

ANALISIS KINERJA MODEL DETEKSI OBJEK YOLO, SSD, DAN FASTER R-CNN PADA CITRA PENGLIHATAN MALAM UNTUK PENGENALAN TINDAK KEJAHATAN

MS Hendriyawan Achmad^{*1}, Arif Pramudwiatmoko², Muhammad Satrio Gumilang³, Bahrul Al Karim⁴,
Hadi Wiyono⁵

^{1,2,3,4,5}Universitas Teknologi Yogyakarta, Yogyakarta

Email: ¹hendriyawanachmad@uty.ac.id, ²arif.pramudwiatmoko@uty.ac.id,

³muhammad.5200411155@student.uty.ac.id; ⁴bahrul.5200411396@student.uty.ac.id,

⁵hadi.5200411183@student.uty.ac.id

*Penulis Korespondensi

(Naskah masuk: 29 Desember 2023, diterima untuk diterbitkan: 10 Februari 2025)

Abstrak

Kejahatan *klitih* di wilayah Yogyakarta telah menimbulkan kekhawatiran serius bagi pemerintah dan masyarakat, sehingga mengancam keamanan dan kenyamanan publik. Dalam upaya penanganan permasalahan ini, penelitian ini mengajukan solusi implementasi teknologi keamanan yang berfokus pada kamera *night vision* dan *machine learning* guna mendeteksi kejahatan *klitih* dengan efektif, khususnya pada rentang waktu malam. Data yang dikumpulkan untuk penelitian ini terdiri dari 1006 gambar yang direkam dari aksi kejahatan *klitih*. Proses pengolahan data melibatkan beberapa tahap, dimulai dengan preprocessing di mana seluruh gambar diubah ukurannya menjadi 640x640 piksel. Selanjutnya, dilakukan augmentasi data untuk meningkatkan variasi dan ketangguhan model, berupa rotasi sebesar 90°, crop dengan variasi zoom dari 0% hingga 20%, penambahan noise dilakukan hingga 5%. Hasil penelitian menunjukkan bahwa model YOLOv6 memberikan kinerja terbaik dalam mendeteksi label senjata, dengan akurasi sebesar 0,9 dan F1-score mencapai 0,91. Sementara itu, dalam mengenali kejahatan fisik, YOLOv6 juga menunjukkan performa unggul dengan akurasi 0,63 dan F1-score 0,73. Model *Faster R-CNN* dan *SSD* juga memberikan hasil yang baik, namun YOLOv6 mempertahankan dominasi dalam deteksi kejahatan *klitih* berdasarkan akurasi dan evaluasi metrik lainnya. Pemanfaatan teknologi pendeteksian *klitih* di masa depan dapat memberikan kontribusi positif dalam menciptakan lingkungan yang lebih aman dan nyaman bagi seluruh masyarakat.

Kata kunci: *klitih*, *Deep Learning*, *YOLOv6*, *SSD*, *Faster R-CNN*

ANALYSIS OF THE PERFORMANCE OF YOLO, SSD, AND FASTER R-CNN OBJECT DETECTION MODELS APPLIED TO NIGHT VISION IMAGERY FOR CRIMINAL ACTIVITY IDENTIFICATION

Abstract

The prevalence of *klitih* criminal activity within the Yogyakarta region has engendered significant apprehension among governmental authorities and the public alike, thereby posing a substantial risk to community safety and well-being. In an effort to address this pressing concern, the present research advocates for the deployment of advanced security technologies, specifically emphasizing the utilization of night vision surveillance cameras in conjunction with machine learning algorithms to proficiently identify *klitih*-related offenses, particularly during nocturnal hours. The dataset utilized for this investigation comprises 1,006 photographic images obtained from various *klitih* crime occurrences. The data processing procedures encompassed multiple phases, commencing with preprocessing wherein all images were standardized to dimensions of 640x640 pixels. Subsequently, data augmentation techniques were employed to bolster the diversity and resilience of the model, incorporating transformations such as 90° rotations, cropping with zoom variations ranging from 0% to 20%, and the introduction of noise levels of up to 5%. The findings of this study indicate that the YOLOv6 model exhibited the most favorable performance in the detection of weapon classifications, achieving an accuracy rate of 0.9 and an F1-score of 0.91. Furthermore, in the context of identifying physical crimes, YOLOv6 similarly showcased outstanding efficacy, attaining an accuracy of 0.63 and an F1-score of 0.73. Although the *Faster R-CNN* and *SSD* models yielded commendable results, YOLOv6 sustained its preeminence in the realm of *klitih* crime detection, as evidenced by its superior accuracy and other evaluative metrics. The prospective implementation of *klitih* detection technology holds the potential to make a constructive impact in fostering a safer and more secure environment for the entire community.

