GeoSeverity-RDD: *Framework* Deteksi dan Estimasi Keparahan Kerusakan Jalan Indonesia Berbasis *Multi-Task* YOLOv11 dengan Proxy Severity Geometris

*Puteri Tonisa[1], Mega Suci Lestari [2]
[1,2] Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Tangerang Selatan, Banten
e-mail: putritonisa@gmail.com[1]; mega.sucilestari@gmail.com[2]

***Abstrak*— Penelitian ini mengusulkan GeoSeverity-RDD, kerangka kerja multi-tugas yang secara simultan mendeteksi jenis kerusakan jalan dan mengestimasi tingkat keparahannya (ringan/sedang/berat) dalam satu inferensi tunggal pada dataset jalan Indonesia. *Novelty* utama terletak pada Proxy Severity Geometris (PSG) yaitu mekanisme pemberian label *severity* otomatis yang diturunkan dari fitur geometri *bounding box* berupa area untuk *pothole* dan panjang dominan untuk *cracking* tanpa memerlukan anotasi manual tambahan. Dataset Road Damage Indonesia v5 (format COCO) yang digunakan terdiri dari 3.321 citra beresolusi 640×640 piksel dengan 6.999 anotasi mencakup empat kelas yaitu *pothole*, *longitudinal cracking*, *lateral cracking*, dan *alligator cracking*. Setelah penerapan PSG dengan ambang batas persentil P_{33}/P_{67}, distribusi label *severity* mencapai keseimbangan 33% per level yang berhasil mengeliminasi *class imbalance*. Arsitektur GeoSeverity-RDD diimplementasikan dengan menambahkan *Severity Estimation Head* (SEH) ringan secara paralel pada *detection head* YOLOv11m, serta berbagi *backbone* C3k2+SPPF dan *neck* C2PSA+FPN/PAN. Hasil eksperimen menunjukkan model mencapai mAP@0.5 sebesar 71,3% untuk *detection task* dengan degradasi minimal 1,1% dari *baseline*. Pada *severity task*, model menghasilkan akurasi sebesar 78,6% dan Macro F1 78,2% dengan kecepatan inferensi 34,7 FPS (*real-time*). Integrasi model ini efisien dengan *overhead* parameter hanya meningkat 6,5% dan penambahan ukuran model sebesar 2,8 MB.**

***Kata Kunci*— GeoSeverity-RDD, Proxy Severity Geometris, Road Damage Detection, *Severity* Estimation, YOLOv11**

I. PENDAHULUAN

Infrastruktur jalan merupakan tulang punggung mobilitas ekonomi dan sosial Indonesia. Dengan total panjang jaringan jalan lebih dari 540.000 km, Indonesia menghadapi tantangan besar dalam mempertahankan kualitas jalan yang memadai. Data Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR) menunjukkan lebih dari 35% jalan nasional dan lebih dari 50% jalan provinsi dalam kondisi rusak [1], [2] berdampak langsung pada kecelakaan lalu lintas, biaya operasional kendaraan, dan efisiensi logistik nasional. Sistem inspeksi jalan konvensional bergantung pada survei manual oleh petugas lapangan yang tidak hanya membutuhkan waktu dan biaya besar, tetapi juga tidak dapat menghasilkan data secara *real-time* dan konsisten. Otomasi inspeksi menggunakan *computer vision* dan *deep learning* telah terbukti mampu mendeteksi jenis-jenis kerusakan jalan secara akurat dan efisien. Namun, terdapat gap mendasar yang belum diatasi oleh sebagian besar penelitian yang ada dimana sistem yang ada hanya menghasilkan keluaran berupa jenis kerusakan dan lokasinya (*bounding box*), tanpa informasi tentang seberapa parah kerusakan tersebut.

Informasi keparahan atau *severity* sangat kritis dalam pengambilan keputusan pemeliharaan jalan. Sebuah retakan kecil (*severity* ringan) dapat ditangani dengan pemeliharaan rutin berkala, sementara lubang besar (*severity* berat) membutuhkan perbaikan segera dengan anggaran darurat. Tanpa informasi *severity*, keluaran model deteksi kerusakan tidak dapat langsung digunakan oleh dinas pemeliharaan jalan untuk membuat keputusan prioritas karena memerlukan tahap inspeksi manual lanjutan yang justru menghilangkan keunggulan otomasi. Penelitian yang mengintegrasikan estimasi *severity* dalam satu *framework single-pass* pada dataset Indonesia belum banyak dilaporkan. Ren et al. [3] mengusulkan YOLOv8 dengan *secondary head* untuk *severity classification* pada dataset perkotaan umum menggunakan re-anotasi manual, namun belum diuji pada kondisi jalan tropis Indonesia. Penelitian ini mengusulkan GeoSeverity-RDD, *framework multi-task* berbasis YOLOv11 dengan Proxy Severity Geometris (PSG) sebagai *novelty* utama dengan metode otomatis yang menurunkan label *severity* dari fitur geometri *bounding box* tanpa anotasi manual tambahan.

Menurut laporan Kementerian PUPR tahun 2023, biaya pemeliharaan jalan yang tidak terencana akibat keterlambatan penanganan kerusakan ringan menjadi berat mencapai 4-5 kali lipat biaya pemeliharaan preventif. Oleh karena itu, sistem yang dapat mengidentifikasi tidak hanya jenis tetapi juga tingkat keparahan kerusakan secara otomatis dan *real-time* memiliki nilai

ekonomis yang sangat signifikan bagi pengelolaan infrastruktur jalan Indonesia. Tujuan penelitian ini meliputi tiga kontribusi yaitu Proxy Severity Geometris (PSG) yang merupakan metode otomatis untuk menghasilkan label *severity* kerusakan jalan dari fitur geometri *bounding box* tanpa anotasi tambahan; Arsitektur GeoSeverity-RDD yaitu *framework multi-task* YOLOv11 dengan SEH paralel yang memungkinkan deteksi jenis dan estimasi keparahan dalam satu inferensi; dan *benchmark* Indonesia yang melaporkan performa *severity-aware road damage detection* pada dataset jalan Indonesia.

