

Robust Deep Learning Framework untuk Deteksi Cacat Permukaan pada Material Logam dalam Kondisi Pencitraan yang Terganggu Noise dan Blur

Santi Rahayu ^{1*}, Achmad Hindasyah ², Nur Annisahaq ³

^{1*,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Provinsi Banten, Indonesia.

³ Magister Teknik Informatika, Program Pascasarjana, Universitas Pamulang, Kota Tangerang Selatan, Provinsi Banten, Indonesia.

Corresponding Email: dosen02666@unpam.ac.id ^{1*}

Histori Artikel:

Dikirim 10 Juni 2026; *Diterima dalam bentuk revisi* 15 Juni 2026; *Diterima* 30 Juli 2026; *Diterbitkan* 10 September 2026. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STM IK Indonesia Banda Aceh.

Abstrak

Deteksi cacat permukaan memegang peranan penting dalam pengendalian kualitas industri, terutama pada manufaktur logam, di mana cacat dapat mempengaruhi kekuatan struktural dan keandalan produk. Meskipun sistem inspeksi berbasis deep learning menunjukkan hasil yang menjanjikan, sebagian besar pendekatan mengasumsikan kondisi pencitraan yang ideal dan mengalami penurunan performa yang signifikan saat terjadi gangguan seperti noise dan blur. Penelitian ini mengusulkan framework deep learning yang tahan terhadap degradasi citra untuk deteksi cacat permukaan. Kerangka kerja ini mengintegrasikan simulasi degradasi citra secara terkendali dan model segmentasi berbasis U-Net untuk pelokalan cacat secara akurat pada tingkat piksel. Evaluasi ketahanan dilakukan dengan menambahkan Gaussian noise, salt-and-pepper noise, dan motion blur. Hasil eksperimen pada Severstal Steel Defect Dataset menunjukkan bahwa metode yang diusulkan mengungguli baseline segmentasi YOLOv8n-seg. Pada kondisi citra bersih, model mencapai precision sebesar 0,6328 dan F1-score sebesar 0,3199. Dalam kondisi terdegradasi, performa model tetap stabil, meskipun terjadi penurunan, sementara metode lain menunjukkan penurunan signifikan. Selain itu, keluaran segmentasi memungkinkan estimasi area cacat secara kuantitatif, memberikan informasi penting untuk aplikasi industri. Pendekatan berbasis segmentasi terbukti lebih robust dan andal dalam kondisi pencitraan yang tidak ideal.

Kata Kunci: Deteksi Cacat Permukaan; Deep Learning; Segmentasi Citra; Degradasi Citra; Noise dan Blur; Inspeksi Industri.

Abstract

Surface defect detection plays a crucial role in industrial quality control, especially in metal manufacturing, where defects can impact structural integrity and product reliability. Although deep learning-based inspection systems have shown promising results, most approaches assume ideal imaging conditions and experience significant performance degradation when faced with disturbances such as noise and blur. This study proposes a robust deep learning framework for surface defect detection under degraded imaging conditions. The framework integrates controlled image degradation simulation and a segmentation-based model using U-Net to enable precise pixel-level localization of defects. Robustness is evaluated through the addition of Gaussian noise, salt-and-pepper noise, and motion blur. Experimental results on the Severstal Steel Defect Dataset demonstrate that the proposed method outperforms the baseline segmentation model YOLOv8n-seg. Under clean conditions, the model achieves a precision of 0.6328 and an F1-score of 0.3199. In degraded conditions, the model remains stable despite performance drops, whereas other methods show significant declines. Additionally, the segmentation output allows for quantitative estimation of defect areas, providing valuable information for industrial applications. The findings indicate that segmentation-based approaches offer greater robustness and more reliable defect localization in challenging imaging environments.

Keyword: Surface Defect Detection; Deep Learning; Image Segmentation; Image Degradation; Noise and Blur; Industrial Inspection.

