

Segmentasi Citra Cacat Industri Menggunakan CNN dengan Contrastive Learning

Fajar Rahmat Akbar¹, Arief Suryadi Satyawan², Mokh Mirza Etnisa Haqiqi³, Andika Muhammad Nur Kholiq⁴, Iasya Faiqoh Nurrohmah⁵, Aulia Adawiyah⁶, Esti Fitria Wulandari⁷, Rendi Tri Sugian⁸, Nurul Fazri⁹

^{1,3,4,5,6,7,8,9} Teknik Elektro, Universitas Garut

² Badan Riset dan Inovasi Nasional

¹24052122044@ftechnik.uniga.ac.id

²arie021@brin.go.id

³mirza@uniga.ac.id

⁴24052122040@ftechnik.uniga.ac.id

⁵24052122011@ftechnik.uniga.ac.id

⁶24052122002@ftechnik.uniga.ac.id

⁷24052122003@ftechnik.uniga.ac.id

⁸24052122025@ftechnik.uniga.ac.id

⁹24052320032@ftechnik.uniga.ac.id

*Corresponding author email: 24052122044@ftechnik.uniga.ac.id

Abstrak— Penelitian ini dilatarbelakangi oleh kebutuhan peningkatan akurasi segmentasi citra cacat pada inspeksi visual industri, terutama karena model CNN konvensional masih cenderung kurang optimal dalam membedakan area cacat dan latar belakang. Penelitian ini mengusulkan integrasi pembelajaran kontrasitif berbasis InfoNCE ke dalam arsitektur CNN encoder-decoder melalui penambahan projection head dan Proposal Contrastive Loss untuk memperkuat pemisahan representasi antara area cacat dan latar belakang. Model diuji melalui pelatihan dan validasi, kemudian dievaluasi menggunakan IoU, Dice, Precision, dan Recall. Hasil eksperimen menunjukkan bahwa pendekatan yang diusulkan secara konsisten meningkatkan performa segmentasi dibandingkan model baseline, ditunjukkan oleh peningkatan IoU dari 0,21 menjadi 0,66, Dice dari 0,31 menjadi 0,77, Precision dari 0,67 menjadi 0,78, dan Recall dari 0,21 menjadi 0,81. Secara kualitatif, hasil segmentasi menunjukkan kesesuaian yang lebih tinggi terhadap ground truth dengan batas objek yang lebih konsisten. Temuan ini menunjukkan bahwa integrasi pembelajaran kontrasitif mampu meningkatkan kualitas segmentasi citra cacat pada sistem inspeksi visual industri.

Kata Kunci— Segmentasi Citra Cacat, CNN, Contrastive Learning, InfoNCE, ComboLoss, Inspeksi Visual Industri, Deep Learning.

I. PENDAHULUAN

Citra cacat (misalnya cacat permukaan pada komponen industri) memegang peranan penting dalam kontrol kualitas manufaktur, karena deteksi dini cacat dapat mencegah kerugian lebih lanjut [1]. Metode berbasis *deep learning* telah menjadi sorotan utama dalam bidang ini. Secara khusus, jaringan saraf konvolusional (CNN) terbukti mampu menangkap fitur visual kompleks dari citra, sehingga banyak diterapkan pada

segmentasi area cacat [1]. Misalnya, studi terbaru menunjukkan bahwa CNN dan teknik segmentasi piksel tingkat lanjut memungkinkan pendeteksian cacat permukaan seperti retakan atau penyimpangan dimensi dengan presisi tinggi [1]. Meski demikian, tantangan nyata masih ada, seperti keterbatasan data berlabel dan kebutuhan komputasi tinggi untuk arsitektur yang kompleks [2]. Oleh karena itu, perlu pendekatan baru yang meningkatkan kualitas segmentasi sambil menjaga efisiensi, terutama dalam konteks pengolahan data besar atau real-time. Berbagai arsitektur CNN encoder-decoder populer (misalnya variasi U-Net atau DeepLab) telah banyak digunakan dalam segmentasi citra medis dan industri. Teknik *loss* kombinasi seperti *ComboLoss* (gabungan *Dice Loss* dan *Cross-Entropy*) kerap dipakai untuk mengatasi ketidakseimbangan kelas pada segmen kecil [3]. Namun, peningkatan presisi segmentasi juga telah dikejar dengan strategi pembelajaran representasi tambahan. Salah satunya adalah *contrastive learning*, yaitu metode pembelajaran tanpa pengawasan yang memperkaya fitur dengan membedakan contoh “positif” (mirip) dan “negatif” (tidak mirip) [4]. Dalam konteks segmentasi, Zhao et al. (ICCV 2021) mengintegrasikan *contrastive loss* berbasis label piksel untuk *semantic segmentation*, sehingga model belajar menarik bersama piksel dari kelas sama dan menjauhkan piksel kelas berbeda, terbukti meningkatkan akurasi segmentasi [5]. Quan et al. (ICCV 2023) juga mengusulkan algoritma *contrastive* piksel-ke-piksel untuk segmentasi, dan melaporkan pencapaian *mIoU* 79.2% pada dataset *Cityscapes* lebih baik 1.1 poin dibanding *DeepLabv3* sebelumnya [6]. Begitu pula, Wang et al. (CVPR 2022) menggunakan *contrastive learning* untuk instansiasi mask segmentasi (*ContrastMask*), membuktikan pendekatan mereka unggul dan melampaui metode *state-of-the-art* pada dataset COCO [7]. Temuan ini konsisten dengan prinsip SimCLR yang memaksimalkan *agreement* antar