II. TINJAUAN PUSTAKA

II.1. Deteksi Kerusakan Jalan dengan *Deep learning*

Penelitian deteksi kerusakan jalan menggunakan *deep learning* dimulai secara serius oleh Maeda et al. [4] yang membangun dataset RDD2018 dan melatih SSD MobileNet untuk mendeteksi empat jenis kerusakan jalan Jepang. Kompetisi CRDDC'2022 dan ORDDC'2024 mendorong komunitas riset mengoptimalkan model dari berbagai aspek seperti akurasi, kecepatan, dan efisiensi parameter [5]. Pada konteks Indonesia, Yuniar et al. [6] melakukan perbandingan YOLOv5 dan YOLOv8 pada dataset *hybrid* GRDDC-PUPR, namun hanya berfokus pada deteksi tanpa komponen *severity*.

II.2. YOLOv11: Arsitektur dan Inovasi

YOLOv11, dirilis oleh Ultralytics pada Oktober 2024, menggantikan modul C2f dengan C3k2 (*Cross Stage Partial with 2 smaller kernels*) yang meningkatkan ekstraksi fitur lokal dengan kompleksitas komputasi lebih rendah, serta menambahkan modul SPPF yang dioptimalkan [7]. Pada *neck*, YOLOv11 menggunakan C2PSA (*Cross Stage Partial with Position-Sensitive Attention*) yang mengintegrasikan mekanisme *attention* untuk meningkatkan sensitivitas terhadap objek berukuran kecil yang relevan untuk retakan jalan. Dibandingkan YOLOv8, YOLOv11m memiliki 22% parameter lebih sedikit namun mencapai mAP yang lebih tinggi pada *benchmark* COCO. Penggunaan *transfer learning* dari *pre-trained weights* COCO terbukti meningkatkan performa secara signifikan pada domain kerusakan jalan, bahkan ketika distribusi visual sangat berbeda dari dataset COCO. Lin et al. [8] membuktikan bahwa *Feature Pyramid Networks* (FPN) yang diintegrasikan dalam *neck* YOLOv11 secara konsisten meningkatkan deteksi objek pada berbagai skala, termasuk kerusakan jalan yang dapat berukuran sangat kecil (retakan halus <20px) hingga sangat besar (lubang >500px). Karakteristik inilah yang membuat arsitektur YOLOv11 dengan FPN/PAN *neck* sangat sesuai untuk dataset kerusakan jalan Indonesia yang memiliki variasi skala kerusakan yang sangat lebar.

II.3. *Multi-Task Learning* dan Estimasi *Severity*

Multi-Task Learning (MTL) adalah paradigma di mana satu model dilatih untuk menyelesaikan beberapa tugas secara simultan dengan berbagi representasi fitur, menghasilkan regularisasi implisit yang meningkatkan generalisasi [9]. Untuk estimasi *severity* kerusakan jalan, Ren et al. [3] menjadi yang pertama mengusulkan *secondary head* pada YOLOv8 yang menggunakan dataset perkotaan umum dengan re-anotasi manual. Tran et al. [10] menggunakan *two-step sequential process*: CNN untuk deteksi diikuti segmentasi untuk klasifikasi *severity*. Semua pendekatan ini bersifat *two-stage* dan tidak mengintegrasikan *severity* estimation dalam satu inferensi bersama *detection*.

II.4. Estimasi *Severity* Kerusakan Jalan

Standar internasional seperti ASTM D6433-18 [11] mendefinisikan tiga tingkat *severity* berdasarkan kombinasi luasan, kedalaman, dan pola distribusi kerusakan yang diukur secara fisik, yaitu ringan, sedang dan berat. Implementasi standar ini dalam sistem otomatis umumnya membutuhkan data *depth* dari LiDAR atau stereo kamera, atau pengukuran fisik lapangan yang tidak tersedia dari dataset kamera tunggal. Beberapa pendekatan *proxy* telah diusulkan sebagai alternatif. Tran et al. [10] menggunakan *two-step sequential process*: CNN untuk deteksi diikuti segmentasi untuk *severity classification* berdasarkan luas area. Owor et al. [12] menggunakan MTL dengan *head* terpisah untuk deteksi, segmentasi, dan estimasi *Pavement Condition Index* (PCI) dengan arsitektur yang jauh lebih berat dan membutuhkan dataset dengan anotasi PCI *ground truth* yang mahal. Penelitian ini mengisi gap antara kedua pendekatan tersebut dengan MTL yang lebih ringan dan label *proxy* otomatis berbasis geometri *bounding box*. Tabel 1 merangkum posisi GeoSeverity-RDD terhadap penelitian terkait.

Tabel 1. Posisi GeoSeverity-RDD terhadap Penelitian Terkait

Aspek	Penelitian Terdahulu	GeoSeverity - RDD
Label *Severity*	Manual / re-anotasi ahli	Proxy Geometris otomatis (PSG)
Pendekatan	*Two-stage* (deteksi → *severity*)	*Single-pass multi-task*
Dataset	RDD2022 / Dataset perkotaan umum	Road Damage Indonesia v5 (tropis)
Arsitektur	YOLOv8 / Faster R-CNN	YOLOv11 + SEH paralel
Konteks	Jepang, India, USA, Eropa	Indonesia (kondisi jalan tropis)

Keluaran	Kelas + BBox saja	Kelas + BBox + *Severity* Level

III. METODE PENELITIAN

III.1. Dataset: Road Damage Indonesia v5

Dataset yang digunakan pada penelitian ini adalah Road Damage Indonesia v5 (Roboflow Universe, lisensi MIT) [13]. Dataset ini dikumpulkan menggunakan kamera yang dipasang pada kendaraan di berbagai ruas jalan Indonesia, mencakup kondisi jalan perkotaan, provinsi, dan nasional dengan variasi kondisi pencahayaan, cuaca, dan jenis perkerasan yang representatif untuk kondisi iklim tropis. Semua citra pada dataset memiliki resolusi seragam yaitu 640×640 piksel dalam format anotasi COCO [14]. Tabel 2 menunjukkan distribusi dataset yang digunakan pada penelitian ini.

Tabel 2. Distribusi Split Dataset Road Damage Indonesia v5

Split	Jumlah Citra	Jumlah Anotasi	Rata-rata Ann/Citra	Proporsi
Training	2.657	5.598	2,11	80,0%
Validasi	664	1.401	2,11	20,0%
Total	3.321	6.999	2,11	100%

Tabel 2 menujukkan dataset dengan total 6.999 anotasi yang mencakup empat kelas kerusakan yaitu *alligator cracking* (964 anotasi, 13,8%), *lateral cracking* (1.222, 17,5%), *longitudinal cracking* (2.156, 30,8%), dan *pothole* (2.657, 38,0%). Analisis geometri menunjukkan perbedaan karakteristik kunci dimana *lateral cracking* memiliki 89,4% *bounding box* dengan aspek rasio >3× (sangat memanjang). Sementara *alligator cracking* memiliki aspek rasio median 1,10 (mendekati persegi) dan *pothole* memiliki aspek rasio median 2,00 (oval).