1. Pendahuluan

Inspeksi kualitas permukaan memainkan peran penting dalam industri manufaktur modern, khususnya dalam produksi baja, rekayasa otomotif, dan komponen mekanis presisi. Cacat permukaan seperti retakan, goresan, inklusi, dan pitting dapat secara signifikan menurunkan integritas struktural, daya tahan, dan kinerja produk. Oleh karena itu, deteksi cacat permukaan yang akurat dan robust sangat penting untuk memastikan kualitas dan keandalan produk dalam lingkungan industri (Qiao *et al.*, 2025). Metode inspeksi permukaan tradisional sangat bergantung pada pemeriksaan visual manual yang dilakukan oleh operator manusia. Meskipun pendekatan tersebut dapat mengidentifikasi cacat yang terlihat jelas, metode ini memiliki beberapa keterbatasan, termasuk subjektivitas, kelelahan operator, dan skalabilitas yang rendah dalam lingkungan industri dengan throughput tinggi. Keterbatasan tersebut mendorong pengembangan sistem inspeksi otomatis berbasis computer vision dan machine learning (Luo *et al.*, 2020). Pendekatan otomatis pada tahap awal terutama menggunakan teknik pengolahan citra konvensional, termasuk deteksi tepi, thresholding, dan analisis berbasis tekstur. Deskriptor fitur seperti *Local Binary Pattern* (LBP), *Gray Level Co-occurrence Matrix* (GLCM), dan filter Gabor banyak digunakan untuk mengkarakterisasi ketidakteraturan permukaan. Namun, metode berbasis fitur yang dirancang secara manual tersebut sangat sensitif terhadap noise, variasi pencahayaan, dan tekstur permukaan yang kompleks sehingga membatasi robustness-nya dalam skenario industri nyata (Chen *et al.*, 2023). Kemajuan terbaru dalam *deep learning*, khususnya *Convolutional Neural Networks* (CNN), telah meningkatkan kinerja deteksi cacat secara signifikan dengan memungkinkan ekstraksi fitur secara otomatis dari citra mentah. Arsitektur berbasis deteksi seperti Faster R-CNN dan YOLO telah banyak diterapkan karena efisiensi dan kemampuannya untuk bekerja secara real-time. Namun, pendekatan tersebut umumnya menggunakan representasi bounding box yang kurang memadai untuk menangkap batas cacat secara rinci serta bentuk tidak beraturan yang umum ditemukan pada permukaan logam (Ren *et al.*, 2024; Zhou *et al.*, 2023). Selain itu, sebagian besar penelitian yang ada mengasumsikan bahwa citra masukan diperoleh dalam kondisi pencitraan yang ideal. Dalam praktiknya, proses akuisisi citra industri sering dipengaruhi oleh berbagai degradasi, termasuk noise sensor, motion blur, dan pencahayaan yang tidak stabil. Faktor-faktor tersebut dapat mengaburkan pola cacat dan secara signifikan menurunkan akurasi deteksi. Oleh karena itu, peningkatan robustness model deteksi cacat pada kondisi pencitraan yang terdegradasi tetap menjadi tantangan penting. Meskipun arsitektur kompleks seperti DeepLab, PSPNet, maupun Attention U-Net menawarkan mekanisme ekstraksi fitur tingkat tinggi, U-Net dipilih dalam penelitian ini sebagai baseline segmentasi piksel yang ringan dan efisien. Struktur encoder-decoder simetris dengan skip-connections pada U-Net sangat ideal untuk mentransfer fitur spasial resolusi tinggi yang dibutuhkan dalam mendeteksi cacat garis/retakan tipis pada permukaan logam tanpa beban komputasi yang berlebihan. Berbeda dengan penelitian sebelumnya, framework ini tidak hanya menyediakan pipeline segmentasi terisolasi, tetapi menyediakan modul terpadu yang *reusable* mencakup: (1) pemrosesan awal dan simulasi degradasi terstandarisasi, (2) model segmentasi U-Net piksel biner, dan (3) modul estimasi area cacat otomatis beserta kuantifikasi error. Kontribusi utama penelitian ini adalah sebagai berikut:

- 1) *Framework deep learning* yang robust untuk deteksi cacat permukaan pada kondisi pencitraan terdegradasi, termasuk noise dan blur.
- 2) Pendekatan berbasis segmentasi menggunakan U-Net untuk pelokalan area cacat secara akurat pada tingkat piksel.
- 3) Analisis komparatif yang komprehensif yang menunjukkan bahwa metode berbasis segmentasi mengungguli model berbasis deteksi seperti YOLOv8n-seg dalam hal robustness dan kinerja deteksi.
- 4) Kemampuan tambahan untuk mengestimasi area cacat yang diperoleh dari keluaran segmentasi guna mendukung analisis industri.

Studi awal mengenai deteksi cacat permukaan otomatis terutama mengandalkan pengolahan citra tradisional dan ekstraksi fitur yang dirancang secara manual. Pendekatan tersebut umumnya menggunakan metode seperti deteksi tepi, thresholding, operasi morfologi, dan analisis tekstur untuk mengidentifikasi pola tidak beraturan pada permukaan logam. Deskriptor tekstur seperti *Local Binary Pattern* (LBP), *Gray Level Co-occurrence Matrix* (GLCM), dan filter Gabor banyak digunakan untuk merepresentasikan karakteristik cacat. Tinjauan komprehensif yang dilakukan oleh (Luo *et al.*, 2020) menganalisis berbagai pendekatan deteksi cacat yang digunakan dalam inspeksi permukaan baja datar. Penelitian tersebut mengelompokkan metode yang ada menjadi pendekatan statistik, spektral, berbasis model, dan machine learning. Para penulis melaporkan bahwa metode tradisional sering mengalami kesulitan dalam menangani tekstur permukaan yang kompleks dan kondisi pencahayaan yang bervariasi di lingkungan industri. Studi survei selanjutnya juga mengonfirmasi bahwa pendekatan konvensional umumnya sensitif terhadap noise dan variasi pencahayaan sehingga membatasi robustness-nya dalam skenario manufaktur dunia nyata (Chen *et al.*, 2023; Qiao *et al.*, 2025). Keterbatasan tersebut mendorong peneliti untuk mengeksplorasi teknik machine learning dan deep learning guna meningkatkan sistem inspeksi otomatis. Dengan berkembang pesatnya *deep learning*, *Convolutional Neural Networks* (CNN) telah menjadi pendekatan dominan untuk deteksi cacat otomatis dalam sistem inspeksi industri. Model berbasis CNN dapat mempelajari fitur hierarkis secara otomatis dari citra mentah sehingga memungkinkan deteksi yang lebih akurat dibandingkan metode fitur yang dirancang secara manual. Beberapa penelitian telah menerapkan arsitektur CNN untuk tugas deteksi cacat permukaan. Sebagai contoh, (Zhao *et al.*, 2020) mengusulkan *framework* deteksi cacat multi-tahap yang mengintegrasikan region proposal network yang ditingkatkan dengan feature fusion dan teknik *super-resolution*. Pendekatan tersebut menunjukkan kinerja yang menjanjikan pada dataset cacat kereta berkecepatan tinggi dan dataset cacat permukaan NEU. Demikian pula, (Guan *et al.*, 2020) mengembangkan visual surface defect network berdasarkan arsitektur VGG19 yang telah ditingkatkan dan dikombinasikan dengan visualisasi fitur serta evaluasi structural similarity. Model tersebut mencapai akurasi klasifikasi yang lebih tinggi dibandingkan arsitektur CNN konvensional. Teknik transfer learning juga banyak digunakan untuk meningkatkan kinerja deteksi cacat ketika data pelatihan terbatas. (Wan *et al.*, 2021) menerapkan transfer learning dengan arsitektur VGG19 yang ditingkatkan dan augmentasi *region-of-interest*, serta mencapai akurasi pengenalan sebesar 97,8% pada dataset cacat baja campuran. Selain itu, pendekatan *semi-supervised learning* telah dieksplorasi untuk mengurangi ketergantungan terhadap dataset berlabel dalam jumlah besar. (Zheng *et al.*, 2020) mengusulkan framework deep learning semi-supervised berdasarkan MixMatch dan residual CNN networks yang menunjukkan kinerja kuat bahkan dengan jumlah sampel berlabel yang terbatas. Penelitian terbaru telah mengeksplorasi model segmentasi dan generatif tingkat lanjut untuk meningkatkan kinerja deteksi cacat. Sebagai contoh, IA-Mask R-CNN diusulkan untuk meningkatkan deteksi cacat berukuran kecil dan besar melalui desain anchor yang dioptimalkan (Zhu *et al.*, 2022). Selain itu, pendekatan berbasis GAN yang dikombinasikan dengan arsitektur CNN telah digunakan untuk menghasilkan data cacat sintetis dan meningkatkan akurasi klasifikasi (Mustafaev *et al.*, 2023). Selanjutnya, model state-of-the-art seperti MeDERT menunjukkan kinerja yang kuat pada berbagai dataset melalui strategi feature fusion tingkat lanjut (Wang and Xie, 2023). Penelitian terbaru semakin berfokus pada framework object detection yang dapat melakukan lokalisasi dan klasifikasi cacat secara bersamaan. Arsitektur deteksi populer seperti Faster R-CNN, Mask R-CNN, dan YOLO telah banyak diterapkan pada tugas inspeksi industri. Sebagai contoh, (Selamet *et al.*, 2022) menggabungkan Faster R-CNN dengan teknik Shape-from-Shading untuk menghasilkan label secara otomatis dan mendeteksi cacat pada permukaan logam. Pendekatan tersebut mencapai *mean average precision* (mAP) sebesar 0,83 pada dataset cacat permukaan NEU. Model deteksi berbasis YOLO juga mendapatkan perhatian yang signifikan karena kecepatan deteksi dan kemampuan real-time yang tinggi. (C. Wang *et al.*, 2022) mengusulkan model YOLOv4 yang ditingkatkan dengan *attentive feature fusion* dan *mosaic data augmentation* untuk meningkatkan kinerja deteksi. Metode tersebut menghasilkan mAP yang lebih tinggi dibandingkan beberapa model deteksi baseline. Penelitian yang lebih baru mengusulkan berbagai peningkatan terhadap arsitektur YOLO.

