KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 1, Januari (2024), pp. 47-60

# Deteksi Objek Bahasa Isyarat Huruf Bisindo Menggunakan SSD-Mobilenet

*Gunawan Abdillah[1], Ridwan Ilyas[2]*
*[1,2]Jurusan Teknik Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani, Indonesia*
*E-mail: gna@unjani.ac.id[1], rdi@if.unjani.ac.id[2]*

***Abstract***

*Automatic sign language recognition systems help the hearing-impaired community communicate better. The pure sign language system developed by deaf Indonesians is called BISINDO. BISINDO is used by deaf friends based on their knowledge of their environment. SSD is an abbreviation for Single Shot MultiBox Detector, a method of detecting objects in images using a neural network in one stage. With SSD, objects of any size and shape can be identified easily and accurately without the need for object suggestions or complex resampling steps. This research uses SSD Mobile Net to identify bisindo sign language for the Letter category. The evaluation results show that the best model is SSD Mobilenet V1 FPN 640 x 640.*

***Keywords****: objec detection, sign language, bisindo, SSD Mobilenet*

***Abstrak***

*Sistem pengenalan bahasa isyarat otomatis membantu komunitas dengan gangguan pendengaran berkomunikasi dengan lebih baik. Sistem bahasa isyarat murni yang dikembangkan oleh tunarungu Indonesia disebut BISINDO. BISINDO digunakan oleh teman tuli berdasarkan pengetahuan mereka tentang lingkungan mereka. SSD adalah singkatan dari Single Shot MultiBox Detector, sebuah metode deteksi objek pada gambar menggunakan jaringan saraf dalam satu tahap. Dengan SSD, objek dari berbagai ukuran dan bentuk dapat diidentifikasi dengan mudah dan akurat tanpa memerlukan saran objek atau tahap resampling yang rumit. Penelitian ini menggunakan SSD Mobile Net untuk mengidentifikasi bahasa isyarat bisindo untuk kategori Huruf. Hasil evaluasi menunjukkan bahwa model terbaik adalah SSD Mobilenet V1 FPN 640 x 640.*

***Kata kunci****: deteksi objek, bahasa isyarat, bisindo, SSD Mobilenet.*

## 1. Pendahuluan

Bahasa isyarat memiliki peran sentral dalam memfasilitasi komunikasi efektif bagi komunitas dengan gangguan pendengaran. Pentingnya bahasa isyarat tidak hanya tercermin dalam kemampuannya untuk memungkinkan individu dengan gangguan pendengaran berpartisipasi secara penuh dalam interaksi sehari-hari, tetapi juga dalam menyediakan medium esensial untuk mereka menyampaikan pemikiran, perasaan, dan ide-ide mereka. Bahasa isyarat bukan sekadar alat komunikasi alternatif; ini adalah bagian integral dari identitas budaya komunitas tuli. Dengan memanfaatkan ekspresi visual dan gerakan tubuh, bahasa isyarat tidak hanya menawarkan sarana efektif bagi komunikasi verbal, tetapi juga merangsang kreativitas dan inovasi dalam menyampaikan makna.

Keberadaan bahasa isyarat memainkan peran penting dalam mempromosikan inklusivitas, meruntuhkan batasan komunikasi, dan memberdayakan individu dengan gangguan pendengaran untuk berpartisipasi sepenuhnya dalam berbagai aspek kehidupan sosial, akademis, dan profesional. Dengan memahami dan menghargai pentingnya bahasa isyarat, masyarakat dapat menciptakan lingkungan yang lebih inklusif dan menghormati keberagaman cara berkomunikasi.

ISSN: 2720-992X

Pentingnya penelitian dalam pengembangan sistem pengenalan bahasa isyarat secara otomatis tidak hanya tercermin dalam kemajuan teknologi, tetapi juga dalam dampak sosial yang signifikan. Sistem pengenalan bahasa isyarat otomatis membuka pintu bagi inklusivitas yang lebih luas bagi komunitas dengan gangguan pendengaran. Dengan adopsi teknologi ini, individu yang menggunakan bahasa isyarat sebagai bentuk komunikasi utama dapat lebih mudah terlibat dalam berbagai aspek kehidupan, mulai dari pendidikan hingga pekerjaan. Selain meningkatkan aksesibilitas, sistem ini juga memiliki potensi untuk mengurangi kesenjangan komunikatif antara komunitas tuli dan pendengar. Melalui penelitian ini, kita dapat mencapai tingkat akurasi dan efisiensi yang lebih tinggi dalam pengenalan bahasa isyarat, membuka peluang untuk mengintegrasikan teknologi ini ke dalam perangkat sehari-hari seperti smartphone atau perangkat pintar lainnya.