Keywords: *klitih*, *Deep Learning*, *YOLOv6*, *SSD*, *Faster R-CNN*

1. PENDAHULUAN

Klitih adalah tindakan kejahatan yang dilakukan oleh pelaku remaja dibawah umur di wilayah Yogyakarta dengan melukai atau bahkan membunuh korban secara acak (Sarmini, Kurniyatuti and Sukartiningsih, 2018). Fenomena kejahatan malam *klitih* di Yogyakarta semakin dikenal luas melalui media sosial dengan adanya kejadian baru-baru ini. Beberapa pendekatan sudah dilakukan oleh pemerintah Provinsi Daerah Istimewa Yogyakarta untuk menekan angka kejadian kejahatan *klitih*, antara lain dengan pendekatan budaya masyarakat melalui pendidikan kepada anak dan orang tua (Sukirno, 2018). Kejahatan *klitih* ini sudah bukan lagi kasus kenakalan remaja namun sudah masuk kategori tindakan kriminal yang sudah meresahkan masyarakat dan dapat mengganggu visi Provinsi Yogyakarta sebagai kota Pelajar dan Pariwisata (Wijanarko and Ginting, 2021). Meskipun pemerintah Provinsi Yogyakarta sudah melakukan edukasi masyarakat namun tindak kejahatan *klitih* di malam hari masih saja terjadi karena para pelaku merasa tindakan mereka tidak terdeteksi.

Sehingga diperlukan teknologi sistem keamanan yang mampu bekerja sepanjang malam dan tersebar yang mampu memberikan peringatan kepada aparat berwajib untuk melakukan penindakan secara cepat. Teknologi yang ada umumnya hanya mengenakan pengawasan oleh manusia melalui kamera litar tertutup (CCTV) untuk pemantauan keamanan dan tidak mampu menjalankan fungsi deteksi secara real-time (Socha and Kogut, 2020).

Dalam menghadapi permasalahan kejahatan *klitih* yang sulit diatasi oleh sistem keamanan swadaya masyarakat, perlu dilakukan pendeteksian kejahatan tersebut secara efektif. Penggunaan kamera *night vision* yang menggunakan infrared dapat menjadi solusi efektif untuk merekam keadaan gelap, memungkinkan pemantauan yang lebih akurat bahkan pada malam hari (J and Thinakaran, 2023). Selain itu, dengan memanfaatkan pemodelan *machine learning*, suatu sistem deteksi dapat dibangun untuk bekerja sepanjang malam. Dilengkapi dengan kemampuan pembelajaran mesin yang memungkinkan identifikasi pola perilaku mencurigakan (Patel and Upla, 2020). Penggabungan kamera *night vision* dan teknologi *machine learning* tidak hanya meningkatkan kemampuan deteksi di kondisi gelap, tetapi juga memberikan respons yang lebih cepat dan presisi dalam mengidentifikasi potensi kejahatan (Singh *et al.*, 2021). Dengan demikian, implementasi solusi ini dapat membantu meningkatkan keamanan di lingkungan *klitih*, memberikan dukungan yang lebih efektif bagi upaya pencegahan dan penanggulangan kejahatan. Pemilihan model deteksi yang tepat akan menjadi kunci keberhasilan dalam menangani permasalahan keamanan yang semakin kompleks.

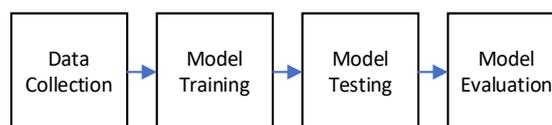
Pemodelan deteksi kejahatan *klitih*, dilakukan perbandingan dengan beberapa model deteksi objek.

Yaitu, model YOLO (*You Only Look Once*), *Faster RCNN* (*Faster Region-based Convolutional Neural Network*), dan SSD (*Single Shot Multibox Detector*). Ketiga model ini umum digunakan dalam penelitian komputer vision dan telah terbukti efektif dalam mendeteksi objek dalam konteks pengujian (Wang, Wang and Liu, 2020), sehingga dapat diimplementasikan untuk mendeteksi tindak kejahatan *klitih*. Evaluasi kinerja akan dilakukan dengan membandingkan tingkat akurasi, kecepatan deteksi, dan kemampuan dalam mengenali situasi kejahatan malam *klitih*. Dengan adanya perbandingan ini, diharapkan dapat diidentifikasi model terbaik yang mampu memberikan solusi efektif dalam mengatasi permasalahan kejahatan *klitih* di wilayah Yogyakarta.

2. METODE PENELITIAN

2.1 Langkah Penelitian

Penelitian ini, diawali proses mengumpulkan dataset gambar *klitih* yang secara khusus peneliti hasilkan melalui reka adegan pada malam hari dengan memperhatikan berbagai situasi dan kondisi yang mungkin dihadapi. Langkah selanjutnya yaitu melakukan pelatihan model deteksi objek menggunakan tiga arsitektur yang berbeda, yaitu YOLO (*You Only Look Once*), *Faster R-CNN* (*Region-based Convolutional Neural Network*), dan SSD (*Single Shot Multibox Detector*).



Gambar 1. Langkah penelitian

Penelitian ini dimulai dengan tahap pengumpulan data, di mana aksi kejahatan *klitih* direkam dan dihasilkan 1006 gambar yang digunakan sebagai dataset. Selanjutnya, dilakukan tahap pelatihan model menggunakan tiga pendekatan yang berbeda untuk mengembangkan model deteksi kejahatan. Setelah proses pelatihan selesai, model-model tersebut kemudian diuji dalam tahap testing menggunakan dataset yang tidak pernah dilihat sebelumnya. Evaluasi dilakukan dengan menganalisis metrik kinerja seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai kemampuan deteksi masing-masing model. Proses pada Gambar 1 memberikan gambaran menyeluruh tentang efektivitas dan keunggulan relatif dari setiap model dalam mendeteksi kejahatan *klitih*.