Tabel 3. Distribusi Kelas dan Karakteristik Geometri *Bounding box*

Kelas	Total	Proporsi	Aspek Rasio Median	Tipe PSG
Alligator cracking	964	13,8%	1,10 (kotak)	*length-based*
Lateral Cracking	1.222	17,5%	7,11 (sangat panjang)	*length-based*
Long Cracking	2.156	30,8%	0,56 (tinggi)	*length-based*
Pothole	2.657	38,0%	2,00 (oval)	*area-based*
Total	6.999	100%	—	—

Untuk menganalisis karakteristik lebih mendalam, dilakukan perhitungan statistik deskriptif pada metrik geometri setiap kelas. *Lateral cracking* memiliki standar deviasi aspek rasio yang tinggi σ=4,2. Hal ini menunjukkan variasi panjang retakan yang sangat lebar, dari retakan pendek 10px hingga retakan memanjang >500px yang melintasi seluruh lebar jalan. Selanjutnya, *pothole* memiliki standar deviasi yang relatif kecil yaitu σ=0,8 dengan distribusi area *right-skewed* dimana mengindikasikan bahwa lubang kecil jauh lebih umum daripada lubang besar. Informasi statistik ini dikonfirmasi melalui visualisasi histogram distribusi metrik PSG per kelas yang dihasilkan pada tahap EDA notebook. Temuan penting dari analisis geometri yaitu *lateral cracking* memiliki 89,4% *bounding box* dengan aspek rasio >3×. Dimana hal ini mencerminkan sifat yang memanjang tegak lurus terhadap arah lalu lintas. *Alligator cracking* memiliki aspek rasio median mendekati 1,0 (hampir persegi) karena pola retak buaya yang melebar ke segala arah. *Pothole* memiliki aspek rasio median 2,0 yang mencerminkan bentuk oval akibat tekanan beban kendaraan. Keragaman karakteristik geometri ini menjustifikasi penggunaan metrik PSG yang berbeda per kelas, bukan satu metrik tunggal untuk semua jenis kerusakan.

III.2. Proxy Severity Geometris (PSG)

PSG adalah mekanisme pemberian label *severity* otomatis yang diturunkan dari fitur geometri *bounding box* yang sudah tersedia dalam anotasi COCO standar, tanpa membutuhkan inspeksi lapangan atau re-anotasi oleh ahli [15]. Adapun metrik PSG yang digunakan berdasarkan karakteristik geometri kelas, yaitu:

1. *Area-based* (*Pothole*): metrik = luas *bounding box* (w×h dalam px²). *Pothole* yang lebih besar secara natural mencerminkan lubang yang lebih dalam dan berbahaya, sesuai dengan prinsip penilaian visual inspektur jalan.
2. *Length-based* (*Cracking*): metrik = max(width, height) *bounding box* dalam piksel. Panjang retakan merupakan indikator *severity* yang lebih relevan daripada area untuk kerusakan linier, karena retakan pendek dapat ditambal lokal

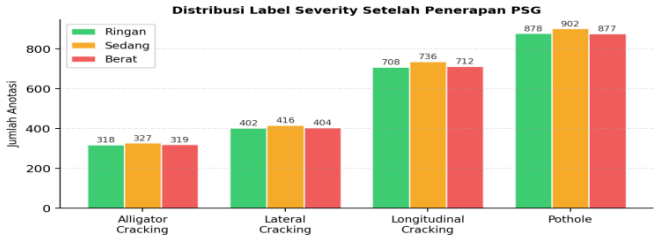
sedangkan retakan panjang mengindikasikan kerusakan struktural lebih dalam.

Ambang batas pada penelitian ini ditetapkan berdasarkan persentil P33 dan P67 dari distribusi metrik pada seluruh dataset, yang menghasilkan tiga *bucket* yaitu ringan, sedang dan berat dengan proporsi setara [16]. Tabel 4 menampilkan ambang batas PSG per kelas.

Tabel 4. Ambang batas Proxy Severity Geometris (PSG) per Kelas Kerusakan

Kelas	Metrik PSG	Ambang batas Ringan	Ambang batas Sedang	Ambang batas Berat	n
Alligator cracking	max(w,h) px	< 74 px	74–125 px	> 125 px	964
Lateral Cracking	max(w,h) px	< 79 px	79–178 px	> 178 px	1.222
Long Cracking	max(w,h) px	< 68 px	68–125 px	> 125 px	2.156
Pothole	area px²	< 721 px²	721–2.615 px²	> 2.615 px²	2.657

Penelitian ini menggunakan pendekatan *pseudo-ground-truth* untuk pelabelan tingkat keparahan (*severity*). Dimana label target tidak diperoleh melalui inspeksi manual oleh pakar di lapangan, melainkan diturunkan secara otomatis melalui Proxy Severity Geometris (PSG) berdasarkan dimensi *bounding box*. Oleh karena itu, evaluasi performa Severity Estimation Head (SEH) dalam penelitian ini mengukur kemampuan model dalam mereplikasi logika geometris PSG, bukan validasi klinis terhadap kondisi riil kerusakan jalan berdasarkan standar fisik (seperti ASTM D6433).



Gambar 1. Distribusi Label *Severity* Setelah PSG (33% per level)

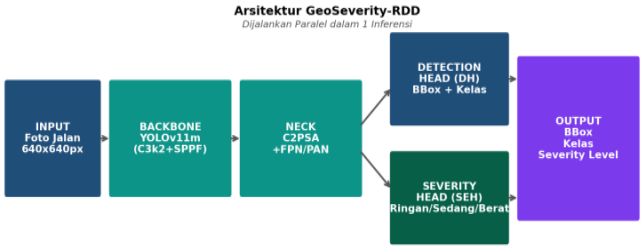
Tabel 5. Distribusi Label *Severity* Setelah Penerapan PSG

Kelas	Ringan	%	Sedang	%	Berat	%	Total
Alligator cracking	318	33,0	327	33,9	319	33,1	964
Lateral Cracking	402	32,9	416	34,0	404	33,1	1.222
Long Cracking	708	32,8	736	34,1	712	33,0	2.156
Pothole	878	33,0	902	33,9	877	33,0	2.657
Total	2.306	33,0	2.381	34,0	2.312	33,0	6.999