augmentasi citra melalui loss InfoNCE [4], sehingga representasi fitur menjadi lebih diskriminatif. Penelitian lain juga menggabungkan pembelajaran konsistensi dengan *contrastive* untuk segmentasi semi-terawasi, menunjukkan bahwa kombinasi ini menjanjikan dalam kondisi data terbatas [8], [9]. Dari telaah literatur singkat di atas, diketahui bahwa meski *contrastive learning* banyak dimanfaatkan dalam segmentasi umum atau medis, penerapannya khusus untuk segmentasi citra cacat industri belum begitu luas dilaporkan.

Berdasarkan kajian di atas, terdapat kesenjangan signifikan dalam penelitian segmentasi cacat: sebagian besar model *CNN* tradisional masih mengandalkan loss piksel-ke-piksel tanpa secara eksplisit memperkuat pembelajaran perbedaan fitur antar wilayah cacat dan latar belakang. Sementara itu, metode *contrastive* yang mampu meningkatkan diskriminasi fitur belum banyak diadaptasi untuk tugas segmentasi cacat industri. *Novelty* penelitian ini adalah menggabungkan arsitektur *CNN encoder-decoder baseline* dengan cabang pembelajaran representasi kontrasif berbasis InfoNCE. Dalam pendekatan yang diusulkan, seluruh jalur utama *encoder-decoder* dipertahankan untuk segmentasi tradisional, namun ditambahkan kepala (*head*) pembelajaran kontrasif yang memproses patch citra sebagai pasangan positif-negatif [10]. Pendekatan ini memungkinkan model mempelajari representasi fitur yang lebih diskriminatif sehingga perbedaan antara area cacat dan area non-cacat dapat dikenali secara lebih efektif. Pendekatan ini berbeda dengan studi sebelumnya karena menekankan sekaligus pengoptimalan segmentasi piksel dan representasi fitur melalui *contrastive learning* dalam satu kerangka *end-to-end*. Kontribusi utama penelitian ini terletak pada integrasi Proposal Contrastive Loss berbasis InfoNCE dengan *ComboLoss* dalam satu kerangka pembelajaran *end-to-end* untuk segmentasi citra cacat industri [10], [7].

Permasalahan utama penelitian ini adalah bagaimana meningkatkan ketelitian segmentasi area cacat pada citra sambil meminimalkan kesalahan prediksi latar belakang. Dengan tambahan loss kontrasif, diasumsikan model dapat lebih baik membedakan pola cacat dan non-cacat. Dari permasalahan tersebut muncul beberapa hipotesis: (1) Hipotesis 1: Menambahkan term *loss* kontrasif (InfoNCE) ke dalam fungsi objektif segmentasi akan meningkatkan metrik evaluasi (misalnya *Dice* atau *IoU*) dibanding model *baseline* tanpa kontrasif [7]. (2) Hipotesis 2: Fitur yang dipelajari menjadi lebih stabil dan *discriminative*, sehingga pelatihan model lebih cepat konvergen dan hasil segmentasi lebih konsisten [7]. Dengan demikian, pembelajaran kontrasif diharapkan mampu meningkatkan sensitivitas model terhadap variasi pola cacat sekaligus mempertahankan kemampuan diskriminasi terhadap area latar belakang.

Untuk menguji hipotesis di atas, digunakan dua varian model *CNN*: satu adalah *baseline encoder-decoder CNN* konvensional untuk segmentasi (menggunakan *ComboLoss*) dan yang kedua adalah *CNN + Contrastive*, yaitu *baseline* yang dilengkapi kepala (*head*) *contrastive learning*

berbasis InfoNCE [10]. Diagram arsitektur keseluruhan digambarkan secara skematis dimana setelah proses *encoding feature map* dilewatkan ke dua *branch: decoder segmentation* serta *branch* kontrasif. *Loss* total pada varian kontrasif merupakan gabungan antara *ComboLoss segmen* dan InfoNCE *loss*, dengan bobot pengatur λ untuk menyeimbangkan kontribusinya [10], [4]. Langkah-langkah implementasi meliputi:

- Membangun jaringan *encoder-decoder CNN* (misalnya arsitektur *U-Net* atau sejenis) sebagai *baseline* segmentasi. Guna menanggulangi ketidakseimbangan kelas (cacat seringkali lebih kecil area), digunakan *ComboLoss* (gabungan *Dice Loss + Binary Cross-Entropy*) pada *baseline* [3].
- Mengintegrasikan cabang kontrasif dengan konsep paired InfoNCE: setiap patch piksel (sebagai anchor) dicocokkan dengan augmentasi (positif) serta patch latar (negatif). Fungsi loss InfoNCE mendorong fitur *anchor* dan positif mendekat, serta menjauhkan *anchor* dari *negative* [10], [4].
- Melatih model secara simultan dengan meminimalkan loss gabungan:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{segmentasi}} + \lambda \mathcal{L}_{\text{contrastive}} \quad (1)$$

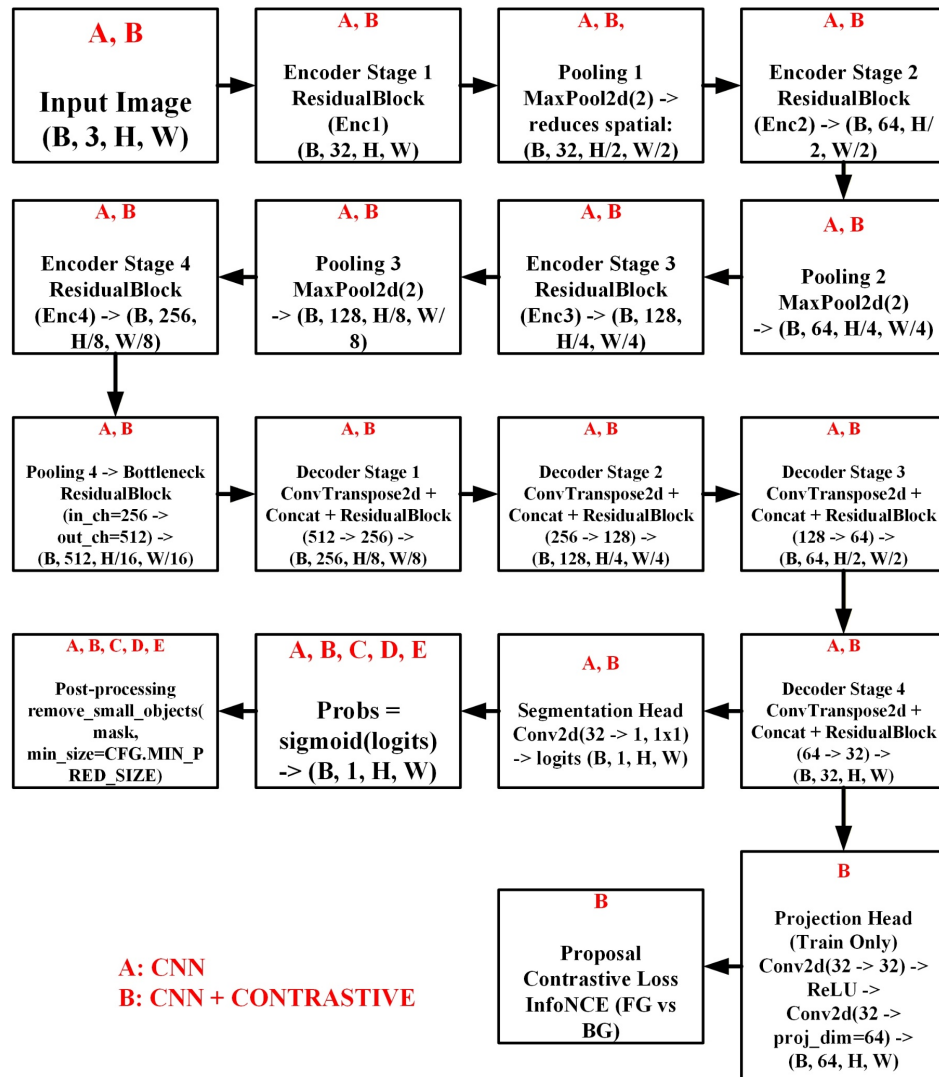
di mana λ adalah koefisien pengimbang [10]. Parameter diperbarui dengan *stochastic gradient descent* atau algoritma sejenis.

- Melakukan optimasi dan validasi model dengan memonitor *loss* dan metrik *IoU/Dice* pada data pelatihan dan validasi.