2.2 Test-bed Penelitian

Penelitian ini melibatkan penggunaan alat-alat dan bahan (*test-bed*) baik untuk pengumpulan dataset, pelatihan, maupun untuk pengujian kinerja model. Gambar 2 menunjukkan *test-bed* alat dan bahan pengujian.



Gambar 2. Test-bed penelitian

2.3 Pengumpulan Dataset

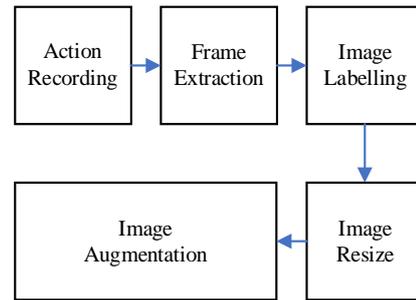
Dalam upaya meningkatkan keamanan masyarakat, penelitian deteksi objek kejahatan *klitih* menjadi suatu aspek penting yang membutuhkan pendekatan teknologi canggih. Pengumpulan data yang akurat dan representatif menjadi landasan utama dalam merancang model deteksi objek yang efektif. Terdapat dua jenis kelas dataset, yaitu: senjata tajam (*weapon*) dan kekerasan fisik (*physical abuse*).



Gambar 3. Contoh aksi yang dijadikan dataset untuk pelatihan model YOLOv6, Faster R-CNN, dan SSD.

Gambar 3 menunjukkan contoh beberapa aksi yang identik dengan tindakan kejahatan yang diambil melalui kamera penglihatan malam. Berdasarkan gambar-gambar tersebut selanjutnya akan

dikumpulkan sebagai dataset latih model. Alur pengumpulan dataset ditunjukkan oleh Gambar 4.



Gambar 4. Langkah pengumpulan dataset

Terlihat pada Gambar 4 proses pengumpulan data, perekaman aksi menjadi langkah pertama untuk mengabadikan kejadian-kejadian terkait *klitih*. Penggunaan berbagai perangkat perekam video yang tersedia membantu mendokumentasikan aksi-aksi yang relevan untuk analisis lebih lanjut (Zhang, Wieland and Basran, 2022). Selanjutnya, penggunaan metode framing extraction melalui perangkat lunak VLC (*Video LAN Client*) memungkinkan pengambilan frame gambar secara teratur dari rekaman video. Hal ini membentuk dasar dari dataset yang akan digunakan dalam pelatihan model deteksi objek.

Setelah pengumpulan data awal, langkah selanjutnya melibatkan proses labeling menggunakan fasilitas *Roboflow*. Proses ini memiliki peran krusial dalam mengidentifikasi dan memberikan anotasi objek, termasuk kejahatan *klitih*, sehingga dataset menjadi lebih informatif untuk pembelajaran model. Dalam rangka mempersiapkan data untuk pelatihan, dilakukan pula proses resize dengan mengubah dimensi gambar menjadi 640x640 piksel, sesuai dengan format input yang diharapkan oleh model deteksi objek yang akan diterapkan.



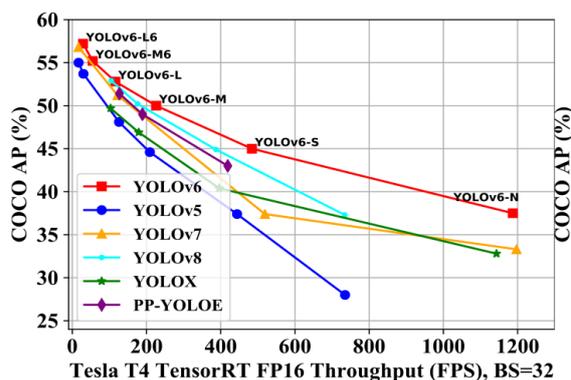
Gambar 5. Hasil augmentasi dataset

Proses terakhir adalah *image augmentation* seperti yang terlihat pada Gambar 5. Augmentasi data dilakukan untuk meningkatkan keragaman dataset dan mencegah *overfitting*. Metode augmentasi yang diterapkan mencakup rotasi sebesar 90° searah jarum jam dan berlawanan jarum jam, memotongan gambar (*cropping*) dengan variasi dari 0% hingga 20%, minimum dan maksimum zoom, serta penambahan *noise* hingga 5% dari total piksel pada gambar.

Dengan menerapkan serangkaian langkah tersebut, dataset yang dihasilkan menjadi lebih berkualitas dan dapat mencerminkan variasi kondisi sebenarnya yang mungkin terjadi dalam lingkungan deteksi kejahatan *klitih* di Yogyakarta.

2.4 Model YOLOv6

YOLO (*You Only Look Once*) adalah algoritma deteksi objek yang dirancang untuk mendeteksi objek berupa gambar atau video dengan cepat dan akurasi yang tinggi (Susanti, Daulay and Intan, 2023). YOLOv6 adalah versi terbaru dari arsitektur YOLO yang dirancang untuk aplikasi industri dan dikenal sebagai model yang efisien dan memiliki kinerja yang tinggi. YOLOv6 bekerja secara single-stage, atau dalam satu tahap, dan tidak memerlukan pengetahuan tentang latar belakang objek (*anchor*) sebelumnya (Norkobil Saydirasulovich *et al.*, 2023). YOLOv6 tersedia dalam enam jenis berdasarkan Gambar 6, yaitu YOLOv6-L6, YOLOv6-M6, YOLOv6-L, YOLOv6-M, YOLOv6-S dan YOLOv6-N (Li *et al.*, 2022).