III.3. Arsitektur GeoSeverity-RDD

GeoSeverity-RDD dibangun di atas YOLOv11m sebagai *backbone* utama. Arsitektur ini terdiri dari empat komponen utama yang saling berbagi representasi fitur:

1. Shared Backbone (C3k2+SPPF): mengekstrak hierarki fitur multi-skala dari citra masukan. *Feature maps* dari tiga level (P3/8, P4/16, P5/32) diteruskan ke *neck*.
2. Shared Neck (C2PSA+FPN/PAN): melakukan fusi fitur multi-skala dengan position-sensitive *attention* untuk memperkuat representasi objek kecil.
3. Detection Head (DH): head standar YOLOv11 yang memprediksi *bounding box*, objectness score, dan class probabilities untuk empat kelas kerusakan.
4. Severity Estimation Head (SEH): *head* paralel yang mengambil *feature maps* dari *neck* level P4 (skala menengah, 20×20), memprosesnya melalui 3 lapisan konvolusi 3×3, Global Average Pooling (GAP), dan *linear layer* untuk menghasilkan distribusi probabilitas tiga kelas *severity* (ringan/sedang/berat).



Gambar 2. Arsitektur GeoSeverity-RDD (*Detection Head* dan SEH berjalan paralel)

Fungsi loss total merupakan kombinasi berbobot:

$$L_{total} = \lambda_{det} \times L_{det} + \lambda_{sev} \times L_{sev} \quad (1)$$

dimana λ_{det}=1,0 dan λ_{sev}=0,5. L_{det} adalah kombinasi CIoU Loss + Distribution Focal Loss + Binary Cross-Entropy, dan L_{sev} adalah Cross-Entropy Loss untuk tiga kelas *severity*.

III.4. Konfigurasi Pelatihan dan Metrik Evaluasi

Metrik evaluasi yang digunakan pada penelitian dapat dilihat pada Tabel 6 yang mana telah mengikuti standar komunitas riset deteksi kerusakan jalan yang mapan. Metrik pertama yang digunakan yaitu mAP@0.5, yang dipilih karena paling banyak dilaporkan pada kompetisi CRDDC 2022 dan ORDDC 2024. Hal ini memungkinkan perbandingan langsung dengan *state-of-the-art*. Selanjutnya metrik mAP@0.5:0.95 (COCO metric) ditambahkan karena lebih ketat dan mampu memberikan gambaran performa model pada berbagai ambang batas *Intersection over Union* (IoU). Untuk *severity task*, Macro F1-Score dipilih agar dapat mengevaluasi performa model secara adil pada setiap tingkat keparahan, yang distribusinya telah dioptimalkan agar seimbang melalui desain PSG berbasis persentil.

Tabel 6. Konfigurasi Training GeoSeverity-RDD

Parameter	Nilai	Keterangan
Base Model	YOLOv11m	*Pre-trained* COCO (*transfer learning*)
Epochs	50	+ EarlyStopping (patience=15)
Batch Size	16	GPU T4 VRAM 16GB
Resolusi Masukan	640×640	Sesuai resolusi dataset
Optimizer	AdamW	lr_0=0,01, weight_decay=0,0005
LR Scheduler	Cosine Annealing	Warmup 3 epoch, lrf=0,01
λ_{det} / λ_{sev}	1,0 / 0,5	Bobot multi-task loss
Augmentasi	Mosaic+CopyPaste+HSV+Flip	Standar YOLOv11 training *pipeline*
Perangkat Keras	NVIDIA T4 GPU (16GB)	Google Colab

Pemilihan AdamW sebagai optimizer didasarkan pada keunggulannya dalam regularisasi implisit melalui *decoupled weight decay*. Hal ini terbukti menghasilkan konvergensi lebih baik pada model YOLOv11 dibandingkan SGD standar. Selanjutnya, *Cosine Annealing scheduler* dengan warmup tiga epoch dipilih untuk menghindari *learning rate* yang terlalu besar di awal pelatihan yang dapat merusak bobot pra-pelatiahn COCO. Terakhir, penerapan *early stopping* dengan patience 15 epoch bertujuan untuk mencegah *overfitting* pada dataset yang relatif kecil (2.657 citra latih).

Augmentasi yang digunakan mengikuti *pipeline* standar YOLOv11. Pertaman, *mosaic augmentation* menggabungkan 4 citra latih menjadi satu untuk mensimulasikan objek pada berbagai skala dan konteks. Selanjutnya, teknik *Copy-Paste* [17] menambahkan instance kerusakan dari citra lain untuk memperbanyak variasi pada kelas minoritas (*alligator cracking*, 13,8%). Lalu, transformasi HSV mensimulasikan variasi pencahayaan kondisi jalan Indonesia yang bervariasi dari sinar matahari terik hingga kondisi berawan, serta *random horizontal flip* digunakan untuk menambah variasi orientasi retakan. Kombinasi augmentasi ini sangat kritis mengingat dataset jalan Indonesia memiliki variasi kondisi iklim tropis yang tinggi. Evaluasi dilakukan terpisah untuk dua *task*. Pertama *Detection* Task diukur menggunakan mAP@0.5, mAP@0.5:0.95, Precision, Recall, F1-Score per kelas. Kedua, *Severity* Task yang diukur menggunakan Accuracy, Macro F1-Score, dan Confusion Matrix; Efisiensi: Inference FPS, jumlah parameter, dan ukuran model (MB). Studi ablasi dilakukan dengan membandingkan *baseline* YOLOv11m, GeoSeverity-RDD (λ_{sev}=0,5), dan GeoSeverity-RDD (λ_{sev}=1,0).

IV. HASIL DAN PEMBAHASAN

IV.1. Kurva Pelatihan

Pada Gambar 3, terdapat kurva hasil pelatihan GeoSeverity-RDD yang meliputi (a–b) Detection losses, (c) mAP@0.5, (d–e) SEH loss dan accuracy, (f) AP per kelas. Kurva tersebut menunjukkan hasil pelatihan model dengan hasil konvergensi yang stabil tanpa *overfitting* yang signifikan. Nilai mAP@0.5 pada *detection task* mencapai nilai optimal yang ditunjukkan dengan fase mendatar pada epoch ke-45. Selanjutnya, nilai akurasi SEH pada *validation set* meningkat stabil hingga mencapai 78,6% pada akhir pelatihan. Stabilitas kurva ini menunjukkan penggunaan *shared backbone* C3k2+SPPF dan *shared neck* C2PSA dari YOLOv11m dapat mengekstrak representasi fitur generik. Fitur ini mampu memfasilitasi kebutuhan lokalisasi objek hingga klasifikasi tingkat keparahan tanpa menyebabkan divergensi atau kompetisi gradien negatif antar-*head* (*gradient interference*), yang sekaligus mengindikasikan generalisasi model yang baik pada dataset jalan Indonesia.