Dari pendekatan tersebut diharapkan peningkatan kinerja segmentasi area cacat dibandingkan *baseline*. Secara kuantitatif, diharapkan nilai metrik *overlap (Dice, IoU)* meningkat secara signifikan, sebagaimana metode *contrastive* sebelumnya terbukti meningkatkan akurasi prediksi segmentasi [7]. Secara kualitatif, diharapkan visualisasi prediksi menunjukkan area cacat yang lebih lengkap dan sedikit noise latar [7]. Tujuan akhir penelitian ini adalah menunjukkan bahwa tambahan pembelajaran kontrasif dapat membantu model *CNN* untuk memahami konteks semantik area cacat lebih baik, sehingga segmentasi cacat menjadi lebih presisi. Dengan demikian, penelitian ini bertujuan menyumbangkan teknik baru yang dapat diterapkan pada sistem inspeksi visual untuk meningkatkan deteksi cacat pada proses produksi.

II. METODE

Sebagaimana ditunjukkan pada Gbr 1., alur kerja penelitian ini terdiri atas dua varian model, yaitu model *CNN* sebagai *baseline* dan model *CNN + Contrastive* sebagai pengembangan dari *baseline* tersebut. Kedua model menerima citra masukan berukuran $B \times 3 \times H \times W$, dengan B menyatakan ukuran batch, tiga kanal warna, serta dimensi spasial tinggi dan lebar citra. Pada tahap awal, citra diproses melalui *encoder* bertingkat yang terdiri atas *residual block* dan operasi *downsampling* menggunakan *max pooling*.



Gbr 1. Pipeline Arsitektur Model

Konfigurasi encoder bertingkat digunakan untuk mengekstraksi representasi multi-level yang mencakup karakteristik lokal maupun informasi semantik tingkat tinggi. Pada arsitektur *CNN baseline*, *encoder* menghasilkan representasi fitur yang semakin kaya pada setiap tahap, mulai dari jumlah kanal 32, 64, 128, hingga 256, sebelum masuk ke bottleneck dengan 512 kanal.

Setelah proses *encoding*, fitur melewati *bottleneck* sebagai representasi paling padat dari citra input. Bottleneck berfungsi sebagai representasi laten yang merangkum informasi spasial dan semantik sebelum proses rekonstruksi pada decoder. Dari *bottleneck*, fitur kemudian diteruskan ke *decoder* melalui serangkaian blok *upsampling* menggunakan *ConvTranspose2d*, disertai *skip connection* dari *layer encoder* yang bersesuaian. Mekanisme *concatenation* pada *skip connection* mempertahankan detail spasial yang hilang selama proses *downsampling*, sehingga prediksi segmentasi yang dihasilkan menjadi lebih presisi pada batas objek. Alur ini terlihat jelas

pada Gbr 1., di mana *decoder* bergerak bertahap dari resolusi rendah kembali ke resolusi awal hingga menghasilkan keluaran dengan dimensi spasial yang sama seperti input.

Pada tahap akhir arsitektur *CNN baseline*, fitur hasil *decoder* dipetakan melalui *segmentation head* berupa konvolusi 1×1 untuk menghasilkan logit dengan ukuran $B \times 1 \times H \times W$. Logit ini kemudian diubah menjadi probabilitas menggunakan fungsi *sigmoid* saat inferensi. Secara konseptual, keluaran *sigmoid* merepresentasikan probabilitas setiap piksel termasuk kelas cacat atau latar belakang. Setelah itu dilakukan *post-processing* menggunakan *remove_small_objects* untuk menghilangkan komponen prediksi berukuran sangat kecil yang berpotensi merupakan *noise*, sehingga hasil segmentasi menjadi lebih bersih dan lebih konsisten secara morfologis. Ringkasan proses utama masing-masing model juga telah dirangkum pada Tabel I., yang memperlihatkan bahwa model *CNN* menggunakan *pipeline encoder-decoder* standar,

sedangkan model *CNN + Contrastive* menambahkan cabang pembelajaran representasi tambahan.

Berbeda dengan *baseline*, model *CNN + Contrastive* mempertahankan seluruh jalur utama *encoder-decoder* yang sama, namun menambahkan *projection head* pada tahap pelatihan. Berdasarkan Gbr 1., *projection head* hanya aktif saat *training* dan menerima fitur dari keluaran *decoder* untuk diproyeksikan ke ruang *embedding* berdimensi lebih rendah. *Projection head* digunakan untuk memetakan fitur ke ruang *embedding* sehingga hubungan antar representasi dapat dioptimalkan menggunakan objective function berbasis *contrastive learning*. Dengan demikian, model tidak hanya belajar melakukan segmentasi berdasarkan kesesuaian piksel, tetapi juga belajar memisahkan distribusi fitur antar-kelas secara eksplisit. Pada pendekatan ini, *loss* segmentasi tetap digunakan sebagai objektif utama, sedangkan *loss* kontrasif berperan sebagai *regularizer* representasional yang memperkuat kemampuan model dalam membedakan area cacat dari area non-cacat.

