

YOLO: Deteksi Objek Terpadu dan Real-Time

Prames Ray Lopian

140810210059



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PADJADJARAN**

2023

ABSTRAK

YOLO (You Only Look Once) adalah pendekatan baru untuk deteksi objek yang mengatasi masalah sebagai tugas regresi, memprediksi kotak pembatas dan probabilitas kelas langsung dari gambar penuh dalam satu evaluasi. Arsitektur terpadu YOLO dirancang untuk pemrosesan waktu nyata, mencapai kecepatan luar biasa 45 frame per detik. Sistem ini mengungguli sistem deteksi objek real-time lainnya, menunjukkan lebih dari dua kali rata-rata presisi rata-rata. Terinspirasi oleh GoogLeNet, YOLO dilatih menggunakan kumpulan data ImageNet, mengoptimalkan kesalahan jumlah kuadrat selama proses pelatihan. Makalah ini juga membahas trade-off dan keterbatasan YOLO dalam hal akurasi dan kesalahan lokalisasi, serta menyoroti potensinya sebagai model deteksi objek yang cepat dan efisien.

Makalah ini menyajikan eksperimen dan hasil yang menunjukkan YOLO sebagai pendeteksi objek tercepat yang ada, melampaui sistem deteksi real-time lainnya. Laporan ini juga menganalisis kesalahan yang dibuat oleh YOLO dan Fast R-CNN, yang menunjukkan bahwa menggabungkan kedua model ini dapat meningkatkan kinerja secara signifikan. Selain itu, kinerja model deteksi objek yang berbeda, termasuk Fast R-CNN dan YOLO, dievaluasi pada kumpulan data VOC 2007 dan VOC 2012. Perbandingannya mencakup metrik seperti mAP (rata-rata presisi) dan presisi rata-rata per kelas. Generalisasi model-model ini pada karya seni dan deteksi real-time menggunakan webcam juga diperiksa. YOLO ditekankan sebagai pendeteksi objek yang cepat dan akurat, khususnya dalam aplikasi real-time.

Kesimpulannya, YOLO diperkenalkan sebagai model terpadu untuk deteksi objek, menawarkan kesederhanaan dalam konstruksi dan pelatihan langsung pada gambar penuh. Ini diakui sebagai pendeteksi objek serba guna tercepat dalam literatur, memajukan deteksi objek terancang dalam waktu nyata. Kemampuan sistem untuk menggeneralisasi dengan baik ke domain baru membuatnya cocok untuk aplikasi yang memerlukan deteksi objek yang cepat dan kuat.

DAFTAR ISI

ABSTRAK	i
DAFTAR ISI	ii
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	2
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
1.6. Metodologi Penelitian	6
1.7. Sistematika Penulisan	10
II TINJAUAN PUSTAKA	14
2.1. Kajian Teori	14
2.1.1. Jaringan Syaraf Tiruan (Artificial Neural Networks - ANN) .	15
2.1.2. Model Bagian Terdeformasi (Deformable Part Models - DPM)	15
2.1.3. Optimasi Algoritma (Optimization Algorithms)	16
2.1.4. Pengaplikasian pada Kasus Klasifikasi Tertentu	17
2.1.5. R-CNN (Region-based Convolutional Neural Network) . . .	17
2.1.6. Faster R-CNN	18

2.1.7. SSD (Single Shot MultiBox Detector)	19
2.1.8. RetinaNet	19
2.1.9. EfficientDet	20
2.1.10. <i>dataset</i> VOC 2007	20
2.1.11. <i>dataset</i> VOC 2012	21
III METODE PENELITIAN	22
3.1. Objek Penelitian	22
3.2. Metode Penelitian	22
3.2.1. Langkah-langkah Penelitian	22
3.2.2. Mekanisme, Proses, dan <i>Software/Library</i>	23
3.2.3. Pengumpulan Data	24
3.2.4. Pengukuran Keberhasilan Metode	25
DAFTAR PUSTAKA	27

BAB I

PENDAHULUAN

1.1. Latar Belakang

Latar belakang penelitian ini melibatkan evolusi pendekatan dalam deteksi objek, sebuah aspek kunci dalam pengembangan sistem pengenalan visual komputer. Pendekatan sebelumnya cenderung memanfaatkan klasifikasi untuk tujuan deteksi, yang melibatkan evaluasi *klasifier* pada berbagai lokasi dan skala dalam gambar uji. Metode terbaru, seperti R-CNN, mengusulkan kotak pembatas potensial terlebih dahulu dalam gambar dan kemudian menjalankan *klasifier* pada kotak-kotak yang diusulkan tersebut. Namun, pendekatan ini cenderung kompleks dan sulit dioptimalkan karena masing-masing komponen harus dilatih secara terpisah.

Penelitian ini mengubah paradigma dengan merancang deteksi objek sebagai masalah regresi langsung dari piksel gambar ke koordinat kotak pembatas dan probabilitas kelas. YOLO (You Only Look Once) menawarkan pendekatan yang sangat cepat, memproses gambar secara real-time tanpa memerlukan pipa deteksi yang rumit. Kecepatan ini diperoleh dengan menjalankan jaringan saraf pada gambar baru selama pengujian untuk memprediksi deteksi. Metode ini memungkinkan pengolahan *video streaming* secara *real-time* dengan latensi kurang dari 25 milidetik, sementara tetap mencapai *mean average precision* lebih dari dua kali lipat dari sistem *real-time* lainnya.