Gambar 6. Karakteristik kinerja model YOLOv6

Beberapa poin penting terkait YOLOv6 adalah kemampuannya untuk dilatih pada dataset kustom. Proses pelatihan pada dataset kustom dapat dilakukan dengan menginstal dependensi YOLOv6, membuat dataset yang disesuaikan, mengonversi anotasi ke format anotasi YOLOv6, dan menyuntikkan data pembelajaran objek deteksi kustom. Fleksibilitas model ini juga terlihat dalam kemampuannya untuk menyesuaikan kecepatan pembelajaran dengan mengubah jumlah *epoch* dan ukuran gambar selama proses pelatihan.

Salah satu keunggulan lain dari YOLOv6 adalah kemampuannya untuk diubah menjadi format *Open Neural Network Exchange* (ONNX) (Li *et al.*, 2023). ONNX merupakan format serialisasi model yang umum digunakan, memungkinkan portabilitas yang lebih baik di berbagai perangkat. Dengan mengonversi model ke format ONNX, YOLOv6 mempermudah integrasinya pada berbagai platform dan infrastruktur sistem yang berbeda. Dengan demikian, YOLOv6 menawarkan solusi deteksi objek yang dapat disesuaikan dengan kebutuhan spesifik

pengguna dan memiliki potensi untuk diterapkan secara luas dalam berbagai konteks aplikasi.

2.5 Model Faster R-CNN

Faster R-CNN (*Region-based Convolutional Neural Network*) adalah sebuah metode deteksi objek yang menggabungkan dua komponen utama, yaitu *Region Proposal Network* (RPN) dan *Fast R-CNN*, yang memungkinkan deteksi objek yang lebih cepat dan akurat (Jiang *et al.*, 2022). RPN digunakan untuk menghasilkan *proposal area* yang berpotensi mengandung objek, sementara *Fast R-CNN* digunakan untuk melakukan klasifikasi dan regresi pada *proposal area* tersebut (Aditya P, Wibowo and Rahmania, 2021). Dengan menggabungkan kedua komponen ini, *Faster R-CNN* mampu mencapai kinerja deteksi objek yang unggul.

Cara kerja *Faster R-CNN* dimulai dengan proses ekstraksi fitur menggunakan *Convolutional Neural Network* (CNN) untuk menghasilkan peta fitur dari gambar input. Selanjutnya, RPN digunakan untuk menghasilkan *proposal area* yang berpotensi mengandung objek berdasarkan peta fitur tersebut. RPN mengambil gambar dari berbagai ukuran sebagai input dan output sekumpulan *proposal area* persegi panjang, masing-masing dengan skor objektivitas (Charli *et al.*, 2020). *Proposal area* ini kemudian disaring dan diklasifikasikan menggunakan *Fast R-CNN* untuk menentukan apakah area tersebut mengandung objek dan untuk melakukan regresi terhadap *bounding box* objek. Proses ini memungkinkan *Faster R-CNN* untuk secara efisien dan akurat mendeteksi objek dalam gambar dengan kinerja yang unggul.

2.6 Model SSD

SSD (*Single Shot Multibox Detector*) adalah algoritma deep learning yang bekerja dengan pendekatan pada *feed-forward convolutional network* yang mendiskritisasi ruang output dari kotak pembatas menjadi satu set kotak standar pada berbagai rasio dan skala aspek per lokasi *feature maps* (Liu *et al.*, 2016). Metode ini memungkinkan deteksi objek dilakukan dalam satu langkah, tanpa memerlukan tahap *region proposal* seperti pada metode deteksi lainnya. SSD memanfaatkan *grid cell hierarchy* yang memungkinkan deteksi objek pada berbagai ukuran dan lokasi dalam gambar. Selain itu, SSD juga menggunakan *multiple feature maps* dengan resolusi yang berbeda untuk menghasilkan prediksi yang lebih akurat.

SSD memiliki dua komponen utama, yaitu *backbone* model dan *SSD head* (Popescu *et al.*, 2022). *Backbone* model biasanya merupakan jaringan syaraf yang telah dilatih sebelumnya sebagai *feature extractor*. Sedangkan *SSD head* digunakan untuk menghasilkan prediksi objek berdasarkan *feature maps* yang dihasilkan oleh *backbone* model. *SSD head* menggunakan *multiple feature maps* dengan

resolusi yang berbeda untuk menghasilkan prediksi yang lebih akurat. Selain itu, SSD juga menggunakan *default boxes* yang memungkinkan deteksi objek pada berbagai skala dan *aspect ratio* (Dompeipen *et al.*, 2021).