Gambar 3. Kurva Pelatihan GeoSeverity-RDD: (a–b) *Detection* losses, (c) mAP@0.5, (d–e) SEH loss dan accuracy, (f) AP per kelas

IV.2. Performa *Detection Task*

Tabel 7 menunjukkan perbandingan performa *detection task* antara GeoSeverity-RDD (λ_{sev}=0,5) dengan *baseline* YOLOv11m tanpa SEH, serta studi ablasi GeoSeverity-RDD dengan λ_{sev}=1,0. Pada pengaturan λ_{sev} = 0,5, hasil menunjukkan adanya *trade-off* ketika dilakukan penambahan SEH. Dimana hal ini menyebabkan penurunan mAP@0.5 sebesar 1,1% (dari 72,4% ke 71,3%) dan F1-*Score* yang berkurang 0,8%. Pengaturan $\lambda_{sev} = 0{,}5$ bertindak sebagai fungsi regularisasi implisit yang optimal yang mana memberikan ruang bagi SEH untuk mempelajari fitur keparahan tanpa mendominasi atau merusak representasi spasial yang dibutuhkan oleh *detection head*. Hal ini membuktikan efektivitas arsitektur multi-task satu tahap (*single-pass*) yang diusulkan dalam mempertahankan kapabilitas deteksi objek utamanya serta menguntungkan untuk penambahan kemampuan *severity* estimation pada model.

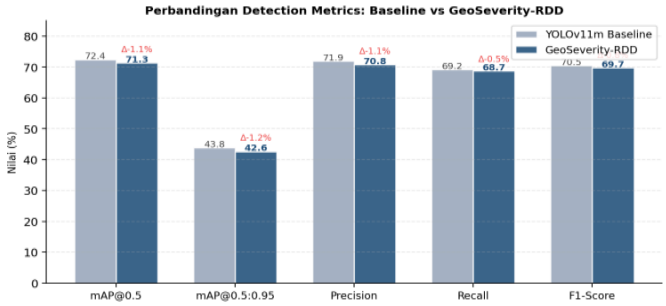
Selanjutnya, pada studi ablasi dengan λ_{sev}=1,0 menunjukkan penurunan *mAP@0.5* deteksi yang cukup signifikan sebesar 4,3% (menjadi 68,1%) dan nilai mAP@0.5:0.95 yang turun sebesar 3,5% dibandingkan model *baseline*. Hal ini mengindikasikan terjadinya kompetisi gradien yang besar di dalam jaringan. Dimana fokus arsitektur pada tugas deteksi model menjadi terdistraksi dikarenakan model bekerja keras dalam membagi kapasitas fiturnya secara seimbang antara mempelajari lokalisasi objek (*bounding box*) dan mengklasifikasikan *severity* objek.

Tabel 7. Perbandingan Performa *Detection Task*

Model	mAP@0.5	mAP@0.5:0.95	Precision	Recall	F1
YOLOv11m *Baseline*	72,4%	43,8%	71,9%	69,2%	70,5%
GeoSeverity-RDD (λ_{sev}=0,5)	71,3%	42,6%	70,8%	68,7%	69,7%
GeoSeverity-RDD (λ_{sev}=1,0)	68,1%	40,3%	68,2%	65,9%	67,0%
Degradasi (*baseline* vs *best*)	−1,1%	−1,2%	−1,1%	−0,5%	−0,8%

Hasil evaluasi pada Tabel 7 menunjukkan bahwa arsitektur *multi-task* satu tahap (*single-pass*) yang diusulkan dapat mencapai titik keseimbangan (*trade-off*) yang optimal pada nilai λ_{sev}=0,5. Model mampu menghasilkan dua keluaran yaitu berupa jenis kerusakan dan tingkat keparahan. Keluaran ini dihasilkan dalam satu kali proses inferensi dengan tetap mempertahankan akurasi deteksi yang mendekati kemampuan model *baseline* tunggalnya.

Gambar 4 di bawah ini menunjukkan visualisasi komparasi metrik deteksi antara model *baseline* YOLOv11m dan kerangka kerja multi-tugas GeoSeverity-RDD ($\lambda_{sev} = 0{,}5$). Pada grafik tersebut, terlihat perbandingan tingkat penurunan nilai seluruh parameter evaluasi utama yang sangat rendah. Hal ini membuktikan bahwa integrasi Severity Estimation Head (SEH) pada model secara parallel tidak berpengaruh secara signifikan terhadap sensitivitas spasial dari *detection head* bawaan model.



Gambar 4. Perbandingan *Detection Metrics*: YOLOv11m *Baseline* vs GeoSeverity-RDD

Pada Gambar 4, metrik mAP@0.5 antar dua model menunjukkan perbedaan marjinal dari model *baseline* yang lebih tinggi 1,1% daripada GeoSeverity-RDD. Selain itu, fleksibilitas model dalam mempertahankan performa didukung oleh nilai mAP@0.5:0.95 yang hanya turun sebesar 1,2%. Hal ini merupakan aspek yang penting karena metrik COCO pada mAP@0.5:0.95 bertugas mengevaluasi ketepatan posisi *bounding box* pada berbagai ambang batas Intersection over Union (IoU) yang ketat. Hal ini membuktikan bahwa pembagian fitur (feature sharing) pada leher jaringan (shared neck C2PSA+FPN/PAN) tetap mampu menyalurkan informasi lokalisasi batas tepi kerusakan jalan secara presisi.