Salah satu keunggulan utama YOLO adalah kemampuannya untuk merasion-

alkan secara global tentang gambar saat membuat prediksi. Berbeda dengan teknik berbasis jendela geser dan proposal wilayah, YOLO melihat seluruh gambar selama pelatihan dan pengujian, secara implisit mengkodekan informasi kontekstual tentang kelas dan penampilan objek. Hal ini meminimalkan jumlah kesalahan latar belakang dibandingkan dengan metode seperti Fast R-CNN.

Selain kecepatan dan keunggulan kontekstual, YOLO juga mampu mempelajari representasi objek yang dapat digeneralisasi dengan baik. Hasil eksperimen menunjukkan bahwa YOLO dapat mengungguli metode deteksi lainnya, termasuk DPM dan R-CNN, ketika diuji pada domain yang berbeda, seperti seni. Keunggulan generalisasi ini membuat YOLO lebih tahan terhadap perubahan domain dan *input* yang tidak terduga. Meskipun demikian, penelitian juga mengidentifikasi beberapa *trade-off*, terutama dalam lokalilasi objek yang presisi, yang perlu dieksplorasi lebih lanjut.

1.2. Identifikasi Masalah

Berdasarkan uraian latar belakang yang telah disampaikan sebelumnya, penelitian ini akan difokuskan pada upaya penyelesaian terhadap beberapa masalah berikut:

1. Apa yang menjadi kelemahan YOLO dalam hal akurasi dan kesalahan lokal?
2. Bagaimana YOLO membandingkan dirinya dengan sistem deteksi lainnya dalam hal kecepatan dan akurasi?
3. Apa yang menjadi keunggulan YOLO dalam mempelajari representasi objek yang sangat umum?

4. Bagaimana YOLO memproses gambar secara *real-time* pada 45 *frame* per detik?
5. Apa yang menjadi kompromi dan batasan YOLO dalam hal akurasi dan kesalahan lokal?

1.3. Batasan Masalah

Dengan mengenali masalah-masalah yang telah ada, terdapat beberapa batasan masalah yang akan menjadi fokus utama dari penelitian ini. Berikut diantaranya:

1. Penelitian ini terbatas pada analisis dan deteksi objek dalam gambar menggunakan pendekatan YOLO.
2. Fokus utama penelitian ini adalah pada evaluasi kinerja YOLO dalam mendeteksi objek dalam gambar secara *real-time*.
3. Fokus utama penelitian ini juga pada penerapan YOLO dalam deteksi objek secara *real-time*, terutama dalam aplikasi *webcam*.
4. Penelitian ini membedakan perbandingan YOLO dengan sistem deteksi lainnya dalam hal kecepatan dan akurasi.
5. Analisis kesalahan yang dilakukan oleh YOLO dan Fast R-CNN menjadi fokus utama dalam penelitian ini.
6. Penelitian ini terbatas pada pengujian kinerja model deteksi objek, terutama pada dataset VOC 2007 dan VOC 2012.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memperkenalkan YOLO sebagai pendekatan baru untuk mendeteksi objek yang sangat cepat dan efisien, serta membandingkannya dengan sistem deteksi objek lainnya. Tujuan utama penelitian ini adalah untuk menunjukkan bahwa YOLO adalah model deteksi objek waktu nyata tercepat yang ada, dengan hasil yang lebih baik daripada sistem deteksi waktu nyata lainnya.

Tujuan Penelitian:

1. Memperkenalkan YOLO sebagai pendekatan baru untuk mendeteksi objek yang sangat cepat dan efisien.
2. Membandingkan YOLO dengan sistem deteksi objek lainnya.
3. Mendeskripsikan proses pelatihan, inferensi, keterbatasan, dan perbandingan YOLO dengan sistem lain.
4. Menyajikan eksperimen dan hasil yang menunjukkan YOLO sebagai detektor objek waktu nyata tercepat yang ada.
5. Menganalisis kesalahan yang dilakukan oleh YOLO dan Fast R-CNN, serta menunjukkan bahwa kombinasi keduanya dapat meningkatkan kinerja.
6. Meneliti generalisabilitas model-model tersebut pada karya seni dan deteksi waktu nyata menggunakan *webcam*.

1.5. Manfaat Penelitian

Jurnal penelitian ini diharapkan dapat memberikan manfaat yang signifikan dalam beberapa aspek. Jurnal ini memperkenalkan YOLO sebagai pendekatan baru untuk mendeteksi objek yang sangat cepat dan efisien, memberikan kontribusi penting bagi pengembang dan peneliti dalam memahami dan menggunakan metode deteksi objek yang inovatif dan canggih. Selain itu, jurnal ini memberikan wawasan yang berharga tentang keunggulan dan kelemahan YOLO dibandingkan dengan pendekatan lain, membantu dalam pemilihan metode deteksi objek yang paling sesuai untuk aplikasi tertentu. Analisis kinerja YOLO dan perbandingannya dengan sistem deteksi lainnya juga memberikan wawasan yang berguna bagi pengembangan model deteksi objek. Selain itu, jurnal ini memberikan kontribusi terhadap literatur dengan memperkenalkan YOLO sebagai model deteksi objek yang cepat, efisien, dan dapat diandalkan, menjadi referensi penting bagi peneliti dan praktisi dalam bidang deteksi objek.