Secara matematis, fungsi *loss* pada model *baseline* dirumuskan sebagai *ComboLoss*, yaitu gabungan antara *Binary Cross-Entropy (BCE)* dan *Dice Loss*. Untuk N piksel, *BCE* dituliskan sebagai:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \quad (2)$$

dengan $y_i \in \{0,1\}$ sebagai label ground truth dan p_i sebagai probabilitas prediksi pada piksel ke- i . Sementara itu, *Dice Loss* diformulasikan sebagai:

$$\mathcal{L}_{DICE} = 1 - \frac{2 \sum_{i=1}^N p_i y_i + \varepsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N y_i + \varepsilon} \quad (3)$$

dengan ε sebagai konstanta kecil untuk menghindari pembagian dengan nol. Kombinasi keduanya dapat ditulis sebagai:

$$\mathcal{L}_{combo} = \alpha \mathcal{L}_{BCE} + \beta \mathcal{L}_{DICE} \quad (4)$$

di mana α dan β merupakan bobot penyeimbang. Kombinasi *BCE* dan *Dice Loss* dipilih untuk menyeimbangkan kemampuan klasifikasi piksel individual dan kemampuan menangani ketidakseimbangan kelas yang umum ditemukan pada segmentasi cacat industri.

Pada eksperimen ini, *ComboLoss* dinyatakan sebagai kombinasi *Binary Cross-Entropy* dan *Dice Loss* dengan bobot masing-masing $\alpha = 0,3$ dan $\beta = 0,7$. Pada model *CNN + Contrastive*, fungsi objektif total juga ditambah dengan *loss* kontrasif berbasis *InfoNCE* dengan koefisien pengimbang $\lambda = 0,1$ dan temperatur $\tau = 0,1$.

Untuk model *CNN + Contrastive*, *loss* total diperluas dengan *Proposal Contrastive Loss* berbasis *InfoNCE*, sesuai alur pada

Gbr 1. Dengan *embedding* hasil *projection head* dinyatakan sebagai z_i , *loss* kontrasif dapat dirumuskan sebagai:

$$\mathcal{L}_{InfoNCE} = -\frac{1}{M} \sum_{i=1}^B \log \frac{\exp(\frac{\text{sim}(z_i, z_i^+)}{\tau})}{\sum_{j=1}^K \exp(\frac{\text{sim}(z_i, z_j^+)}{\tau})} \quad (5)$$

dengan $\text{sim}(\cdot)$ menyatakan ukuran kemiripan, z_i^+ adalah pasangan positif, z_k adalah sampel negatif, τ adalah parameter temperatur, dan M jumlah pasangan data yang digunakan. Pada konteks penelitian ini, pasangan positif merepresentasikan fitur yang sejenis, sedangkan pasangan negatif digunakan untuk mendorong pemisahan *embedding* antara *foreground* dan *background*. Karena itu, fungsi objektif akhir model *contrastive* menjadi:

$$\mathcal{L}_{Total} = \mathcal{L}_{Combo} + \lambda_{Con} \mathcal{L}_{InfoNCE} \quad (6)$$

dengan λ sebagai koefisien pengimbang kontribusi *loss* kontrasif terhadap *loss* segmentasi.

Proses optimasi dilakukan dengan meminimalkan fungsi objektif tersebut melalui pembaruan parameter jaringan secara iteratif. Secara umum, pembaruan parameter dapat dituliskan sebagai:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} \mathcal{L}_{Total} \quad (7)$$

dengan θ_t adalah parameter pada iterasi ke- t , η laju pembelajaran, dan $\nabla_{\theta} \mathcal{L}_{total}$ gradien *loss* terhadap parameter model. Pada penelitian ini proses optimasi dilakukan menggunakan optimizer Adam sehingga pembaruan parameter memanfaatkan estimasi momen pertama dan kedua dari gradien untuk meningkatkan stabilitas konvergensi. Dengan demikian, perbedaan utama kedua model tidak hanya terletak pada struktur keluaran, tetapi juga pada strategi pembelajaran: *baseline CNN* berfokus pada optimasi segmentasi piksel, sedangkan *CNN + Contrastive* memperkaya proses pembelajaran dengan pemisahan representasi fitur *foreground* dan *background* secara eksplisit.