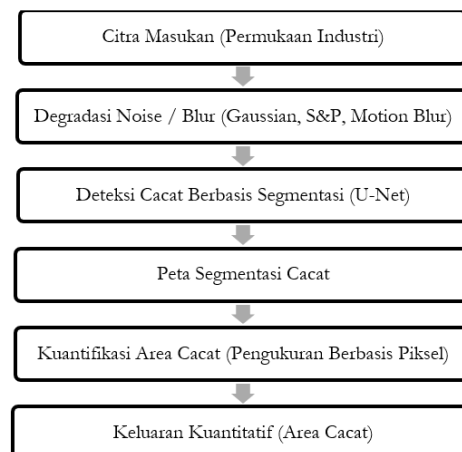
(Zhou *et al.*, 2023) memperkenalkan model berbasis YOLOv5 yang mengintegrasikan global attention mechanism untuk meningkatkan akurasi deteksi. Demikian pula, (Ren *et al.*, 2024) mengusulkan arsitektur YOLOv5 yang ditingkatkan dengan menggunakan deformable convolution dan efficient attention modules. Jaringan deteksi ringan juga telah dikembangkan untuk mendukung sistem inspeksi industri secara real-time. Sebagai contoh, (H. Wang *et al.*, 2025) mengusulkan model deteksi berbasis YOLOv8 dengan feature fusion yang ditingkatkan dan kompleksitas komputasi yang dikurangi, sehingga menghasilkan peningkatan kinerja deteksi sekaligus mempertahankan kecepatan inferensi yang efisien. Model berbasis YOLO terbaru telah memperkenalkan berbagai peningkatan arsitektur untuk meningkatkan akurasi dan efisiensi deteksi. Sebagai contoh, IE-YOLOv5 menggunakan mekanisme *attention* dan *loss function* yang ditingkatkan untuk memperoleh kinerja deteksi yang lebih tinggi (Zhao *et al.*, 2023). Demikian pula, LCG-YOLO mengintegrasikan modul ringan dan mekanisme attention untuk meningkatkan akurasi serta kecepatan inferensi (Yu *et al.*, 2024). Pendekatan lainnya, seperti YOLOv5 dengan BiFPN, meningkatkan multi-scale feature fusion untuk menghasilkan deteksi cacat yang lebih baik (Gao *et al.*, 2024). Perkembangan yang lebih baru, termasuk arsitektur FD-YOLO, semakin meningkatkan kinerja deteksi sekaligus mempertahankan efisiensi komputasi (Dang & Wang, 2025). Meskipun telah terjadi kemajuan yang signifikan pada metode deteksi cacat berbasis deep learning, masih terdapat beberapa tantangan dalam sistem inspeksi industri praktis. Pertama, banyak penelitian yang ada mengasumsikan bahwa citra masukan diperoleh dalam kondisi pencitraan ideal. Namun, dalam lingkungan industri nyata, akuisisi citra sering dipengaruhi oleh berbagai gangguan, termasuk noise sensor, motion blur, pencahayaan yang tidak stabil, dan pantulan permukaan. Degradasi tersebut dapat mengaburkan pola cacat dan menurunkan akurasi deteksi (Zhou *et al.*, 2023). Kedua, sebagian besar penelitian sebelumnya terutama berfokus pada tugas klasifikasi atau deteksi cacat, sedangkan pengukuran kuantitatif ukuran atau tingkat keparahan cacat jarang dibahas. Dalam proses pengendalian kualitas industri, kemampuan untuk mengukur area cacat sangat penting untuk menentukan apakah suatu komponen memenuhi standar manufaktur (Lin & Wibowo, 2021). Ketiga, banyak model deteksi terbaru memperoleh peningkatan kinerja dengan meningkatkan kompleksitas jaringan melalui penambahan modul attention, lapisan multi-scale feature fusion, atau komponen berbasis transformer. Meskipun peningkatan tersebut dapat meningkatkan akurasi deteksi, pendekatan tersebut sering meningkatkan biaya komputasi dan mengurangi kepraktisannya untuk sistem inspeksi *real-time*. Keterbatasan tersebut menunjukkan perlunya framework inspeksi yang robust, yang mampu mendeteksi cacat permukaan dan melakukan pengukuran kuantitatif sekaligus mempertahankan kinerja yang andal dalam kondisi pencitraan terdegradasi seperti noise dan blur. Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan framework machine learning yang robust yang mengintegrasikan simulasi degradasi citra, deteksi cacat, dan kuantifikasi cacat dalam arsitektur terpadu untuk mendeteksi dan mengukur cacat permukaan dalam kondisi pencitraan yang noise dan blur. Selain itu, penelitian terbaru telah mengeksplorasi cross-domain generalization untuk meningkatkan robustness model pada berbagai dataset industri. Namun, pendekatan tersebut masih menghadapi tantangan ketika menangani degradasi citra yang parah seperti noise dan blur (Chen *et al.*, 2024).