Manfaat Penelitian:

1. Jurnal ini memperkenalkan YOLO sebagai pendekatan baru untuk mendeteksi objek yang sangat cepat dan efisien, memberikan kontribusi penting bagi pengembang dan peneliti dalam memahami dan menggunakan metode deteksi objek yang inovatif dan canggih.
2. Jurnal ini membandingkan YOLO dengan sistem deteksi objek lainnya, memberikan wawasan yang berharga tentang keunggulan dan kelemahan YOLO dibandingkan dengan pendekatan lain, membantu dalam pemilihan metode

deteksi objek yang paling sesuai untuk aplikasi tertentu.

3. Jurnal ini menganalisis kinerja YOLO dan membahas kesalahan yang dilakukan oleh model tersebut, memberikan wawasan yang berguna bagi pengembangan model deteksi objek lainnya.
4. Jurnal ini melakukan evaluasi kinerja model deteksi objek seperti Fast R-CNN dan YOLO pada *dataset* VOC 2007 dan VOC 2012, memberikan pemahaman tentang kinerja relatif dari berbagai model deteksi objek.
5. Jurnal ini mengeksplorasi generalisabilitas YOLO pada domain baru dan aplikasi deteksi objek waktu nyata, memberikan wawasan tentang kemungkinan penerapan YOLO dalam berbagai konteks dan aplikasi.
6. Jurnal ini memberikan kontribusi terhadap sastra dengan memperkenalkan YOLO sebagai model deteksi objek yang cepat, efisien, dan dapat dipercaya, menjadi referensi penting bagi peneliti dan praktisi dalam bidang deteksi objek.

1.6. Metodologi Penelitian

Berikut merupakan beberapa tahapan penelitian yang dilakukan:

Jenis Penelitian:

Riset dan Pengembangan.

Langkah-langkah Penelitian:

1. Studi Literatur

2. Pengumpulan Data:

- Pengumpulan data citra dari berbagai sumber, termasuk *dataset* publik dan citra dari internet.
- Pemilihan *dataset* yang representatif untuk melatih dan menguji model deteksi objek.
- Pemrosesan data citra untuk mempersiapkan *dataset* latihan dan pengujian.
- Anotasi data citra dengan label objek yang tepat untuk melatih model deteksi objek.
- Verifikasi kualitas dan keberagaman *dataset* untuk memastikan representasi yang baik dari objek yang akan dideteksi.

3. Pembahasan Metodologi Penelitian:

- Penelitian ini menggunakan pendekatan literatur untuk mempelajari pendekatan deteksi objek yang ada, mengidentifikasi kelemahan dan keunggulan pendekatan tersebut, dan memperkenalkan YOLO sebagai pendekatan yang memiliki kecepatan dan efisiensi tinggi dalam deteksi objek
- Penelitian ini mengumpulkan data hasil kinerja YOLO dan sistem deteksi lainnya, melakukan analisis perbandingan kinerja antara YOLO dan sistem deteksi lainnya, serta menunjukkan keunggulan YOLO dalam hal kecepatan, akurasi, dan generalisabilitas.

- Penelitian ini melakukan analisis kinerja dan kesalahan YOLO, menjelaskan proses pelatihan dan inferensi YOLO, serta membandingkan batasan dan keunggulan YOLO dengan sistem deteksi lainnya. .
- Penelitian ini melakukan pengujian kinerja YOLO pada dataset yang representatif, mengukur kinerja model dalam hal akurasi dan kecepatan, serta membandingkan hasil evaluasi dengan model deteksi objek lainnya.
- Penelitian ini melakukan analisis kinerja dan kesalahan YOLO dan Fast R-CNN, serta menunjukkan bahwa kombinasi keduanya dapat meningkatkan kinerja deteksi objek

4. Penerapan Metode:

- Pengenalan YOLO sebagai pendekatan baru untuk mendeteksi objek dilakukan dengan melakukan studi literatur tentang pendekatan deteksi objek yang ada, mengidentifikasi kelemahan dan keunggulan pendekatan tersebut, serta memperkenalkan YOLO sebagai pendekatan yang memiliki kecepatan dan efisiensi tinggi dalam deteksi objek.
- Perbandingan YOLO dengan sistem deteksi lainnya dilakukan dengan mengumpulkan data hasil kinerja YOLO dan sistem deteksi lainnya, melakukan analisis perbandingan kinerja antara YOLO dan sistem deteksi lainnya, serta menunjukkan keunggulan YOLO dalam hal kecepatan, akurasi, dan generalisabilitas
- Analisis kinerja dan kesalahan YOLO dilakukan dengan mengumpulkan

data hasil deteksi YOLO, menganalisis kesalahan yang terjadi, dan mengidentifikasi faktor-faktor yang mempengaruhi kinerja YOLO dalam mendeteksi objek. .

- Evaluasi model deteksi objek dilakukan dengan menguji model YOLO pada dataset yang representatif, mengukur performa model dalam hal akurasi dan kecepatan, serta membandingkan hasil evaluasi dengan model deteksi objek lainnya.
- Eksplorasi generalisabilitas YOLO pada domain baru dilakukan dengan menguji model YOLO pada domain atau dataset yang berbeda, mengamati kinerja model dalam domain baru, dan mengeluarkan kemampuan generalisasi model YOLO.