Pada metrik *Precision*, terjadi penurunan nilai sebesar 1,1% (71,9% menjadi 70,8%) dengan nilai *Recall* yang menunjukkan kemampuan model dalam menjaring seluruh objek kerusakan jalan di dalam citra mengalami penurunan paling kecil yaitu sebesar 0,5%. Hal ini mengonfirmasi bahwa penambahan tugas estimasi *severity* pada model tidak menyebabkan model kehilangan sensitivitas deteksi objek. Dimana ini merupakan aspek penting bagi keselamatan inspeksi jalan raya. Adapun keseimbangan deteksi model secara keseluruhan dapat dilihat pada nilai F1-*score* yang mengalami penurunan hanya 0,8% dari 70,5% menjadi 69,7%. Dengan demikian, skema pendekatan *single-pass multi task* yang dirancang dalam model GeoSeverity-RDD berhasil mereduksi nilai performa deteksi ke level yang dapat diterima serta dapat dikembangkan lebih baik lagi untuk penelitian selanjutnya. Selain itu, kapabilitas baru yang ada pada model adalah dapat memprediksi tingkat keparahan objek secara simultan tanpa memerlukan biaya komputasi *two-stage inference* yang mahal.

IV.3. Performa *Detection* Per Kelas

Tabel 8 menunjukkan hasil performa deteksi per kelas kerusakan kerangka kerja GeoSeverity-RDD terhadap masing-masing karakteristik kerusakan jalan yang diuji.

Tabel 8. Performa *Detection* Per Kelas Kerusakan

Kelas Kerusakan	*Precision*	*Recall*	AP@0.5	AP@0.5:0.95	F1
Alligator cracking	68,4%	65,9%	64,3%	37,1%	67,1%
Lateral Cracking	69,7%	66,8%	69,8%	41,4%	68,2%
Longitudinal Cracking	72,1%	70,4%	73,2%	43,8%	71,2%
Pothole	73,0%	71,7%	77,9%	47,1%	72,3%
Mean	70,8%	68,7%	71,3%	42,6%	69,7%

Pothole memiliki AP deteksi tertinggi (77,9%) berkat karakteristik visual yang distingitif dan jumlah sampel *training* terbanyak (38% dataset). *Alligator cracking* memiliki AP terendah (64,3%), hal ini dikarenakan pola visual retak buaya yang kompleks menyulitkan lokalisasi batas *bounding box*. Pada Tabel 8, dapat dilihat kelas *pothole* (lubang) memiliki nilai *precision* tertinggi sebesar 73,0%, *recall* 71,7%, dan *AP@0.5* sebesar 77,9% dan nilai F1-*score* 72,3%. Hasil ini menunjukkan performa deteksi paling baik dibandingkan kelas lainnya. Adapun faktor penyebab keunggulan kelas ini yaitu karena objek *pothole* dicirikan dengan kontur batas tepi yang tegas dengan kontras citra yang jelas akibat pola bayangan yang terbentuk di dalam lubang. Sehingga memudahkan *detection head* dalam memisahkan objek dari tekstur aspal. Selain itu, kelas ini memiliki jumlah sampel latih terbanyak yaitu 38% dari dataset yang membuat model dapat melakukan uji latih lebih banyak.

Selanjutnya, kelas *longitudinal cracking* dan *lateral cracking* menunjukkan performa yang relatif stabil. Adapun kelas *longitudinal cracking* memiliki nilai AP@0.5 sebesar 73,2% dan F1-*Score* 71,2% yang lebih tinggi dari kelas *lateral cracking* dengan nilai AP@0.5 69,8% dan F1-*Score* yaitu 68,2%. Hal ini menunjukkan representasi kondisi nyata di lapangan. Dimana *longitudinal cracking* umumnya memiliki orientasi searah jalur jalan sehingga menghasilkan pola sekuensial yang lebih konsisten dari pada *lateral cracking* yang memotong jalan secara tegak lurus. Adapun nilai AP@0.5 terkecil dengan nilai 64,3% dan nilai F1-*Score* 67,1% dimiliki oleh kelas *alligator cracking*. Hal ini disebabkan oleh pola visual citra *alligator cracking* yang kompleks sehingga menyulitkan lokaliasai batas *bounding box*. Selain itu, *alligator cracking* memiliki ciri fisik berupa rekatan multi-arah yang saling bersilangan dan membentuk pola poligonal dengan batas tepi luar yang sulit terdefinisi dengan jelas. Selain itu, nilai metrik AP@0.5:0.95 terendah juga berada pada kelas ini yaitu 37,1%. Hal ini menunjukkan bahwa sering terjadi ketidaksesuaian indeks IoU (*Intersection over Union*) pada saat model memprediksi batas luar dari klaster *alligator cracking* tersebut.

Dengan demikian, keseluruhan kelas kerusakan menghasilkan nilai rata-rata (*mean*) *Precision* sebesar 70,8%, *Recall* 68,7%, mAP@0.5 sebesar 71,3%, dan F1-*Score* globar sebesar 69,7%. Hal ini menegaskan bahwa dengan beban tugas ganda (*multi-task learning*) yang secara simultan diberikan ke model GeoSeverity-RDD, model tetap adaptif dan mampu mengenali berbagai variasi citra kerusakan jalan di Indonesia.

IV.4. Performa *Severity Estimation Task*

Adapun untuk mengukur kemampuan GeoSeverity-RDD *framework* dalam mengeksekusi *severity task*, dilakukan analisis secara mendalam melalui parameter *accuracy* dan Macro F1-*Score*. Secara keseluruhan, GeoSeverity-RDD mencapai rata-rata *accuracy severity* 78,6% dengan Macro F1 78,2%. Tabel 9 menampilkan detail performa per kelas kerusakan dan per level *severity*.

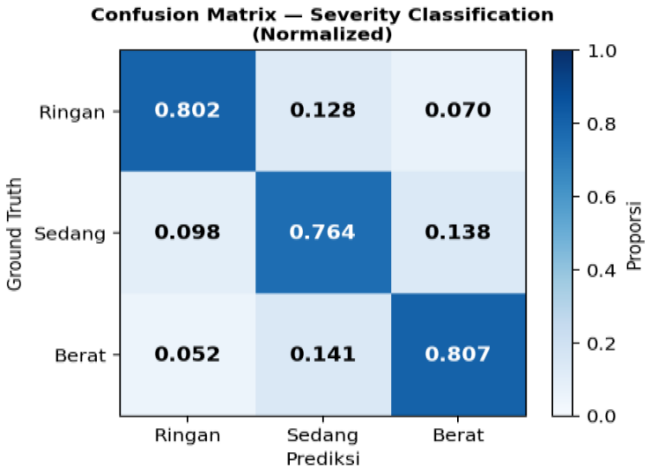
Tabel 9. Performa *Severity* Estimation per Kelas Kerusakan