2. Metode Penelitian

2.1 Gambaran Umum Framework yang Diusulkan

Framework yang diusulkan bertujuan untuk melakukan deteksi dan analisis cacat permukaan secara robust pada kondisi pencitraan yang terdegradasi. Pipeline keseluruhan terdiri atas tiga tahap utama: degradasi citra, deteksi cacat, dan kuantifikasi cacat. Pertama, citra masukan diperoleh dari *Severstal Steel Defect Dataset*. Untuk mensimulasikan kondisi industri yang realistis, degradasi sintesis berupa *Gaussian noise*, *salt-and-pepper noise*, dan *motion blur* diterapkan pada citra masukan. Degradasi tersebut diperkenalkan selama evaluasi (*test-time stress test* saja, tanpa melibatkan citra terdegradasi pada proses pelatihan) untuk menilai robustness model. Selanjutnya, jaringan saraf konvolusional berbasis U-Net

digunakan untuk melakukan segmentasi area cacat pada tingkat piksel. Model menghasilkan peta segmentasi biner yang menunjukkan area cacat dan area tidak cacat. Terakhir, hasil segmentasi digunakan untuk menghitung area cacat dengan menghitung jumlah piksel yang diklasifikasikan sebagai cacat. Hal ini memungkinkan dilakukannya analisis kuantitatif terhadap tingkat keparahan cacat. *Framework* keseluruhan ditunjukkan pada Gambar 1, yang memperlihatkan alur dari citra masukan yang telah mengalami degradasi hingga segmentasi cacat dan pengukuran area

Gambar 1. *Pipeline Framework*

2.2 Model Degradasi Citra

Untuk mengevaluasi robustness, tiga jenis degradasi citra dipertimbangkan dengan parameter spesifik yang ditetapkan secara eksplisit:

1) Gaussian Noise:

Gaussian noise ditambahkan untuk mensimulasikan noise sensor:

$$I_{noisy} = I + \mathcal{N}(0, \sigma^2)$$

Di mana varians/standar deviasi ditetapkan sebesar $\sigma = 0,10$. I adalah citra asli dan $\mathcal{N}(0, \sigma^2)$ merepresentasikan Gaussian noise dengan nilai rata-rata nol dan varians σ^2 .

2) Salt-and-Pepper Noise:

Salt-and-pepper noise secara acak mengganti nilai piksel dengan intensitas minimum atau maksimum:

$$I_{sp}(x, y) = \begin{cases} 0 & \text{with probability } p \\ 255 & \text{with probability } p_v \\ I(x, y) & \text{otherwise} \end{cases}$$

Ditentukan dengan probabilitas $p = 0,05$ (amount 5%) dan rasio salt vs pepper sebesar 0,50 : 0,50.

3) Motion Blur:

Motion blur disimulasikan menggunakan kernel linear:

$$I_{blur} = I * K$$

Dimana K merupakan kernel horizontal berukuran 5×5 . $*$ menunjukkan operasi konvolusi dan K merupakan kernel motion blur.

2.3 Deteksi Cacat Berbasis U-Net

Modul deteksi cacat diimplementasikan menggunakan arsitektur U-Net sederhana yang banyak digunakan untuk tugas segmentasi citra karena kemampuannya menangkap konteks global dan detail spasial yang lebih halus. Jaringan terdiri atas struktur *encoder-decoder*. Encoder mengekstraksi fitur hierarkis melalui lapisan konvolusi dan downsampling, sedangkan decoder membangun kembali resolusi spasial melalui operasi upsampling dan konvolusi.

Proses forward jaringan dapat dinyatakan sebagai:

$$Y = \sigma(f(X))$$

Dimana X merupakan citra masukan, $f(\cdot)$ merepresentasikan model U-Net, dan σ merupakan fungsi aktivasi sigmoid yang menghasilkan peta probabilitas pada tingkat piksel. Peta probabilitas ini diubah menjadi mask biner menggunakan nilai threshold tetap sebesar 0,50 ($THRESHOLD = 0.5$).

2.4 Loss Function

Model dilatih menggunakan Binary Cross-Entropy (BCE) loss:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

Dimana y_i merupakan label ground truth dan $\hat{y}_i$ merupakan probabilitas prediksi.