III. HASIL DAN PEMBAHASAN

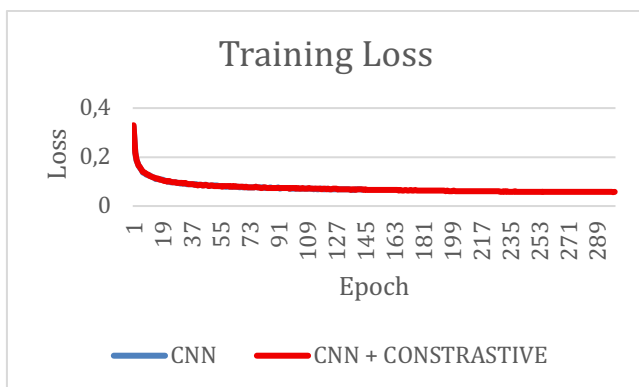
Merujuk Gbr 1. dan Tabel I., evaluasi eksperimen pada penelitian ini dilakukan dengan membandingkan dua rancangan model, yaitu *CNN baseline* dan *CNN + Contrastive*. Pada model *baseline*, alur kerja dimulai dari citra masukan yang diproses oleh *encoder* berbasis *residual* dan *SE block*, diteruskan ke *bottleneck*, kemudian direkonstruksi melalui *decoder* dengan *skip connection* hingga menghasilkan prediksi segmentasi melalui *segmentation head*. Model ini dioptimalkan menggunakan *ComboLoss* dan hasil akhirnya diperhalus dengan *remove_small_objects*. Sementara itu, model *CNN + Contrastive* mempertahankan jalur segmentasi utama yang sama, tetapi menambahkan *projection head* pada tahap

TABEL I.
MODIFIKASI ARSITEKTUR

Model	Proses ringkas (langkah utama)
CNN	Input → Encoder (Residual + SE Block) → Bottleneck → Decoder (ConvTranspose + Skip Connection) → Segmentation Head (1x1 Conv) → logit → sigmoid saat inferensi. Loss: ComboLoss (BCE + Dice). Post-processing: remove small objects.
CNN + Contrastive	Input → Encoder → Bottleneck → Decoder → Mask Prediksi. Cabang tambahan (Hanya saat Training): Fitur akhir Decoder → <i>Projection Head</i> (ekstraksi fitur). Loss gabungan: ComboLoss + Proposal Contrastive Loss (InfoNCE) untuk menjauhkan vektor fitur <i>Foreground</i> (cacat) dari <i>Background</i> (papan). Post-processing: remove_small_objects.

pelatihan serta menggabungkan *ComboLoss* dengan Proposal *Contrastive Loss* (InfoNCE) untuk memperkuat pemisahan representasi *foreground* dan *background*. Dengan demikian, perbedaan utama kedua model terletak pada keberadaan cabang pembelajaran representasi tambahan pada model *contrastive*.

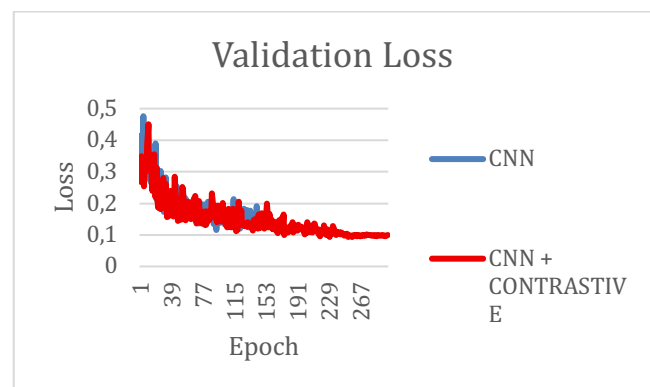
Berdasarkan Gbr 2. dan Gbr 3., dinamika pelatihan menunjukkan bahwa kedua model mengalami penurunan loss secara bertahap selama proses training dan validasi. Kurva training loss memperlihatkan kecenderungan konvergensi pada kedua model, meskipun model *CNN + Contrastive* memulai dengan nilai *loss* yang sedikit lebih tinggi pada fase awal pelatihan. Setelah beberapa *epoch*, *loss* model *contrastive* turun lebih stabil dan cenderung mencapai kondisi yang lebih baik pada fase akhir. Pola yang serupa juga terlihat pada validation loss, di mana model *baseline* memperlihatkan fluktuasi yang lebih terasa dibanding model *contrastive*. Secara umum, hasil ini menunjukkan bahwa penambahan loss kontrasif tidak menghambat proses optimasi, melainkan membantu model belajar representasi yang lebih stabil selama pelatihan.



Gbr 2. Perbandingan Training Loss

Hasil kuantitatif pada Tabel II. memperlihatkan peningkatan yang sangat jelas ketika pendekatan *contrastive* diterapkan. Model *CNN* hanya mencapai *IoU* 0,21, *Dice* 0,31, *Precision* 0,67, dan *Recall* 0,21. Sebaliknya, model *CNN + Contrastive* menghasilkan *IoU* 0,66, *Dice* 0,77, *Precision* 0,78, dan *Recall* 0,81. Peningkatan ini menunjukkan bahwa strategi *contrastive*

learning memberikan dampak paling besar pada kemampuan model dalam menangkap area target secara utuh, yang tercermin dari kenaikan *Recall* yang paling menonjol, yaitu +0,60, disusul oleh peningkatan *IoU* +0,45 dan *Dice* +0,47 sebagaimana dilaporkan pada tabel. Dengan kata lain, model *contrastive* tidak hanya lebih akurat dalam memprediksi piksel positif, tetapi juga lebih konsisten dalam menyeimbangkan *overlap* prediksi terhadap *ground truth*.