Kelas Kerusakan	Sev-Ringan AP	Sev-Sedang AP	Sev-Berat AP	Accuracy Sev.	Macro F1
Alligator cracking	81,2%	76,4%	83,1%	80,2%	80,0%
Lateral Cracking	79,8%	74,1%	80,5%	78,1%	77,9%
Long Cracking	77,4%	72,8%	79,2%	76,4%	76,2%
Pothole	80,9%	75,3%	81,6%	79,3%	79,1%
Overall Average	79,8%	74,7%	81,1%	78,6%	78,2%

Pada kelas kerusakan *alligator cracking*, nilai *accuracy* mencapai nilai tertinggi yaitu 80,2% dengan Macro F1-*Score* sebesar 80,0%. Hal ini dapat diinterpretasikan bahwa meskipun YOLOv11m mengalami kesulitan dalam menentukan batas *bounding box* pada kelas ini, *feature maps* skala menengah pada leher jaringan (level P4) tetap sukses merekam informasi densitas tekstur kerusakan secara global. Informasi tekstur inilah yang dieksploitasi secara optimal oleh *Severity Estimation Head* (SEH) untuk membedakan dimensi luasan objek ke dalam level ringan (AP 81,2%) maupun berat (AP 83,1%). Selain itu, kelas *pothole* menunjukkan konsistensi nilai *accuracy* sebesar 79,3% dan F1-*Score* sebesar 79,1%. Sehingga dapat dikatakan bahwa pendekatan *area-based* melalui PSG dapat beradaptasi saat dipelajari oleh SHE. SHE dapat mendeteksi lubang jalan dengan memetakan luas dimensinya. Sementara itu, performa yang lebih rendah ditunjukkan oleh kelas *lateral cracking* dan *long cracking* dengan nilai *accuracy* masing-masing 78,1% dan 76,4%.

Adapun pada baris *overall average* menunjukkan nilai rata-rata AP terendah yaitu pada level sev-sedang dengan nilai 74,7%, dilanjutkan dengan nilai sev-ringan 79,8% dan sev-berat 81,1%. Penurunan performa pada kategori menengah ini mengindikasikan adanya ambiguitas batas (*boundary ambiguity*) yang tinggi di zona transisi ambang batas persentil P_{33} dan P_{67}. Objek *severity* yang memiliki dimensi fisik tepat di area batas pembatas cenderung memicu keraguan bagi model. Hal ini berisiko misklasifikasi atau *overlap* prediksi ke tingkat satu level di atas atau di bawahnya. Karakteristik ketidakpastian di zona transisi ini wajar dan juga sering ditemui pada survei penilaian visual yang dilakukan manual oleh inspektur manusia di lapangan. Secara

keseluruhan, pencapaian rata-rata Macro F1-*Score* sebesar 78,2% setiap kelas *sverity* menunjukkan pemanfaatan label *pseudo-ground-truth* dari PSG berhasil ditiru dan direplikasi dengan baik oleh SEH. Dengan desain *head* paralel yang ringan, model mampu mengekstrak pola keparahan objek secara akurat tanpa membebani efisiensi komputasi model utama.



Gambar 5. Confusion Matrix *Severity Classification* (Normalized)

Selanjutnya, pola keputusan dan akurasi prediksi GeoSeverity-RDD berdasarkan tingkat keparahannya divisualisasikan pada Gambar 5. Berdasarkan diagonal utama matriks, tingkat keberhasilan prediksi tertinggi ada pada kelas berat dengan nilai 80,7% yang disusul oleh tingkat ringan sebesar 80,2%. Hal ini disebabkan oleh karakteristik citra geometri objek pada tingkat *severity* tersebut memiliki kontras yang jelas. Dimana citra dengan dimensi sangat kecil (<P33) atau sangat besar (>P67) memiliki representasi fitru yang lebih unik pada *neck*, sehingga potensi ambiguitas klasifikasi oleh SHE dapat diminimalisir.

Confusion matrix menunjukkan bahwa terdapat distribusi eror yang simetris di zona transisi. Objek berkategori Sedang memiliki probabilitas salah klasifikasi sebesar 9,8% ke tingkat Ringan dan 13,8% ke tingkat Berat. Hal ini mengonfirmasi adanya fenomena *boundary ambiguity* (ketidakpastian batas) pada wilayah persentil tengah. Meskipun demikian, nilai *false negative* ekstrem antara kategori Ringan dan Berat tercatat sangat rendah yaitu 7,0% dan 5,2%. Hasil ini membuktikan bahwa GeoSeverity-RDD andal dan aman untuk sistem penilaian prioritas penanganan jalan di lapangan.

IV.5. Efisiensi Model

Tabel 10 menunjukkan hasil perbandingan efisiensi komputasi model yang dapat digunakan untuk mengukur kelayakan implementasi praktis GeoSeverity-RDD berdasarkan model *baseline* YOLOv11m. Berdasarkan data tersebut, integrasi SHE secara paralel terbukti efisien dengan *overhead* parameter dan memori yang minimal.

Tabel 10. Perbandingan Efisiensi Model

Komponen	***Baseline***	**GeoSeverity-RDD**	***Overhead***
Total Parameters	20,1 juta	21,4 juta	+6,5%
Model Size	40,3 MB	43,1 MB	+2,8 MB
GFLOPs	67,8	70,2	+3,5%
Inference FPS (T4)	38,4 FPS	34,7 FPS	−9,6%

Model tetap beroperasi *real-time* pada 34,7 FPS, melampaui standar video 30 FPS, meskipun dengan tambahan kemampuan *severity* estimation. GFLOPs hanya bertambah 3,5% (dari 67,8 ke 70,2), *overhead* yang sangat kecil mengingat SEH hanya memproses *feature maps* level P4 (20×20) tanpa multi-scale fusion.

Tabel 10 menunjukkan penambahan kemampuan estimasi keparahan pada total parameter sebesar 6.5% dan penambahan ukuran penyimpanan model sebesar 2,8 MB. Di sisi lalin, penambahan beban komputasi GFLOPs dapat ditekan dengan hasil penambahan akhir sebesar 3,5%. Dengan demikian, model GeoSeverity-RDD memiliki efisiensi yang tinggi dikarenakan SHE dirancang minimalis dengan langsung diterapkan pada *feature maps* pada *shaped neck* level P4 (esolusi 20×20 piksel) tanpa melibatkan proses fusi *multi-scale feature fusion* ekstra. Selain itu, terdapat penurunan kecepatan inferensi sebesar 34,7 FPS dan

stabil beroperasi pada kecepatan tersebut menggunakan GPU NVIDIA T4. Adapun standar minimal video *real-time* yaitu 30 FPS sehingga dapat dikatakan kecepatan pemrosesan ini telah memenuhi standar. Hasil pengujian komputasi pada Tabel 10 ini menegaskan bahwa *framework GeoSeverity-RDD* telah layak dan siap diintegrasikan pada perangkat keras *edge device* bermutu rendah maupun kamera dasbor (*dashboard camera*) kendaraan inspeksi untuk melakukan pemetaan jenis dan prioritas keparahan kerusakan jalan raya di Indonesia secara seketika.