2.5 Kuantifikasi Cacat

Area cacat dihitung berdasarkan mask segmentasi yang diprediksi. Area didefinisikan sebagai jumlah total piksel yang diklasifikasikan sebagai cacat:

$$A = \sum_{x,y} M(x,y)$$

Dimana $M(x,y) \in \{0,1\}$ merupakan mask segmentasi biner.

Untuk mengevaluasi akurasi kuantifikasi, digunakan Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE):

$$MAE = \frac{1}{N} \sum |A_{pred} - A_{gt}|$$

$$RMSE = \sqrt{\frac{1}{N} \sum (A_{pred} - A_{gt})^2}$$

Selain itu, untuk memberikan konteks skala industri, dihitung pula Relative MAE dan Relative RMSE terhadap rata-rata luas cacat Ground Truth (A_{gt}):

$$Relative\ MAE\ (\%) = \frac{MAE}{\bar{A}_{gt}} \times 100\%$$

$$Relative\ RMSE\ (\%) = \frac{RMSE}{\bar{A}_{gt}} \times 100\%$$

2.5 Konfigurasi Pelatihan dan Reproduksiabilitas

Model U-Net diimplementasikan menggunakan PyTorch dan dilatih menggunakan optimizer Adam dengan learning rate sebesar 0,001. Binary Cross-Entropy (BCE) loss digunakan sebagai

objective function. Model dilatih selama 20 epoch dengan batch size 8. Citra masukan diubah ukurannya menjadi 256×256 piksel. YOLOv8n-segYOLOv8n-seg Untuk menjamin reproduosibilitas eksperimen, digunakan random seed tetap yaitu 42 (SEED = 42) pada NumPy, PyTorch, dan CUDA. Eksperimen dijalankan sebanyak 1 kali pelatihan (N_RUNS = 1) tanpa augmentasi buatan pada data pelatihan U-Net.

2.6 Deskripsi Dataset

Eksperimen dilakukan menggunakan *Severstal Steel Defect Dataset* (Kaggle, 2019). Setelah tahap pra-pemrosesan awal dan pembersihan data RLE, diperoleh total 6.666 citra unik. Dataset dibagi secara acak dengan rasio 70% : 15% : 15% menjadi:

- 1) Data Pelatihan (*Training*): 4.666 citra (70,00%)
- 2) Data Validasi (*Validation*): 1.000 citra (15,00%)
- 3) Data Pengujian (*Testing*): 1.000 citra (15,00%)

Seluruh 1.000 citra pada set pengujian memuat sekurang-kurangnya satu cacat (100% images with defect). Distribusi ukuran cacat Ground Truth (GT) pada set pengujian memiliki nilai Rata-rata (Mean) sebesar 3.774,02 piksel (5,76% dari total piksel citra), Median sebesar 1.638,50 piksel, rentang Minimum 33,00 piksel, Maksimum 55.814,00piksel, Percentile-25 (P_{25}) sebesar 748,00 piksel, dan Percentile-75 (P_{75}) sebesar 4.446,25 piksel.

2.7 Pra-pemrosesan Data dan Penanganan Class Imbalance

Seluruh citra dikonversi menjadi grayscale dan dinormalisasi ke rentang [0, 1]. Mask ground truth yang sesuai juga dinormalisasi dan dikonversi ke format biner (1 untuk cacat, 0 untuk latar belakang). Pada tingkat piksel, proporsi piksel cacat pada test set hanya sebesar 5,76%, sedangkan piksel latar belakang mencapai 94,24%. Pada eksperimen baseline ini, ketidakseimbangan kelas (*class imbalance*) sengaja tidak ditangani melalui *weighted loss*, *Focal loss*, maupun *oversampling* untuk mengevaluasi kinerja murni arsitektur standar. Hal ini menjadi faktor utama rendahnya nilai Recall (0,2614).

2.8 Detail Implementasi

Seluruh eksperimen dilakukan menggunakan Python dalam lingkungan Jupyter Notebook. Model deep learning diimplementasikan menggunakan PyTorch dan Ultralytics YOLOv8n-seg.

2.9 Protokol Evaluasi

Kinerja framework yang diusulkan dievaluasi dari tiga aspek:

- 1) Kinerja Deteksi
Diukur menggunakan precision, recall, F1-score, dan *Intersection over Union* (IoU).
- 2) Robustness terhadap Degradasi
Dievaluasi dengan menerapkan Gaussian noise, salt-and-pepper noise, dan motion blur pada citra masukan selama pengujian.
- 3) Akurasi Kuantifikasi Cacat
Diukur menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE) antara area cacat yang diprediksi dan area cacat ground truth.

2.10 Pengaturan Baseline YOLOv8n-seg

Model segmentasi YOLOv8n-seg (dilatih menggunakan pembagian dataset yang sama (70:15:15) dan pipeline pra-pemrosesan yang sama dengan model U-Net. Untuk memastikan perbandingan yang adil, mask segmentasi dikonversi menjadi anotasi poligon yang diperlukan oleh YOLOv8n-seg. Selain itu, metrik evaluasi yang mencakup precision, recall, F1-score, dan IoU dihitung berdasarkan mask segmentasi yang diprediksi, bukan hanya mengandalkan mAP.