3. HASIL DAN PEMBAHASAN

Hasil evaluasi model pendeteksian *klitih* yang melibatkan pencampuran gambar yang mengandung label dengan gambar yang tidak mengandung label seperti pada Gambar 7 menunjukkan sejumlah temuan yang signifikan. Pada dasarnya, model tersebut berhasil dalam mengidentifikasi dan membedakan antara objek atau situasi yang relevan dengan kejahatan *klitih* yang diinginkan dan konteks lainnya yang tidak relevan. Dengan mempertimbangkan campuran gambar yang mencakup skenario nyata yang kompleks, di mana objek yang dicari bersamaan dengan unsur-unsur yang tidak terkait, model menunjukkan tingkat akurasi yang memuaskan.



Gambar 7. Contoh dataset tanpa label (netral)

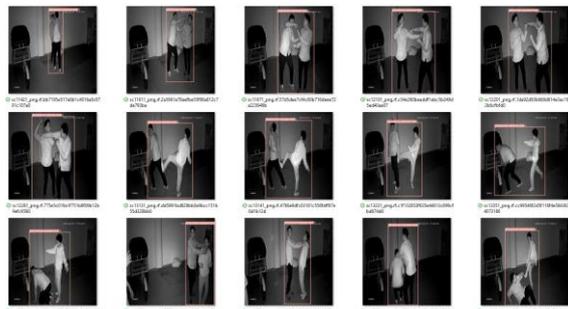
Evaluasi yang dilakukan ini melibatkan situasi kehidupan nyata di mana *klitih* mungkin tersembunyi di tengah konteks yang kompleks dan beragam. Hasil evaluasi tersebut membuktikan kemampuan model dalam memfilter informasi dengan cermat dan mengidentifikasi *klitih* bahkan dalam situasi yang rumit. Dalam keseluruhan, hasil evaluasi ini memberikan wawasan berharga untuk pengembangan dan peningkatan model pendeteksian *klitih*, yang dapat diterapkan secara efektif dalam pengawasan dan pemantauan keamanan swadaya masyarakat.

3.1 Hasil Uji Peforma Model

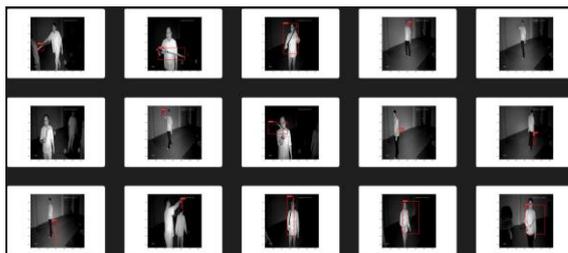
Hasil pengujian menggunakan YOLOv6 menunjukkan performa yang paling optimal dilihat dari penggunaan CPU dan Memory, serta *inference time* meskipun dari sisi akurasi model *Faster R-CNN* masih lebih baik. Gambar 8-13 menunjukkan contoh hasil deteksi senjata tajam dan gestur oleh model YOLOv6, *Faster R-CNN*, dan SSD.



Gambar 8. Hasil deteksi “*weapon*” oleh model YOLOv6



Gambar 9. Hasil deteksi “*physical abuse*” oleh model YOLOv6



Gambar 10. Hasil deteksi “*weapon*” oleh model *Faster R-CNN*



Gambar 11. Hasil deteksi “*physical abuse*” oleh model *Faster R-CNN*



Gambar 12. Hasil deteksi “*weapon*” oleh model SSD



Gambar 13. Hasil deteksi “physical abuse” oleh model SSD

Hasil dari confusion matrix untuk tiga model deteksi objek, yaitu YOLOv6, *Faster R-CNN*, dan SSD, dapat memberikan pemahaman yang mendalam tentang kinerja masing-masing model dalam tugas klasifikasi objek. Matriks tersebut mencakup empat kategori utama: *True Positive* (TP), *true negative* (TN), *False Positive* (FP), dan *false negative* (FN).

Tabel 1. Confusion matrix untuk evaluasi model

Model	Matriks Weapon	Matriks Physical Abuse																								
YOLOv6	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>15</td><td>0</td></tr> <tr><td colspan="2"></td><td>3</td><td>12</td></tr> </table>			Actual		Prediction		15	0			3	12	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>15</td><td>0</td></tr> <tr><td colspan="2"></td><td>11</td><td>4</td></tr> </table>			Actual		Prediction		15	0			11	4
		Actual																								
Prediction		15	0																							
		3	12																							
		Actual																								
Prediction		15	0																							
		11	4																							
<i>Faster R-CNN</i>	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>13</td><td>2</td></tr> <tr><td colspan="2"></td><td>4</td><td>11</td></tr> </table>			Actual		Prediction		13	2			4	11	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>14</td><td>1</td></tr> <tr><td colspan="2"></td><td>10</td><td>5</td></tr> </table>			Actual		Prediction		14	1			10	5
		Actual																								
Prediction		13	2																							
		4	11																							
		Actual																								
Prediction		14	1																							
		10	5																							
SSD	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>13</td><td>2</td></tr> <tr><td colspan="2"></td><td>3</td><td>12</td></tr> </table>			Actual		Prediction		13	2			3	12	<table border="1"> <tr><td colspan="2"></td><td colspan="2">Actual</td></tr> <tr><td colspan="2">Prediction</td><td>14</td><td>1</td></tr> <tr><td colspan="2"></td><td>12</td><td>3</td></tr> </table>			Actual		Prediction		14	1			12	3
		Actual																								
Prediction		13	2																							
		3	12																							
		Actual																								
Prediction		14	1																							
		12	3																							