5. Evaluasi Metode:

- Melakukan pengujian kecepatan dan efisiensi YOLO dalam deteksi objek dengan menggunakan berbagai jenis gambar dan ukuran dataset yang berbeda. Pengujian ini dapat dilakukan dengan mengukur waktu yang diperlukan oleh YOLO untuk mendeteksi objek dalam gambar dengan resolusi yang berbeda.
- Melakukan analisis perbandingan kinerja antara YOLO dan sistem deteksi lainnya, terutama dalam hal kecepatan, akurasi, dan generalisabilitas. Pengujian ini dapat dilakukan dengan menggunakan dataset yang representatif dan mengukur kinerja keduanya dalam mendeteksi objek.

- Melakukan evaluasi kinerja model YOLO pada dataset yang representatif, mengukur akurasi dan kecepatan model dalam mendeteksi objek, serta membandingkan hasil evaluasi dengan model deteksi objek lainnya.
- Melakukan eksplorasi generalisabilitas YOLO pada domain atau dataset yang berbeda, mengamati kinerja model dalam domain baru, dan menyebarkan kemampuan generalisasi model YOLO.
- Melakukan analisis kinerja dan kesalahan YOLO dan Fast R-CNN, menjelaskan proses pelatihan dan inferensi keduanya, serta membandingkan batasan dan keunggulan keduanya dalam mendeteksi objek.

6. Penulisan Laporan/Skripsi:

- Menyusun laporan penelitian yang mencakup semua langkah-langkah, hasil evaluasi, temuan, dan kesimpulan dari penelitian ini.

1.7. Sistematika Penulisan

Skripsi ini disusun secara sistematis untuk memandu pembaca melalui tahapan penelitian dengan jelas. Bab "Pendahuluan" memperkenalkan konteks dan tujuan penelitian. Bab "Tinjauan Pustaka" merinci landasan teoritis dan literatur terkait. Bab "Analisis dan Perancangan" membahas langkah-langkah dari studi literatur hingga evaluasi metode deteksi. Terakhir, bab "Daftar Pustaka" mencantumkan sumber referensi yang digunakan dalam pengembangan konsep dan metode peneli-

tian. Dengan sistematika ini, skripsi ini dirancang agar mudah dipahami dan diikuti oleh pembaca.

1. Pendahuluan:

Dalam bagian Pendahuluan, akan dibahas mengenai tujuan penelitian untuk mengevaluasi metodologi deteksi objek YOLO sebagai pendekatan baru dalam deteksi objek, serta keunggulan YOLO dalam hal kecepatan dan efisiensi dalam mendeteksi objek. Selain itu, akan dibahas pula mengenai perbandingan kinerja YOLO dengan sistem deteksi objek lainnya, analisis kinerja dan kesalahan YOLO dalam mendeteksi objek, serta evaluasi kinerja model YOLO pada dataset yang representatif. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang keunggulan dan kelemahan metodologi deteksi objek YOLO, serta potensi penerapannya dalam berbagai domain.

2. Tinjauan Pustaka:

Dalam bagian Tinjauan Pustaka, akan dibahas mengenai pendekatan deteksi objek yang ada, seperti pendekatan menggunakan deep neural networks, serta pendekatan lainnya seperti object detection with discriminatively trained part-based models dan histograms of oriented gradients for human detection. Selain itu, akan dibahas juga mengenai pendekatan lainnya seperti deformable part models (DPM) yang menggunakan pendekatan *sliding window* untuk deteksi objek. Dan juga akan membahas mengenai penelitian terkait yang telah dilakukan sebelumnya, baik mengenai pendekatan deteksi objek maupun eval-

uasi kinerja model deteksi objek, untuk memberikan landasan teoritis yang kuat bagi penelitian ini.

3. Analisis dan Perancangan:

Dalam bagian Analisis dan Perancangan, akan dibahas mengenai skenario pengujian yang akan dilakukan untuk mengevaluasi kinerja YOLO dalam mendeteksi objek. Skenario pengujian meliputi pengujian kecepatan dan efisiensi YOLO dalam mendeteksi objek dengan menggunakan berbagai jenis gambar dan ukuran *dataset* yang berbeda, serta analisis perbandingan kinerja antara YOLO dan sistem deteksi lainnya dalam hal kecepatan, akurasi, dan generalisabilitas. Selain itu, akan dibahas pula evaluasi kinerja model YOLO pada *dataset* yang representatif, mengukur akurasi dan kecepatan model dalam mendeteksi objek, serta membandingkan hasil evaluasi dengan model deteksi objek lainnya.

4. Daftar Pustaka:

Dalam bagian Daftar Pustaka, akan terdapat referensi dari penelitian terkait yang telah dilakukan sebelumnya, seperti penelitian mengenai pendekatan deteksi objek menggunakan deep neural networks, object detection with discriminatively trained part-based models, histograms of oriented gradients for human detection, dan pendekatan lainnya seperti Fast R-CNN yang menggabungkan deteksi objek dan segmentasi semantik. Selain itu, akan terdapat pula referensi mengenai penelitian terkait yang membahas evaluasi kinerja model deteksi objek, seperti penelitian mengenai evaluasi kinerja model

YOLO pada *dataset* yang representatif.