2.11 Metrik Evaluasi

Untuk mengevaluasi kinerja deteksi, digunakan metrik berikut:

Precision:

$$Precision = \frac{TP}{TP + FP}$$

Recall:

$$Recall = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Intersection over Union (IoU):

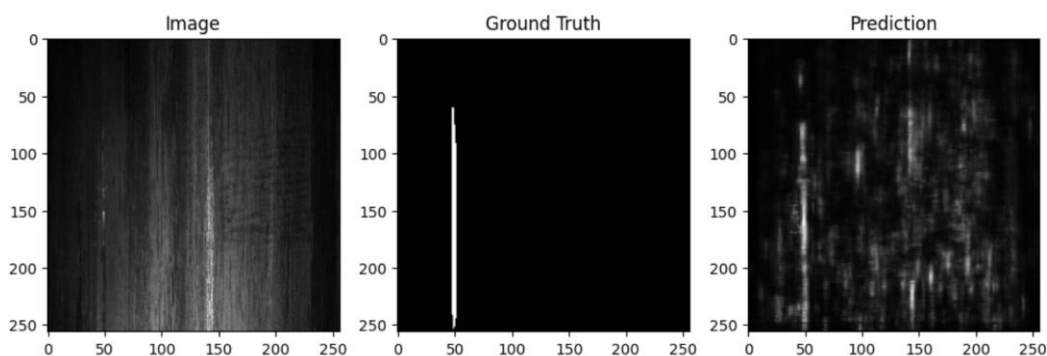
$$IoU = \frac{TP}{TP + FP + FN}$$

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Perbandingan Kinerja Deteksi (Kondisi Bersih)

Kinerja deteksi dari framework yang diusulkan dibandingkan dengan YOLOv8n-seg sebagaimana ditampilkan pada Tabel 1. Hasil menunjukkan bahwa pendekatan segmentasi berbasis U-Net secara signifikan mengungguli model YOLOv8n-seg pada seluruh metrik evaluasi. Secara khusus, model U-Net mencapai precision sebesar 0,6540, recall sebesar 0,3770, F1-score sebesar 0,4641, dan IoU sebesar 0,3098. Sebaliknya, YOLOv8n-seg menghasilkan kinerja yang jauh lebih rendah, dengan precision sebesar 0,2963, recall sebesar 0,2822, F1-score sebesar 0,2611, dan IoU sebesar 0,1971. Perbedaan kinerja tersebut dapat dikaitkan dengan perbedaan mendasar antara pendekatan berbasis segmentasi dan pendekatan berbasis deteksi.



Gambar 2. Hasil kualitatif segmentasi cacat menggunakan metode yang diusulkan, dari kiri ke kanan: citra masukan, mask ground truth, dan segmentasi yang diprediksi

Gambar 2 menunjukkan perbandingan visual antara citra masukan, mask ground truth, dan hasil segmentasi yang diprediksi. Model mampu mengidentifikasi area cacat yang memiliki pola struktural serupa dengan ground truth, terutama pada area cacat yang memanjang. Perbedaan antara mask prediksi dan ground truth menunjukkan kesulitan yang melekat pada segmentasi cacat di lingkungan industri, di mana cacat sering kali bersifat samar dan berada di dalam tekstur yang kompleks. Namun, segmentasi yang diprediksi tampak kurang presisi dan lebih menyebar dibandingkan ground truth.

Hal ini terutama disebabkan oleh rendahnya kontras area cacat serta adanya noise dan variasi tekstur pada permukaan baja yang menyulitkan penentuan batas cacat secara akurat. Meskipun memiliki keterbatasan tersebut, model berhasil menangkap lokasi dan bentuk umum cacat, yang menunjukkan kemampuannya melakukan segmentasi cacat dalam kondisi permukaan yang kompleks. YOLOv8n-seg mengandalkan representasi *bounding box* yang kurang efektif dalam menangkap batas cacat secara rinci pada permukaan logam. Sebaliknya, U-Net melakukan segmentasi pada tingkat piksel sehingga memungkinkan pelokalan area cacat yang tidak beraturan dan berukuran kecil secara lebih akurat. Karakteristik tersebut sangat penting dalam skenario inspeksi industri karena bentuk cacat dapat bersifat kompleks dan samar. Temuan ini menunjukkan bahwa model berbasis segmentasi lebih sesuai dibandingkan model berbasis deteksi untuk analisis cacat permukaan, terutama ketika diperlukan pelokalan yang presisi.

Tabel 1. Perbandingan Kinerja Deteksi (Kondisi Citra Bersih)

Metode	Precision	Recall	F1-score	IoU
U-Net (Diusulkan)	0,6328	0,2614	0,3199	0,2124
YOLOv8n-seg	0,3538	0,3851	0,3347	0,2481

3.1.2 Kinerja pada Kondisi Pencitraan Terdegradasi

Robustness model yang diusulkan pada kondisi pencitraan terdegradasi dievaluasi menggunakan noise dan blur sintetis, yang mencakup Gaussian noise, salt-and-pepper noise, dan motion blur. Hasilnya dirangkum pada Tabel 2. Model mencapai F1-score sebesar 0,4641 dan IoU sebesar 0,3098 pada kondisi bersih. Namun, kinerja menurun seiring meningkatnya degradasi. Pada Gaussian noise, F1-score turun menjadi 0,2721 dan IoU menjadi 0,1632. Degradasi yang lebih parah terlihat pada salt-and-pepper noise dan motion blur, di mana F1-score masing-masing menurun menjadi 0,1768 dan 0,1541, sedangkan IoU turun menjadi 0,0988 dan 0,0924. Meskipun mengalami penurunan tersebut, model tetap mempertahankan tingkat kemampuan deteksi tertentu pada seluruh kondisi. Hal ini menunjukkan robustness framework yang diusulkan dalam menangani skenario pencitraan non-ideal yang umum ditemui pada lingkungan industri dunia nyata. Penting untuk dicatat bahwa F1-score dan IoU dilaporkan sebagai metrik evaluasi utama dalam eksperimen ini karena keduanya memberikan penilaian yang lebih stabil dan representatif untuk tugas segmentasi pada tingkat piksel dalam kondisi terdegradasi dibandingkan precision dan recall.

