: Classification performance of lighting condition by the VSE module

| Ablation | Normal | Bright Line | Deformation | Dent | Scratch | Overall accuracy |
|---|---|---|---|---|---|---|
| w/o ($\mathcal{L}_{\text{view}}$, $\mathcal{L}_{\text{dis}}$) | 93.89 | 70.42 | 82.59 | 97.04 | 92.14 | 87.15 |
| w/o $\mathcal{L}_{\text{view}}$ | 95.56 | 67.61 | 80.60 | 88.89 | 93.68 | 88.83 |
| w/o $\mathcal{L}_{\text{dis}}$ | 91.11 | 78.87 | 86.57 | 95.56 | 94.25 | 90.32 |
| **DiAL (Ours)** | 96.00 | 82.33 | 86.67 | 96.33 | 91.33 | 91.44 |

Table 3: Ablation study to assess the contributions of $\mathcal{L}_{\text{view}}$ and $\mathcal{L}_{\text{dis}}$

## B. Qualitative Analysis

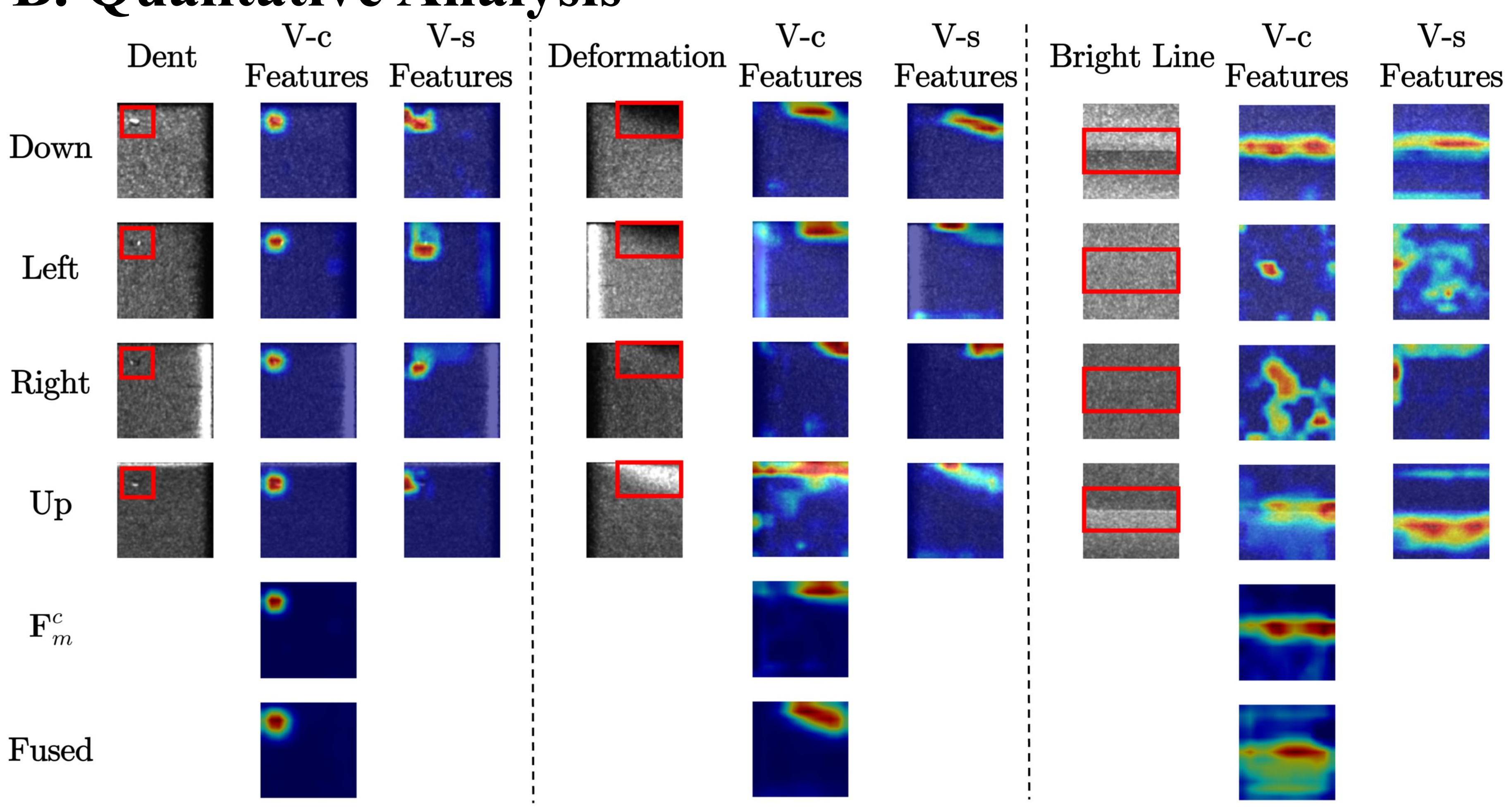


Figure 3: Grad-CAM visualizations for three defect classes (Dent, Deformation, and Bright Line) under four lighting conditions (Down, Left, Right, and Up). Each column shows (1) the input image, (2) Grad-CAM results based on view-consistent (V-c) encoder features, and (3) those based on view-specific (V-s) encoder features. The last row presents the fused attention map obtained by combining V-s features with the averaged V-c attention map ($\mathbf{F}^c_m$).

# IV. Conclusion

- DiAL uses a dual-branch design: a view-consistent encoder (robust to reflection-induced artifacts) and a view-specific encoder (captures lighting-dependent cues).
- This design enabled the model to reduce misclassification caused by lighting-induced misleading reflections while leveraging meaningful lighting-responses around defects.
- Grad-CAM-based visualizations qualitatively validated the interpretability and reliability of the VCE and VSE

# V. References

[1] Guizhong Fu, Shukai Jia, Wenbin Zhu, Jiangxin Yang, Yanlong Cao, Michael Ying Yang, and Yanpeng Cao. Fusion of multi-light source illuminated images for effective defect inspection on highly reflective surfaces. *Mechanical Systems and Signal Processing,* 175:109109, 2022.
[2] Li Yi, Guangyao Li, and Mingming Jiang. An end-to-end steel strip surface defects recognition system based on convolutional neural networks. *steel research international*, 88(2):1600068, 2017.
[3] Guizhong Fu, Peize Sun, Wenbin Zhu, Jiangxin Yang, Yanlong Cao, Michael Ying Yang, and Yanpeng Cao. A deep-learning-based approach for fast and robust steel surface defects classification. *Optics and Lasers in Engineering*, 121:397–405, 2019.
[4] Francisco López de la Rosa, José L Gómez-Sirvent, Rafael Morales, Roberto Sánchez Reolid, and Antonio Fernández-Caballero. Defect detection and classification on semi conductor wafers using two-stage geometric transformation-based data augmentation and squeezenet lightweight convolutional neural network. *Computers & Industrial Engineering*, 183:109549, 2023.
[5] Michele Somero, Federico Urli, Lauro Snidaro, and Alessandro Liani. Defect detection multiheadattention fusion model on images acquired with different light sources. *In 2024 27th International Conference on Information Fusion (FUSION)*, pages 1–6. IEEE, 2024.