Pengujian kinerja model menggunakan *confusion matrix* dengan data uji yang terdapat obyek senjata tajam (*weapon*) sebanyak 15 dan obyek kekerasan fisik (*physical abuse*) sebanyak 15. Data uji tanpa obyek *weapon/physical abuse* masing-masing sebanyak 15 (total 30). Hasil *True Positive* (TP) menyatakan bahwa model berhasil dalam mengenali adanya obyek *weapon/physical abuse* pada data uji. Hasil *True Negative* (TN) menyatakan bahwa model berhasil dalam mengenali ketiadaan obyek *weapon/physical abuse* pada data uji. Hasil *False Positive* (FP) menyatakan bahwa model salah dalam mengenali adanya obyek *weapon/physical abuse* pada data uji. Hasil *False Negative* (FN) menyatakan bahwa model salah dalam mengenali ketiadaan obyek *weapon/physical abuse* pada data uji.

Confusion matrix pada Tabel 1 menunjukkan hasil deteksi *weapon/physical abuse* oleh seluruh model. Pada kasus deteksi senjata tajam, model YOLOv6 menunjukkan kinerjanya berdasarkan nilai TP = 15, TN = 12, FP = 0, dan FN = 3. Dan pada kasus deteksi kekerasan fisik, model YOLOv6 menunjukkan kinerjanya dengan nilai TP = 15, TN = 4, FP = 0, dan FN = 11, hipotesis awal penyebab dari

besarnya nilai FN adalah kekurangan dataset latihan untuk label *physical abuse*.

Model *Faster R-CNN* selama pengujian deteksi senjata tajam menghasilkan nilai TP = 13, TN = 11, FP = 2, dan FN = 4. Dan pada percobaan deteksi kekerasan fisik, model *Faster R-CNN* menunjukkan kinerjanya dengan nilai TP = 14, TN = 5, FP = 1, dan FN = 10.

Sementara itu, model SSD menunjukkan hasil yang hampir sama dengan *Faster R-CNN* dalam deteksi senjata tajam, yaitu nilai TP = 13, TN = 12, FP = 2, FN = 3. Dan untuk percobaan kasus deteksi kekerasan fisik, model SSD menghasilkan nilai TP = 14, TN = 3, FP = 1, dan FN = 12.

Secara keseluruhan, berdasarkan hasil pengujian dengan instrumen *confusion matrix*, model YOLOv6 menunjukkan performa yang lebih unggul dalam mendeteksi senjata tajam dan kekerasan fisik dibandingkan dengan model *Faster R-CNN* dan SSD, model YOLOv6 memiliki tingkat akurasi dan *recall* yang lebih tinggi seperti ditunjukkan oleh tabel 2 dan tabel 3.

Tabel 2. Matrik evaluasi untuk label “weapon”

Model	Result Label Weapon			
	Accuracy	Precision	Recall	F1-score
YOLOv6	0,9	0,83	1	0,91
<i>Faster R-CNN</i>	0,8	0,76	0,87	0,81
SSD	0,83	0,81	0,87	0,84

Tabel 3. Matrik evaluasi untuk label “physical abuse”

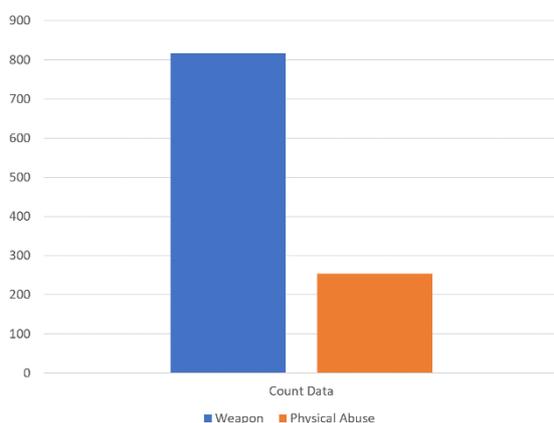
Model	Result Label Physical Abuse			
	Accuracy	Precision	Recall	F1-score
YOLOv6	0,63	0,58	1	0,73
<i>Faster R-CNN</i>	0,63	0,58	0,93	0,72
SSD	0,56	0,54	0,93	0,68

Hasil perhitungan kinerja model deteksi kejahatan *klitih* menunjukkan perbandingan yang menarik antara tiga model yang diuji, yaitu YOLOv6, *Faster R-CNN*, dan SSD. Hasil kinerja model untuk deteksi senjata tajam berdasarkan tabel 2 menunjukkan bahwa kinerja model YOLOv6 lebih unggul dengan akurasi sebesar 90%, presisi 83%, *recall* 100%, dan nilai F1-score 91%, hal ini menunjukkan bahwa kemampuan model YOLOv6 cukup baik dalam mengidentifikasi senjata tajam dengan tepat dengan prosentasi kesalahan prediksi yang rendah. Sementara itu, pada kategori deteksi kekerasan fisik pada table 3, YOLOv6 juga menunjukkan hasil yang baik dengan akurasi 63%, presisi 58%, *recall* 100%, dan nilai F1-score 73%. Meskipun terdapat model lain seperti *Faster R-CNN* dan SSD yang memiliki akurasi yang setara, namun YOLOv6 memiliki nilai *recall* yang lebih tinggi, menandakan kemampuannya untuk mendeteksi kekerasan fisik dengan lebih baik.