Tabel 2. Kinerja pada Kondisi Terdegradasi

Kondisi	Precision	Recall	F1-score	IoU
Citra Bersih (Clean)	0,6328	0,2614	0,3199	0,2124
Gaussian noise ($\sigma = 0,10$)	0,0875	0,7274	0,1387	0,0822
Salt & pepper noise ($p = 0,05$)	0,0694	0,9253	0,1169	0,0683
Motion blur (5×5)	0,2368	0,0321	0,0492	0,0309

Selain itu, YOLOv8n-seg menunjukkan penurunan kinerja yang signifikan pada kondisi yang noise. Pada Gaussian noise, F1-score turun menjadi 0,0973 dan IoU menjadi 0,0572. Pada salt-and-pepper noise, kinerja semakin menurun dengan F1-score sebesar 0,0772 dan IoU sebesar 0,0453. Motion blur menghasilkan kinerja yang relatif sedang dengan F1-score sebesar 0,1450 dan IoU sebesar 0,1101. Hasil tersebut menunjukkan bahwa YOLOv8n-seg lebih sensitif terhadap degradasi citra dibandingkan model U-Net yang diusulkan.

3.1.3 Hasil Kuantifikasi Cacat

Akurasi pengukuran area cacat dievaluasi menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE), sebagaimana ditunjukkan pada Tabel 3. Model U-Net mencapai MAE sebesar 2041,28 piksel dan RMSE sebesar 3284,46 piksel. Hasil tersebut menunjukkan bahwa

pendekatan berbasis segmentasi yang diusulkan mampu memberikan pengukuran kuantitatif area cacat yang dapat diandalkan. Estimasi area yang akurat sangat penting dalam aplikasi industri karena memungkinkan penilaian tingkat keparahan cacat secara objektif dan mendukung proses pengambilan keputusan dalam pengendalian kualitas. Perlu diperhatikan bahwa kuantifikasi area cacat dilakukan menggunakan keluaran segmentasi. Oleh karena itu, hanya model berbasis segmentasi seperti U-Net yang dipertimbangkan dalam evaluasi ini. Model berbasis deteksi, termasuk YOLOv8n-seg, tidak sesuai untuk pengukuran presisi pada tingkat piksel karena bergantung pada representasi *bounding box*.

Tabel 3. Akurasi Pengukuran Area Cacat (U-Net)

Metrik	Nilai Piksel	Keterangan / Konteks
Rata-rata Cacat GT ($\bar{A}_{gt}$)	3.774,02 piksel	Basis skala cacat sebenarnya pada test set
Mean Absolute Error (MAE)	2.542,74 piksel	Rata-rata selisih mutlak prediksi piksel
Root Mean Square Error (RMSE)	4.699,42 piksel	Akar rata-rata kuadrat error agregat
Relative MAE (%)	67,37%	$\frac{MAE}{\bar{A}_{gt}} \times 100\%$
Relative RMSE (%)	124,52%	$\frac{RMSE}{\bar{A}_{gt}} \times 100\%$

3.2 Pembahasan

Meskipun pendekatan berbasis U-Net secara konsisten menunjukkan keunggulan dalam pelokalan piksel dibandingkan dengan YOLOv8n-seg, hasil eksperimen mengindikasikan bahwa performa keseluruhan model tersebut masih tergolong modest untuk langsung diterapkan dalam skala industri nyata tanpa adanya modifikasi atau penguatan lebih lanjut. Capaian *Intersection over Union* (IoU) bersih sebesar 0,2124 dan nilai Relative MAE sebesar 67,37% menunjukkan bahwa model cenderung melakukan under-segmentation, yang tercermin dari recall yang jauh lebih rendah (0,2614) dibandingkan precision (0,6328). Kondisi ini dipicu oleh dominasi piksel latar belakang yang mencapai 94,24%, serta belum digunakannya teknik bobot loss (weighted loss) yang dapat membantu model dalam menangani ketidakseimbangan kelas yang ekstrem tersebut. Hasil ini sejalan dengan temuan Luo *et al.* (2020), yang menyatakan bahwa model segmentasi berbasis CNN cenderung mengalami kesulitan dalam mengatasi ketidakseimbangan data dan tekstur kompleks pada citra industri nyata. Penurunan F1-score sebesar 84,61% saat pengujian pada kondisi motion blur menegaskan bahwa model yang hanya dilatih pada citra bersih sangat rentan terhadap kehilangan informasi frekuensi tinggi akibat blur. Hal ini sesuai dengan hasil studi Guan *et al.* (2020), yang menunjukkan bahwa model deteksi dan segmentasi sangat dipengaruhi oleh degradasi citra seperti noise dan blur, sehingga mengurangi keandalan model dalam kondisi pencitraan terganggu. Oleh karena itu, klaim mengenai "keandalan" model harus dipertimbangkan secara proporsional: U-Net terbukti lebih stabil secara relatif dibandingkan YOLOv8n-seg, namun tetap belum cukup matang untuk pengujian di lini produksi tanpa adanya integrasi modul peningkatan citra seperti image restoration atau augmentasi degradasi selama proses pelatihan. Dengan demikian, diperlukan pengembangan lebih lanjut agar model dapat bekerja secara lebih robust dan adaptif terhadap berbagai kondisi pencitraan yang tidak ideal di lingkungan industri.

4. Kesimpulan

Penelitian ini mengusulkan *framework deep learning* berbasis U-Net untuk deteksi dan kuantifikasi luas cacat permukaan pada material logam dalam kondisi terdegradasi. Hasil menunjukkan bahwa U-Net memberikan pelokalan piksel yang lebih baik dibandingkan baseline YOLOv8n-seg. Kuantifikasi luas cacat menghasilkan MAE sebesar 2.542,74 piksel (Relative MAE 67,37%) dan RMSE sebesar

4.699,42 piksel. Penurunan kinerja yang signifikan di bawah gangguan motion blur (penurunan F1-score sebesar 84,61%) mengidentifikasi fokus utama untuk pengembangan selanjutnya. Untuk penelitian mendatang, validasi harus difokuskan pada: (1) penerapan teknik image restoration (seperti *deblurring* dan *denoising CNN*) sebelum tahap segmentasi, (2) penanganan *class imbalance* menggunakan *Focal Loss* atau *Dice Loss*, serta (3) pengujian segmentasi multi-kelas secara *real-time* pada lini produksi manufaktur aktual.