BAB II

TINJAUAN PUSTAKA

2.1. Kajian Teori

Tinjauan Pustaka pada penelitian ini akan membahas sejumlah konsep kunci yang menjadi dasar teoretis serta landasan pengembangan metode deteksi objek menggunakan YOLO (You Only Look Once). Dimulai dengan pembahasan mengenai Jaringan Syaraf Tiruan (Artificial Neural Networks - ANN) sebagai fondasi utama dalam pemrosesan *visual*, kemudian menjelajahi Optimasi Algoritma yang meningkatkan efisiensi pembelajaran model. Selanjutnya, tinjauan mencakup aplikasi khusus pada kasus klasifikasi tertentu, memberikan pemahaman mendalam mengenai penggunaan YOLO dalam situasi ini.

Selain itu, akan dibahas pula perkembangan terkini dalam deteksi objek, melibatkan evolusi dari metode R-CNN (Region-based Convolutional Neural Network) hingga model-model terkini seperti EfficientDet. Setiap pendekatan akan dianalisis secara mendalam, mencakup kelebihan dan kekurangan serta relevansinya dengan penelitian ini. Penjelasan terinci juga akan diberikan mengenai beberapa *dataset* klasik yang digunakan, yaitu VOC 2007 dan VOC 2012. Keseluruhan tinjauan pustaka ini akan memberikan konteks teoretis yang kokoh dan mendalam untuk mendukung penelitian ini.

2.1.1. Jaringan Syaraf Tiruan (Artificial Neural Networks - ANN)

Jaringan syaraf tiruan (ANN) merupakan pendekatan pemodelan yang terinspirasi oleh struktur dan fungsi otak manusia. Dalam konteks deteksi objek, ANN menjadi elemen kunci dalam pengembangan model yang mampu memahami fitur-fitur kompleks dalam gambar. Terdiri dari lapisan-lapisan neuron buatan, ANN mampu melakukan ekstraksi fitur secara hierarkis, memungkinkan model untuk belajar pola-pola yang lebih kompleks. Dengan kemampuan adaptasinya terhadap data latihan, jaringan ini mampu mengenali objek-objek dalam gambar dengan akurasi yang meningkat seiring dengan peningkatan kompleksitas model.

Selain itu, ANN memiliki aplikasi yang luas dalam pengolahan citra dan visi komputer, termasuk deteksi objek. Model deep learning, seperti Convolutional Neural Networks (CNN) yang merupakan jenis khusus dari ANN, telah menunjukkan keunggulan dalam mengatasi tantangan deteksi objek seperti variasi pose dan perubahan skala. Oleh karena itu, integrasi ANN dalam kerangka deteksi objek memberikan dasar yang kokoh untuk mencapai hasil yang lebih baik.[1]

2.1.2. Model Bagian Terdeformasi (Deformable Part Models - DPM)

Deformable Part Models (DPM) merupakan salah satu pendekatan dalam deteksi objek yang telah digunakan sebagai benchmark dalam penelitian pengembangan sistem deteksi objek. Metode ini pertama kali diperkenalkan oleh Felzenszwalb, Girshick, McAllester, dan Ramanan pada tahun 2010. DPM mengusung ide bahwa objek dapat direpresentasikan sebagai kombinasi dari bagian-bagian deformable

yang dapat disesuaikan dengan variasi bentuk dan posisi. Model ini terdiri dari bagian-bagian yang saling terkait dan dapat "deform" untuk mengakomodasi variasi dalam ukuran dan bentuk objek.

DPM memanfaatkan representasi spasial dan histogram gradien sebagai fitur untuk mewakili objek. Algoritma ini memiliki kemampuan untuk mendeteksi objek dalam berbagai posisi dan skala, membuatnya sesuai untuk tugas deteksi objek pada skenario yang kompleks. Meskipun DPM telah memberikan kontribusi signifikan dalam pengembangan sistem deteksi objek, beberapa kelemahan juga telah diidentifikasi, seperti keterbatasan dalam menangani objek dengan pose yang sangat ekstrim dan variasi deformasi yang tinggi.[2]

2.1.3. Optimasi Algoritma (Optimization Algorithms)

Algoritma optimasi memiliki peran sentral dalam melatih model deep learning untuk deteksi objek. Dalam konteks ini, proses optimasi bertujuan untuk menemukan nilai optimal untuk parameter-model, sehingga model dapat memberikan prediksi yang akurat. Metode optimasi seperti Adam, Stochastic Gradient Descent (SGD), dan RMSprop menjadi kunci dalam menyesuaikan bobot dan bias agar model dapat mempelajari pola-pola yang ada dalam data latihan.

Selain itu, pemilihan algoritma optimasi juga dapat memengaruhi kecepatan konvergensi dan stabilitas model. Misalnya, algoritma Adam sering digunakan karena kemampuannya menyesuaikan laju pembelajaran secara adaptif. Oleh karena itu, pemahaman mendalam tentang berbagai algoritma optimasi sangat penting dalam merancang kerangka deteksi objek yang efisien dan andal. Integrasi al-

goritma optimasi yang tepat dapat mempercepat pelatihan model dan meningkatkan kemampuan generalisasi terhadap data uji.[3]

2.1.4. Pengaplikasian pada Kasus Klasifikasi Tertentu

Pengaplikasian jaringan syaraf tiruan (ANN) pada kasus klasifikasi tertentu, seperti deteksi objek, melibatkan pemahaman mendalam tentang konteks aplikasi. Dalam literatur, banyak penelitian yang fokus pada pengembangan model deteksi objek yang dioptimalkan untuk kasus klasifikasi spesifik, seperti deteksi wajah, mobil, atau objek lainnya. Pemilihan dan penyusunan *dataset* yang relevan dengan kasus klasifikasi tersebut menjadi kunci untuk melatih model dengan baik, meningkatkan keakuratan dan ketepatan deteksi.