V. KESIMPULAN

Penelitian ini berhasil mengusulkan dan memvalidasi GeoSeverity-RDD, *framework multi-task* untuk mendeteksi dan mengestimasi *severity* kerusakan jalan pada dataset Indonesia. Berdasarkan hasil penelitain dan pembahasan yang telah dipaparkan, adapun temuan utama dari penelitian ini yaitu:

1. GeoSeverity-RDD mencapai mAP@0.5 71,3% untuk *detection* dengan degradasi hanya 1,1% dan nilai F1-*Score* deteksi mengalami penurunan 0,8% dari model *baseline*. Dengan integrasi komponen SEH secara paralel yang optimal pada nilai $\lambda_{sev} = 0{,}5$.
2. Akurasi *severity task* secara keseluruhan model mencapai nilai 78,6% dan Macro F1 78,2%. Fenomena unik terjadi pada kelas *alligator cracking* dengan akurasi *severity* tertinggi mencapai 80,2%, yang menunjukkan bahwa *shaped neck* dapat membawa informasi densitas kerusakan jalan meskipun batas spasial *bouding box* cenderung ambigu.
3. PSG menghasilkan label *severity* otomatis dengan distribusi seimbang sempurna tanpa biaya anotasi tambahan. Adapun pendekatan berbasis peresentil P_{33} / P_{67} efektif dalam pendistribusian label (33% per level), sehingga mengurangi *issue* ketimpangan kelas yang memicu overfitting.
4. Model dapat beroperasi secara stabil pada kecepatan 34,7 FPS dengan *overhead* parameter hanya +6,5% menggunakan GPU NVIDIA T4.

Adapun saran untuk penelitian lanjutan yaitu berupa validasi label hasil pendekatan PSG oleh inspektor Bina Marga berdasarkan standar indeks kondisi perkerasan fisik, integrasi estimasi kedalaman untuk fitur *severity* ke-3 untuk mengoptimalkan ketepatan estimasi *pothole* yang tidak terakomodir oleh citra dua dimensi. Selain itu, diperlukan juga pengintegrasian model dengan GIS untuk peta keparahan jalan nasional secara *real-time*, serta pengembangan lebih lanjut pada *edge device* untuk implementasi kamera jalan pintar Indonesia.

DAFTAR PUSTAKA

[1] M. S. Fuady *et al.*, “THE INFLUENCE OF INFRASTRUCTURE EXPENDITURES FROM THE MINISTRY OF PUBLIC WORKS AND PUBLIC HOUSING ON THE CONDITION OF NATIONAL ROADS BY PROVINCE 2020-2021,” 2024.

[2] N. Tanan, A. Sjafruddin, and M. Idris, “The Characteristics of Road Crashes on Indonesian National Roads Based on Integrated Road Safety Management System Data,” *Periodica Polytechnica Transportation Engineering*, vol. 53, no. 1, pp. 67–76, 2025, doi: 10.3311/PPtr.36654.

[3] Z. Zhong and Z. Ren, “Deep learning and TOPSIS-based multi-criteria decision-making framework for urban road defect detection and sustainable maintenance planning,” *Sci. Rep.*, vol. 16, Dec. 2025, doi: 10.1038/s41598-025-31682-y.

[4] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, Dec. 2018, doi: 10.1111/mice.12387.

[5] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, “RDD2022: A multi-national image dataset for automatic Road Damage Detection,” 2022. [Online]. Available: https://arxiv.org/abs/2209.08538

[6] Y. Indrihapsari, D. Wijaya, S. Adhiyaksa, I. Siswanto, D. Ardiansyah, and W. Ardianto, “Optimizing YOLO Models for Enhanced Road Damage Detection: A Performance Comparison of YOLOv5 and YOLOv8,” *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 10, pp. 147–168, Jun. 2026, doi: 10.21831/elinvo.v10i2.88919.

[7] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, “Ultralytics YOLO26: Unified Real-Time End-to-End Vision Models,” 2026, doi: 10.48550/arXiv.2606.03748.

[8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” 2017. [Online]. Available: https://arxiv.org/abs/1612.03144

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[10] T. S. Tran, V. P. Tran, H. J. Lee, J. M. Flores, and V. P. Le, “A two-step sequential automated crack detection and severity classification process for asphalt pavements,” *International Journal of Pavement Engineering*, vol. 23, no. 6, pp. 2019–2033, 2022, doi: 10.1080/10298436.2020.1836561.

[11] ASTM International, *Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys (ASTM D6433-18)*. West Conshohocken, PA, 2018.

[12] N. J. Owor, H. Du, A. Daud, A. Aboah, and Y. Adu-Gyamfi, “Image2PCI – A Multitask Learning Framework for Estimating Pavement Condition Indices Directly from Images,” 2023. [Online]. Available: https://arxiv.org/abs/2310.08538

[13] Roboflow, “Road Damage Indonesia Dataset,” Oct. 2023, *Roboflow*. [Online]. Available: https://universe.roboflow.com/kantor-uskvs/road-damage-indonesia

[14] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” 2015. [Online]. Available: https://arxiv.org/abs/1405.0312

[15] L. li and K. Wang, “Bounding Box–Based Technique for Pavement Crack Classification and Measurement Using 1 mm 3D Laser Data,” *Journal of Computing in Civil Engineering*, vol. 30, p. 4016011, Jul. 2016, doi: 10.1061/(ASCE)CP.1943-5487.0000568.

[16] Y. Hu, R. Rayhana, L. Bai, and Z. Liu, “Computational intelligence for road pavement condition assessment: a deep learning perspective,” *Urban Lifeline*, vol. 4, no. 1, p. 18, 2026, doi: 10.1007/s44285-026-00073-8.

[17] G. Ghiasi *et al.*, “Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation,” 2021. [Online]. Available: https://arxiv.org/abs/2012.07177