5. Daftar Pustaka

- Chen, X., Liu, M., Niu, Y., Wang, X., & Cheng Wu, Y. (2024). Deep-learning-based lithium battery defect detection via cross-domain generalization. *IEEE Access*, 12, 78505–78514. <https://doi.org/10.1109/ACCESS.2024.3408718>.
- Chen, X., Wu, Y., He, X., & Ming, W. (2023). A comprehensive review of deep learning-based PCB defect detection. *IEEE Access*, 11, 139017–139038. <https://doi.org/10.1109/ACCESS.2023.3339561>.
- Dang, Z., & Wang, X. (2025). FD-YOLO11: A feature-enhanced deep learning model for steel surface defect detection. *IEEE Access*, 13, 63981–63993. <https://doi.org/10.1109/ACCESS.2025.3559733>.
- Gao, Y., Lv, G., Xiao, D., Han, X., Sun, T., & Li, Z. (2024). Research on steel surface defect classification method based on deep learning. *Scientific Reports*, 14. <https://doi.org/10.1038/s41598-024-58643-1>.
- Guan, S., Lei, M., & Lu, H. (2020). A steel surface defect recognition algorithm based on improved deep learning network model using feature visualization and quality evaluation. *IEEE Access*, 8, 49885–49895. <https://doi.org/10.1109/ACCESS.2020.2979755>.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8n-seg. GitHub repository:
- Kaggle. (2019). Severstal: Steel defect detection. Retrieved from
- Lin, H. I., & Wibowo, F. S. (2021). Image data assessment approach for deep learning-based metal surface defect-detection systems. *IEEE Access*, 9, 47621–47638. <https://doi.org/10.1109/ACCESS.2021.3068256>.
- Luo, Q., Fang, X., Liu, L., Yang, C., & Sun, Y. (2020). Automated visual defect detection for flat steel surface: A survey. *IEEE Transactions on Instrumentation and Measurement*, 69, 626–644. <https://doi.org/10.1109/TIM.2019.2963555>.
- Mustafaev, B., Kim, S., & Kim, E. (2023). Enhancing metal surface defect recognition through image patching and synthetic defect generation. *IEEE Access*, 11, 113339–113359. <https://doi.org/10.1109/ACCESS.2023.3322734>.
- Qiao, Q., Hu, H., Ahmad, A., & Wang, K. (2025). A review of metal surface defect detection technologies in industrial applications. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3544578>.

- Ren, F., Fei, J., Li, H., & Doma, B. T. (2024). Steel surface defect detection using improved deep learning algorithm: ECA-SimSPPF-SIoU-Yolov5. *IEEE Access*, 12, 32545–32553. <https://doi.org/10.1109/ACCESS.2024.3371584>.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241). Springer.
- Selamet, F., Cakar, S., & Kotan, M. (2022). Automatic detection and classification of defective areas on metal parts by using adaptive fusion of Faster R-CNN and shape from shading. *IEEE Access*, 10, 126030–126038. <https://doi.org/10.1109/ACCESS.2022.3224037>
- Wan, X., Zhang, X., & Liu, L. (2021). An improved VGG19 transfer learning strip steel surface defect recognition deep neural network based on few samples and imbalanced datasets. *Applied Sciences*, 11. <https://doi.org/10.3390/app11062606>.
- Wang, C., & Xie, H. (2023). MeDERT: A metal surface defect detection model. *IEEE Access*, 11, 35469–35478. <https://doi.org/10.1109/ACCESS.2023.3262264>.
- Wang, C., Zhou, Z., & Chen, Z. (2022). An enhanced YOLOv4 model with self-dependent attentive fusion and component randomized mosaic augmentation for metal surface defect detection. *IEEE Access*, 10, 97758–97766. <https://doi.org/10.1109/ACCESS.2022.3203198>.
- Wang, H., Li, W., Zhang, B., & Gu, Z. (2025). n-GSE: Efficient steel surface defect detection method. *IEEE Access*, 13, 166343–166356. <https://doi.org/10.1109/ACCESS.2025.3602360>.
- Yu, J., Shi, X., Wang, W., & Zheng, Y. (2024). LCG-YOLO: A real-time surface defect detection method for metal components. *IEEE Access*, 12, 41436–41451. <https://doi.org/10.1109/ACCESS.2024.3378999>.
- Zhao, B., Dai, M., Li, P., Xue, R., & Ma, X. (2020). Defect detection method for electric multiple units key components based on deep learning. *IEEE Access*, 8, 136808–136818. <https://doi.org/10.1109/ACCESS.2020.3009654>.
- Zhao, Y., Wang, H., Xie, X., Xie, Y., & Yang, C. (2023). An enhanced YOLOv5-based algorithm for metal surface defect detection. *Applied Sciences*, 13. <https://doi.org/10.3390/app132011473>.
- Zheng, X., Wang, H., Chen, J., Kong, Y., & Zheng, S. (2020). A generic semi-supervised deep learning-based approach for automated surface inspection. *IEEE Access*, 8, 114088–114099. <https://doi.org/10.1109/ACCESS.2020.3003588>.
- Zhou, C., Lu, Z., Lv, Z., Meng, M., Tan, Y., Xia, K., Liu, K., & Zuo, H. (2023). Metal surface defect detection based on improved YOLOv5. *Scientific Reports*, 13. <https://doi.org/10.1038/s41598-023-47716-2>.
- Zhu, H., Wang, Y., & Fan, J. (2022). IA-Mask R-CNN: Improved anchor design Mask R-CNN for surface defect detection of automotive engine parts. *Applied Sciences*, 12. <https://doi.org/10.3390/app12136633>.