Selain itu, pengaplikasian ANN pada kasus klasifikasi tertentu memerlukan penyesuaian arsitektur model dan teknik pra-pemrosesan yang sesuai dengan karakteristik objek yang akan dideteksi. Dengan demikian, penelitian terkait pengaplikasian pada kasus klasifikasi tertentu memberikan wawasan penting bagi pengembangan deteksi objek yang spesifik dan efektif.[4]

2.1.5. R-CNN (Region-based Convolutional Neural Network)

Region-based Convolutional Neural Network (R-CNN) adalah salah satu inovasi penting dalam deteksi objek yang mengatasi keterbatasan model sebelumnya dengan mengenali dan memproses wilayah (region) kandidat secara terpisah. Metode ini menggabungkan CNN untuk ekstraksi fitur dengan algoritma pemilihan wilayah (region proposal), seperti Selective Search, untuk meningkatkan efisiensi

dan akurasi deteksi objek.

R-CNN memanfaatkan pendekatan dua tahap, yaitu tahap pemilihan wilayah dan tahap klasifikasi. Pada tahap pertama, wilayah-wilayah kandidat diekstraksi, kemudian dilakukan warping dan disusun menjadi batch untuk tahap klasifikasi. Meskipun berhasil meningkatkan performa deteksi objek, R-CNN memiliki keterbatasan pada kecepatan inferensi yang rendah karena memerlukan banyak waktu untuk mengevaluasi setiap wilayah. Inovasi selanjutnya, seperti Faster R-CNN, mengatasi keterbatasan ini dengan memperkenalkan modul proposal terintegrasi.[5]

2.1.6. Faster R-CNN

Faster R-CNN memperkenalkan Region Proposal Network (RPN) yang terintegrasi langsung dalam arsitektur model. Hal ini mengatasi kekurangan kecepatan pada pendahulunya dengan menyederhanakan proses pemilihan wilayah. Dengan menggunakan RPN, Faster R-CNN dapat menghasilkan proposal wilayah dengan cepat, meningkatkan efisiensi deteksi objek secara signifikan.

Selain itu, Faster R-CNN mempertahankan keunggulan ekstraksi fitur hierarkis dari CNN, yang memungkinkan model untuk memahami konteks dan detail dalam gambar. Kemampuan model untuk secara efisien memproses wilayah kandidat membuatnya menjadi salah satu pendekatan yang sangat sukses dalam deteksi objek pada berbagai kasus pengaplikasian.[6]

2.1.7. SSD (Single Shot MultiBox Detector)

Single Shot MultiBox Detector (SSD) mencapai deteksi objek secara *real-time* dengan menggabungkan beberapa skala fitur dalam satu model. SSD memperkenalkan pendekatan single shot, yang berarti bahwa objek dapat dideteksi dalam satu iterasi feedforward, mengurangi kompleksitas dan waktu inferensi. Dengan memanfaatkan lapisan konvolusi dengan skala berbeda, SSD dapat mendeteksi objek pada berbagai ukuran dan perbandingan aspek.

Kelebihan utama SSD adalah kemampuannya untuk menangani objek dengan skala yang berbeda secara efektif, serta mengurangi kebutuhan akan proses pemilihan wilayah yang terpisah. Hal ini membuat SSD menjadi pilihan yang baik untuk aplikasi deteksi objek yang memerlukan waktu inferensi yang cepat dan akurasi yang tinggi.[7]

2.1.8. RetinaNet

RetinaNet dikenal karena kemampuannya menangani masalah ketidakseimbangan antara jumlah objek dan latar belakang dalam gambar. Masalah ini biasanya dihadapi oleh model deteksi objek karena distribusi objek yang jarang dalam *dataset*. Dengan memperkenalkan modul Focal Loss, RetinaNet dapat memberikan bobot yang lebih tinggi pada objek yang sulit dideteksi, meningkatkan akurasi deteksi secara signifikan.

Focal Loss membantu model untuk fokus pada kasus-kasus yang sulit, mengatasi tantangan dalam mengenali objek dengan tingkat kemunculan yang rendah.

Dengan demikian, RetinaNet menjadi salah satu pilihan unggul untuk kasus pengaplikasian di mana distribusi objek tidak merata dalam *dataset*. [8]

2.1.9. EfficientDet

EfficientDet merupakan evolusi dari pendekatan deteksi objek yang berfokus pada efisiensi komputasional. Model ini dirancang untuk memberikan tingkat akurasi yang tinggi dengan menggunakan sumber daya komputasional yang lebih efisien. Dengan memanfaatkan skala model yang dinamis dan pengoptimalan struktural, EfficientDet berhasil menyajikan solusi yang baik untuk mendeteksi objek dengan performa yang tinggi pada berbagai perangkat.