Gbr 3. Perbandingan Validation Loss

TABEL II.
RATA-RATA METRIKS EVALUASI MODEL

Model	IoU	Dice	Precision	Recall
CNN	0,21	0,31	0,67	0,21
CNN + Contrastive	0,66	0,77	0,78	0,81

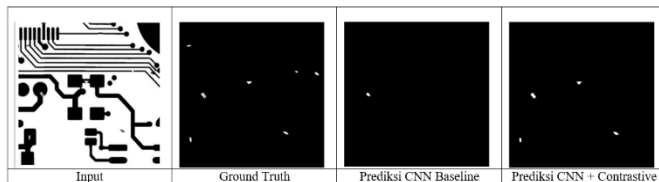
Berdasarkan nilai *precision* dan *recall* pada Tabel II., model *CNN* *baseline* menunjukkan kecenderungan menghasilkan *false negative* yang tinggi. Nilai *recall* yang hanya mencapai 0,21 mengindikasikan bahwa sebagian besar area cacat belum berhasil tersegmentasi dengan baik. Kondisi ini sejalan dengan hasil analisis *confusion matrix* pada Tabel III., di mana jumlah *false negative* pada model *CNN* *baseline* masih relatif tinggi. Sebaliknya, model *CNN + Contrastive* memperoleh *recall* sebesar 0,81 dan *precision* sebesar 0,78, yang menunjukkan

kemampuan lebih baik dalam mengidentifikasi area cacat secara menyeluruh. Penurunan jumlah *false negative* yang terlihat pada Tabel III. menunjukkan bahwa pembelajaran kontrasif mampu meningkatkan sensitivitas model terhadap keberadaan cacat. Dengan demikian, integrasi pembelajaran kontrasif tidak hanya meningkatkan nilai recall, tetapi juga memperbaiki keseimbangan antara kemampuan deteksi dan ketepatan prediksi yang ditunjukkan oleh nilai precision yang tetap tinggi.

TABEL III.
Confusion Matrix

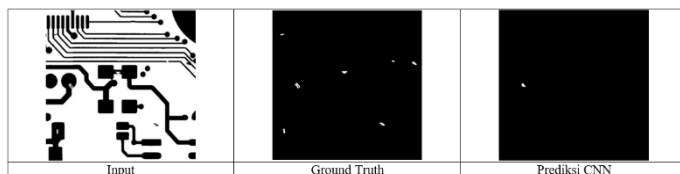
Model	TP	FP	FN	TN
CNN	1.494	51	4.756	2.615.139
CNN + Contrastive	5.196	1.441	1.054	2.613.749

Secara kualitatif, perbandingan pada Gbr 4. mendukung temuan numerik tersebut. Visualisasi sampel hasil uji menunjukkan bahwa keluaran model *CNN baseline* masih kurang representatif dalam menangkap bentuk area segmentasi secara menyeluruh, sedangkan *CNN + Contrastive* menghasilkan mask yang secara visual lebih dekat dengan target segmentasi dan tampak lebih terstruktur. Temuan ini menguatkan interpretasi bahwa penambahan cabang *projection head* dan *loss* kontrasif membantu pemisahan fitur *foreground* dan *background*, sehingga model mampu menghasilkan prediksi yang lebih presisi dan lebih lengkap pada data uji.



Gbr 4. Sampel Hasil Testing

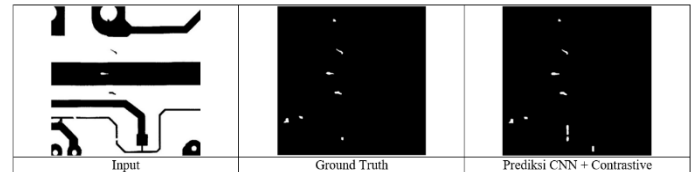
Gbr 5. menunjukkan contoh kasus false negative pada model CNN baseline, di mana sebagian area cacat gagal dikenali dan diklasifikasikan sebagai latar belakang, model CNN gagal mendeteksi sebagian besar area cacat sehingga menghasilkan false negative yang tinggi. Kesalahan ini diduga disebabkan oleh ukuran cacat yang relatif kecil dan kontras yang rendah terhadap latar belakang.