Secara umum, YOLOv6 mempertahankan konsistensi dalam memberikan kinerja terbaik dibandingkan dengan model lain, terutama dalam konteks deteksi senjata tajam dan kekerasan fisik. Performa yang lebih tinggi ini menunjukkan bahwa YOLOv6 dapat menjadi pilihan utama dalam implementasi teknologi pendeteksian kejahatan *klitih* di wilayah Yogyakarta, memberikan kontribusi signifikan dalam upaya meningkatkan keamanan dan pencegahan kejahatan di lingkungan tersebut.

3.2 Pembahasan

Ketidakeimbangan data dalam dataset pelatihan dapat memiliki dampak yang signifikan pada hasil model deteksi objek, dan hal ini terlihat dalam skenario evaluasi yang dilakukan, di mana label "*weapon*" memiliki 817 sampel, sementara label "*physical abuse*" hanya memiliki 254 sampel. Ketidakeimbangan ini dapat memengaruhi kinerja model dengan membuatnya cenderung lebih akurat dalam mengenali objek dengan label yang lebih banyak.



Gambar 14. Perbandingan jumlah dataset

Pada Gambar 14 terlihat ketidak seimbangan dataset, sehingga model cenderung memiliki predisposisi untuk mengenali objek dengan label "*weapon*" dengan lebih baik karena jumlah sampel pelatihan yang lebih besar untuk kategori tersebut. Model mungkin menjadi kurang terlatih dan kurang akurat dalam mengenali objek yang termasuk dalam kategori "*physical abuse*" karena jumlah sampel pelatihan yang lebih sedikit. Sebagai hasilnya, kemampuan model dalam mengidentifikasi dan memprediksi objek dengan label "*physical abuse*" dapat mengalami keterbatasan, dan hasil prediksi pada kategori ini mungkin kurang akurat.

Untuk mengatasi masalah ketidak seimbangan data, langkah-langkah tertentu dapat diambil, seperti *oversampling*, *undersampling*, dan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk memastikan bahwa setiap kategori memiliki representasi yang cukup dalam set pelatihan (Wongvorachan, He and Bulut, 2023). Selain itu, dapat dipertimbangkan juga teknik-teknik seperti

penyesuaian berat kelas atau penggunaan metode ensemble learning untuk menangani ketidak seimbangan data (Kemala and Wijayanto, 2021). Upaya untuk mengatasi ketidakseimbangan data ini dapat membantu meningkatkan akurasi dan keseimbangan model dalam mengenali objek dengan label "*physical abuse*".

4. KESIMPULAN

Kami melakukan perbandingan eksperimental yang cermat dari tiga model deteksi objek berbasis CNN yang sudah mapan, yaitu YOLOv6, *Faster R-CNN*, dan SSD untuk mendeteksi kejadian kriminal *klitih* menggunakan kamera CCTV. Kinerja ketiganya, terutama dalam hal akurasi dan kecepatan deteksi, dievaluasi menggunakan dataset gambar terintegrasi dan independen. Hasil terbaik secara keseluruhan diperoleh dengan model YOLOv6 dalam mendeteksi objek senjata tajam dan pelaku kekerasan fisik berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*. Model *Faster R-CNN* juga memberikan hasil yang menjanjikan terutama pada *recall* atau kemampuan mendeteksi korban kekerasan pelaku *klitih*. Sementara itu, SSD terbukti lebih unggul dalam efisiensi waktu komputasi meskipun kalah dalam akurasi dan presisi deteksi objek dibandingkan dengan YOLOv6 dan *Faster R-CNN*.

YOLOv6 dan *Faster R-CNN* menunjukkan kekuatan masing-masing baik pada akurasi maupun kecepatan untuk deteksi kejahatan *klitih*. Sedangkan untuk SSD, ia memiliki potensi kegunaan untuk kasus penggunaan deteksi real-time mengingat waktu pemrosesan gambar yang cepat. Ketiga model CNN ini dapat diimplementasikan untuk sistem deteksi otomatis guna mencegah dan memerangi kasus kejahatan *klitih*.

Hasil penelitian ini dapat dikembangkan di masa yang akan datang dengan melakukan integrasi terhadap sistem basis data. Informasi yang terkumpul dapat disajikan dalam bentuk *dashboard information system* yang menyajikan bukti gambar, lokasi, dan waktu kejadian. Hal ini akan sangat bermanfaat bagi pihak berwajib untuk menegakkan hukum maupun tindakan pencegahan secara cepat.

5. UCAPAN TERIMAKASIH

Penelitian ini didanai oleh hibah penelitian fundamental dari Kemendikbudristek pada tahun fiskal 2023 dengan nomor kontrak 0423.13/LL5-INT/AL.04/2023. Penulis menyatakan tidak ada konflik kepentingan yang terkait dengan penelitian ini. Semua analisis dan kesimpulan dalam penelitian ini adalah hasil kerja mandiri penulis.