Keunggulan utama EfficientDet terletak pada kemampuannya untuk mencapai tingkat akurasi yang setara dengan model yang lebih besar, tetapi dengan kecepatan inferensi yang lebih cepat. Pendekatan ini menjadi penting dalam konteks deteksi objek *real-time* dan implementasi pada perangkat dengan keterbatasan komputasional. [9]

2.1.10. *dataset* VOC 2007

dataset VOC 2007 (Visual Object Classes Challenge 2007) adalah *dataset* yang sangat populer dan sering digunakan dalam penelitian deteksi objek. *dataset* ini mencakup berbagai kategori objek seperti manusia, kendaraan, dan hewan, dengan anotasi bounding box yang akurat. Keberagaman objek dalam *dataset* memungkinkan peneliti untuk menguji dan melatih model deteksi objek dengan berbagai kasus pengaplikasian.

Anotasi yang terinci memudahkan penelitian dalam mengevaluasi performa model deteksi objek, dan sejumlah besar *dataset* memastikan keberagaman yang diperlukan untuk pengembangan model yang generalis. Oleh karena itu, VOC 2007 menjadi rujukan penting dalam literatur deteksi objek dan digunakan sebagai benchmark untuk mengukur keunggulan model.[10]

2.1.11. *dataset* VOC 2012

Sebagai kelanjutan dari VOC 2007, *dataset* VOC 2012 menyediakan tambahan data dan tantangan untuk penelitian deteksi objek. *dataset* ini menambahkan variasi objek, pose, dan kondisi pencahayaan, meningkatkan tingkat kesulitan dalam tugas deteksi. Dengan menambahkan VOC 2012 sebagai *dataset* pelatihan atau uji, penelitian dapat mengukur kemampuan model untuk mengatasi situasi yang lebih kompleks dan menilai tingkat generalisasi model terhadap variasi kondisi.

Penggunaan bersama *dataset* VOC 2007 dan VOC 2012 memberikan landasan yang kokoh untuk penelitian deteksi objek, memungkinkan peneliti untuk mengembangkan dan mengevaluasi model mereka dalam berbagai konteks. Dengan melibatkan *dataset* ini, penelitian dapat menghasilkan model yang lebih robust dan mampu mengatasi tantangan dalam deteksi objek pada dunia nyata.[10]

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Objek penelitian dari jurnal referensi di atas adalah pendekatan deteksi objek menggunakan YOLO (You Only Look Once). Dalam konteks ini, YOLO diimplementasikan sebagai metode untuk melakukan deteksi objek dalam gambar atau video dengan lebih efisien dan akurat. Objek penelitian dapat didefinisikan sebagai penerapan YOLO dalam mengenali objek-objek yang muncul dalam suatu konteks *visual*.

Pada dasarnya, YOLO dirancang untuk menangkap objek-objek yang ada dalam suatu gambar atau *frame video* tanpa perlu melihatnya secara berulang kali. Oleh karena itu, objek penelitian melibatkan pengembangan dan penerapan metode YOLO, serta analisis terhadap keberhasilan dan efektivitasnya dalam deteksi objek.

3.2. Metode Penelitian

3.2.1. Langkah-langkah Penelitian

Langkah-langkah penelitian dimulai dengan seleksi *dataset* VOC 2007 dan VOC 2012 sebagai pondasi utama. Proses ini memerlukan pemahaman mendalam terhadap karakteristik *dataset*, termasuk distribusi kelas, ukuran gambar, dan variasi lainnya. Pada tahap awal, dilakukan *preprocessing* data yang sangat teliti untuk memastikan kualitas dan integritas *dataset*. Proses ini mencakup penanganan *out-*

liers, normalisasi nilai piksel, dan segmentasi *dataset* menjadi subset pelatihan dan pengujian. *Preprocessing* yang akurat adalah kunci untuk meminimalkan potensi distorsi data dan memastikan konsistensi hasil eksperimen.

Setelah persiapan *dataset* selesai, penentuan konfigurasi parameter YOLO menjadi tahap berikutnya. Ini melibatkan pemilihan parameter kritis, seperti ukuran *grid*, jumlah *anchor boxes*, dan tingkat pembelajaran (*learning rate*). Pentingnya konfigurasi parameter ini sangat berkaitan dengan kinerja deteksi objek YOLO. Selanjutnya, menerapkan algoritma optimasi, khususnya stochastic gradient descent (SGD), untuk meminimalkan fungsi kerugian dan meningkatkan akurasi model. Tahap ini menunjukkan dedikasi terhadap pengaturan yang teliti untuk memastikan model optimal.

3.2.2. Mekanisme, Proses, dan Software/Library

Dalam konteks penerapan formula dan metode YOLO, penelitian ini melibatkan implementasi yang sangat teliti untuk memastikan keakuratan dan kestabilan model. Formula YOLO, terutama terkait perhitungan fungsi kerugian dan parameter utama seperti ukuran *grid* dan *anchor boxes*, diterapkan dengan penuh perhatian. Adopsi *anchor boxes*, sebagai contoh, memerlukan analisis mendalam terhadap distribusi objek dalam dataset untuk menentukan jumlah dan ukuran anchor boxes yang optimal. Selanjutnya, pemanfaatan mekanisme *backpropagation* sebagai bagian dari optimasi model menambah kompleksitas penerapan. Proses ini membutuhkan pemahaman mendalam tentang turunan parsial fungsi kerugian terhadap parameter model, menegaskan kontribusi yang signifikan terhadap peningkatan akurasi deteksi

objek.