Gbr 5. Contoh False Negative pada CNN

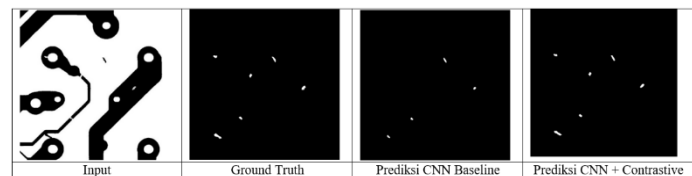
Gbr 6. memperlihatkan contoh false positive pada model CNN + Contrastive yang muncul akibat meningkatnya sensitivitas

model terhadap pola visual yang menyerupai cacat, meskipun model CNN + Contrastive berhasil mendeteksi area cacat utama, masih terdapat beberapa piksel latar belakang yang salah diklasifikasikan sebagai cacat. Fenomena ini menyebabkan peningkatan false positive, namun secara keseluruhan tidak mengurangi kualitas segmentasi secara signifikan.



Gbr 6. Contoh False Positive pada CNN + Contrastive

Gbr 7. menunjukkan contoh segmentasi yang berhasil, di mana hasil prediksi model CNN + Contrastive memiliki kesesuaian yang tinggi terhadap ground truth baik dari sisi bentuk maupun cakupan area cacat, model CNN + Contrastive menghasilkan segmentasi yang lebih mendekati ground truth dibandingkan model CNN baseline. Batas area cacat tampak lebih lengkap dan lebih konsisten.



Gbr 7. Contoh Segmentasi Berhasil

IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penambahan pembelajaran kontrasif pada arsitektur *CNN encoder-decoder* mampu meningkatkan kualitas segmentasi citra cacat secara signifikan dibandingkan model baseline. Model *CNN + Contrastive* yang mengombinasikan *ComboLoss* dan *Proposal Contrastive Loss* (InfoNCE) menunjukkan proses pelatihan yang lebih stabil, serta menghasilkan peningkatan metrik evaluasi yang nyata, yaitu *IoU* dari 0,21 menjadi 0,66, *Dice* dari 0,31 menjadi 0,77, *Precision* dari 0,67 menjadi 0,78, dan *Recall* dari 0,21 menjadi 0,81. Hasil visual juga memperlihatkan bahwa mask segmentasi yang dihasilkan lebih presisi dan lebih mendekati *ground truth*, sehingga pendekatan yang diusulkan efektif dalam memperkuat pemisahan fitur *foreground* dan *background*. Selain meningkatkan nilai *IoU*, *Dice*, *Precision*, dan *Recall*, pendekatan yang diusulkan juga mampu menurunkan jumlah false negative secara signifikan berdasarkan analisis confusion matrix berbasis piksel. Dengan demikian, penelitian ini menegaskan bahwa integrasi *loss* kontrasif ke dalam kerangka segmentasi *CNN* merupakan strategi yang menjanjikan untuk meningkatkan akurasi deteksi area cacat pada citra industri.

UCAPAN TERIMA KASIH

Dalam penelitian ini, penulis menyampaikan terima kasih kepada Fakultas Teknik Universitas Garut dan Badan Riset dan Inovasi Nasional (BRIN) atas dukungan berupa sumber daya serta fasilitas yang telah menunjang kelancaran pelaksanaan penelitian.

REFERENSI

- [1] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, "Surface defect detection methods for industrial products: A review," *Appl. Sci.*, vol. 11, no. 16, p. 7657, 2021.
- [2] C. Sung, W. Kim, J. An, W. Lee, H. Lim, and H. Myung, "Contextextrat: Contextual Contrastive Learning for Semantic Segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 3732–3742.
- [3] R. Azad *et al.*, "Loss functions in the era of semantic segmentation: A survey and outlook," *arXiv Prepr. arXiv2312.05391*, 2023.
- [4] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning (ICML)*, 2020, pp. 1597–1607.
- [5] X. Zhao *et al.*, "Contrastive learning for label efficient semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 10623–10633.
- [6] S. Quan, M. Hirano, and Y. Yamakawa, "Semantic Information in Contrastive Learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 5686–5696.
- [7] X. Wang, K. Zhao, R. Zhang, S. Ding, Y. Wang, and W. Shen, "Contrastmask: Contrastive learning to segment every thing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11604–11613.
- [8] Z. Wang and C. Ma, "Dual-Contrastive Dual-Consistency Dual-Transformer: A Semi-Supervised Approach to Medical Image Segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023, pp. 870–879.
- [9] C. You *et al.*, "Rethinking Semi-Supervised Medical Image Segmentation: A Variance-Reduction Perspective," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [10] J. Huang, S. A. Ariffin, Q. Yang, K. Yong, and W. Xu, "Contrastive learning enhanced MobileMamba for real time industrial defect detection on edge devices," *arXiv Prepr.*, pp. 1–15, 2026.