DAFTAR PUSTAKA

ADITYA P, A., WIBOWO, S. A. AND RAHMANIA, R. 2021. Investigasi pengaruh Step Training pada Skema Same-Padding untuk Metode *Faster R-CNN* dalam

- Teknologi Augmented Reality. *Jurnal Ilmiah FIFO*, 12(2), p. 128. doi: 10.22441/fifo.2020.v12i2.002.
- CHARLI, F. *et al.* 2020. Implementasi Metode *Faster Region Convolutional Neural Network (Faster R-CNN)* Untuk Pengenalan Jenis Burung Lovebird. *Journal of Information Technology Ampera*, 1(3), pp. 185–197. doi: 10.51519/journalita.volume1.issuue3.year2020.page185-197.
- DOMPEIPEN, T. A. *et al.* 2021. Computer Vision Implementation for Detection and Counting the Number of Humans. *Jurnal Teknik Informatika*, 16(1), pp. 65–76. doi: 10.35793/jti.v16i1.31471.
- J, N. T. AND THINAKARAN, K. 2023. Detection of Crime Scene Objects using Deep Learning Techniques', in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, pp. 357–361. doi: 10.1109/IDCIoT56793.2023.10053440.
- JIANG, W. *et al.* 2022. Artificial Neural Networks and Deep Learning Techniques Applied to Radar Target Detection: A Review. *Electronics*, 11(1), p. 156. doi: 10.3390/electronics11010156.
- KEMALA, I. & WIJAYANTO, A. W. 2021. Perbandingan Kinerja Metode Bagging dan Non-Ensemble *Machine learning* pada Klasifikasi Wilayah di Indonesia menurut Indeks Pembangunan Manusia. *Jurnal Sistem dan Teknologi Informasi (Justin)*, 9(2), p. 269. doi: 10.26418/justin.v9i2.44166.
- LI, C. *et al.* 2022. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv preprint arXiv:2209.02976*.
- LI, M. *et al.* 2023. LFD-CD: Peripheral Blood Cells Detection Using a Lightweight Cell Detection Model with Full-Connection and Dropconnect, in, pp. 623–633. doi: 10.1007/978-3-031-46677-9_43.
- LIU, W. *et al.* 2016. SSD: *Single Shot Multibox Detector*, *arXiv preprint arXiv:1512.02325*.
- NORKOBIL SAYDIRASULOVICH, S. *et al.* 2023. A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments. *Sensors*, 23(6), p. 3161. doi: 10.3390/s23063161.
- PATEL, H. & UPLA, K. P. 2020. *Night vision Surveillance: Object Detection using Thermal and Visible Images*, in *2020 International Conference for Emerging Technology (INCET)*. IEEE, pp. 1–6. doi: 10.1109/INCET49848.2020.9154066.
- POPESCU, D. *et al.* 2022. Comparative Study of Neural Networks Used in Halyomorpha Halys Detection, in *2022 30th Mediterranean Conference on Control and Automation (MED)*. IEEE, pp. 182–187. doi: 10.1109/MED54222.2022.9837254.
- SARMINI, M., KURNIYATUTI, N. & SUKARTININGSIH, S. 2018. Klitih: Invisible Crime by Teenagers. *Proceedings of the 1st International Conference on Social Sciences (ICSS 2018)*. Paris, France: Atlantis Press, pp. 1578–1582. doi: 10.2991/icss-18.2018.328.
- SINGH, A. *et al.* 2021. IoT Based Weapons Detection System for Surveillance and Security Using YOLOV4. *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, pp. 488–493. doi: 10.1109/ICCES51350.2021.9489224.
- SOCHA, R. & KOGUT, B. 2020. Urban Video Surveillance as a Tool to Improve Security in Public Spaces. *Sustainability*, 12(15), p. 6210. doi: 10.3390/su12156210.
- SUKIRNO, S. 2018. Pencegahan klitih Melalui Pendekatan Budaya Baca Pada Siswa Di Daerah Istimewa Yogyakarta. *Jurnal IPI (Ikatan Pustakawan Indonesia)*, 3(1), pp. 28–37.
- SUSANTI, L., DAULAY, N. K. & INTAN, B. 2023. Sistem Absensi Mahasiswa Berbasis Pengenalan Wajah Menggunakan Algoritma YOLOv5. *JURIKOM (Jurnal Riset Komputer)*, 10(2), p. 640. doi: 10.30865/jurikom.v10i2.6032.
- WANG, H., WANG, W. & LIU, Y. 2020. X-YOLO: A deep learning based toolset with multiple optimization strategies for contraband detection. *Proceedings of the ACM Turing Celebration Conference - China*. New York, NY, USA: ACM, pp. 127–132. doi: 10.1145/3393527.3393549.
- WIJANARKO, A. & GINTING, R. 2021. Kejahatan Jalanan Klitih Oleh Anak Di Yogyakarta. *Recidive: Jurnal Hukum Pidana dan Penanggulangan Kejahatan*, 10(1), p. 23. doi: 10.20961/recidive.v10i1.58845.
- WONGVORACHAN, T., HE, S. & BULUT, O. 2023. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information*, 14(1), p. 54. doi: 10.3390/info14010054.
- ZHANG, Y., WIELAND, M. AND BASRAN, P. S. 2022. Unsupervised Few Shot Key Frame Extraction for Cow Teat Videos. *Data*, 7(5), p. 68. doi: 10.3390/data7050068.