Pemanfaatan perangkat lunak atau *library*, dalam hal ini TensorFlow, memperkaya aspek implementasi. TensorFlow menyediakan lingkungan yang kaya fitur dan mendukung pembuatan model YOLO dengan abstraksi tinggi. Integrasi TensorFlow dengan keras sebagai antarmuka pemrograman aplikasi (API) memudahkan proses pemodelan tanpa kehilangan detail esensial. Penggunaan *library* ini juga memberikan fleksibilitas dalam penyesuaian arsitektur model dan memfasilitasi eksperimen dengan parameter dan konfigurasi yang berbeda. Pemanfaatan teknologi ini mencerminkan komitmen pada keakuratan dan efisiensi, yang merupakan fondasi dari hasil penelitian ini.

3.2.3. Pengumpulan Data

Pada tahap penggunaan data, penelitian ini mengintegrasikan *dataset* VOC 2007 dan VOC 2012. Sebagai pembuat jurnal, langkah-langkah penggunaan data diarahkan untuk memastikan keakuratan dan representatifitas hasil deteksi objek yang diperoleh dari model YOLO yang dikembangkan. Proses pengumpulan data melibatkan ekstraksi informasi yang relevan dari kedua *dataset*, memastikan keberagaman dan kompleksitas yang sesuai dengan konteks pengujian model.

Pemilihan *dataset* dijelaskan sebagai keputusan yang kritis untuk memastikan model dapat menangani berbagai kondisi dan variasi objek. Penggunaan data dilakukan dengan penuh pertimbangan terhadap perwakilan dunia nyata dan tujuan deteksi objek. Oleh karena itu, mekanisme ini mencerminkan keseriusan peneliti dalam merancang eksperimen yang solid dan dapat diandalkan, menciptakan lan-

dasar yang kuat untuk evaluasi performa model YOLO yang diusulkan.

3.2.4. Pengukuran Keberhasilan Metode

Proses pengumpulan data pada penelitian ini didasarkan pada dua *dataset* utama, yaitu VOC 2007 dan VOC 2012, yang telah terbukti menjadi sumber daya yang berharga dalam pelatihan model deteksi objek. Tahap awal melibatkan penelitian terinci terkait struktur *dataset*, termasuk kelas objek yang relevan dan jumlah sampel yang tersedia. Langkah selanjutnya adalah mengimplementasikan teknik *sampling* yang cermat untuk memastikan distribusi objek yang seimbang dalam setiap subset *dataset*, yang penting untuk mencegah bias dalam model. Pengumpulan data melibatkan tahapan ekstraksi fitur dan anotasi manual yang dilakukan oleh ahli berpengalaman, memastikan kualitas dan akurasi *label* pada setiap objek.

Proses *preprocessing* data menjadi langkah penting dalam mengoptimalkan kualitas dataset. Normalisasi nilai piksel, *resize* gambar, dan teknik augmentasi data, seperti rotasi atau *flipping*, diterapkan untuk memperkaya variasi dan memastikan model dapat mengenali objek dalam berbagai konteks visual. Selain itu, penanganan *outlier* dan kejanggalan dalam *dataset* menjadi fokus, di mana metode statistik dan *visual* digunakan untuk mengidentifikasi dan memperbaiki anomali. Keberlanjutan dataset juga memerlukan upaya periodik untuk pembaruan, memasukkan data baru atau menghilangkan data yang tidak lagi relevan. Keseluruhan, pengumpulan dan pengelolaan data dalam penelitian ini tidak hanya mencerminkan ketelitian dalam mengidentifikasi objek, tetapi juga komitmen untuk menjaga integritas dataset sebagai fondasi eksperimen deteksi objek yang handal dan dapat diandalkan.

DAFTAR PUSTAKA

- [1] T. D. Dan Hendrycks, “Benchmarking neural network robustness to common corruptions and perturbations,” *arXiv preprint*, 2020.
- [2] Y.-W. T. C. L. Hao-Shu Fang, Shuqin Xie, “The devil is in the details: Delving into unbiased data processing for human pose estimation,” *arXiv preprint*, 2020.
- [3] S. K. Sashank J. Reddi, Satyen Kale, “On the convergence of adam and beyond,” *International Conference on Learning*, 2018.
- [4] t. Alexey Bochkovski, “Yolov4: Optimal speed and accuracy of object detection,” *arXiv preprint*, 2020.
- [5] P. D. R. G. Kaiming He, Georgia Gkioxari, “Mask r-cnn,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [6] R. G. K. H. B. H. Tsung-Yi Lin, Piotr Dollár, “Feature pyramid networks for object detection,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [7] D. E. C. S. S. R. C.-Y. F. A. C. B. Wei Liu, Dragomir Anguelov, “Ssd: Single shot multibox detector,” *European Conference on Computer Vision (ECCV)*, 2016.

- [8] R. G. K. H. P. D. Tsung-Yi Lin, Priya Goyal, “Focal loss for dense object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- [9] Q. V. L. Mingxing Tan, “Efficientdet: Scalable and efficient object detection,” *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [10] C. K. I. W. J. W.-A. Z. M. Everingham, L. Van Gool, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, 2010.