Pemilihan metode SSD MobileNet dalam pengembangan model deep learning untuk mendeteksi bahasa isyarat pada gambar didasarkan pada beberapa alasan yang mendasar. Pertama, arsitektur SSD (Single Shot Multibox Detector) memungkinkan deteksi objek yang cepat dan akurat dengan menggunakan lapisan konvolusi yang dilakukan hanya sekali pada seluruh gambar *[1]*. Ini menjadikan SSD MobileNet sangat efisien, ideal untuk aplikasi deteksi objek pada gambar secara real-time, seperti dalam kasus bahasa isyarat yang seringkali melibatkan gerakan cepat.

Kedua, MobileNet merupakan arsitektur jaringan saraf yang dioptimalkan khusus untuk perangkat mobile. Kelebihannya terletak pada kemampuan untuk menjalankan model deep learning dengan kinerja tinggi pada perangkat dengan sumber daya terbatase. Ini menjadikan SSD MobileNet pilihan yang tepat untuk membuat model yang dapat diintegrasikan secara efektif pada perangkat mobile, memungkinkan pengguna untuk mengakses teknologi deteksi bahasa isyarat dengan lebih mudah dan praktis.

Selain itu, SSD MobileNet juga dikenal karena kemampuannya dalam menangani objek dengan berbagai ukuran dan skala, yang relevan ketika berhadapan dengan variasi ukuran tangan dan gerakan dalam bahasa isyarat. Hal ini meningkatkan ketangguhan model terhadap variasi dalam pengambilan gambar bahasa isyarat, menjadikannya solusi yang lebih adaptif dan responsif terhadap kondisi dunia nyata.

## 2. Metodologi Penelitian

Bahasa isyarat adalah bahasa yang tidak menggunakan bunyi ucapan manusia atau tulisan dalam penerapannya [2]. Bahasa isyarat digunakan oleh orang-orang yang memiliki keterbatasan dalam berbicara, seperti penyandang tunarungu, untuk berkomunikasi dengan orang lain. Bahasa isyarat termasuk bahasa yang unik, karena berbeda di setiap negara. Di Indonesia, terdapat dua kategori perkembangan bahasa isyarat yaitu, bahasa isyarat SIBI (Sistem Isyarat Bahasa Indonesia) dan BISINDO (Bahasa Isyarat Indonesia).

BISINDO adalah sistem bahasa isyarat murni yang dikembangkan oleh tunarungu Indonesia. BISINDO digunakan oleh teman tuli sesuai dengan pemahaman mereka dengan lingkungan sekitar. BISINDO ini merupakan isyarat untuk teman tuli yang posisinya lebih tua dari SIBI. Karakteristik BISINDO ketika digunakan sebagai bahasa isyarat yakni memunculkan ekspresi wajah dan mulut. Selain itu, ada lima parameter yang biasa digunakan, yakni lokasi, bentuk tangan, orientasi, gerak tangan, dan ekspresi non-manual.

BISINDO dibentuk dan berkembang secara alami dalam lingkungan tunarungu. Bahasa ini berkembang sesuai dengan pemahaman tunarungu dari berbagai latar belakang. Sehingga bahasa ini menjadi bahasa yang sangat awal bagi mereka sampai disebut juga dengan bahasa ibu bagi tunarungu. BISINDO tidak menggunakan struktur imbuhan bahasa Indonesia layaknya SIBI.

Deteksi objek adalah proses untuk menentukan lokasi dan kategori objek dalam sebuah gambar atau video [3]. Tujuan dari deteksi objek adalah untuk mengidentifikasi objek

yang ada dalam gambar atau video secara otomatis, sehingga dapat digunakan untuk berbagai aplikasi seperti pengawasan keamanan, pengenalan wajah, kendaraan otonom, dan lain sebagainya. Dalam deteksi objek, model deep learning digunakan untuk mempelajari fitur-fitur visual dari objek dan memprediksi lokasi serta kategori objek dalam gambar atau video.

Masalah deteksi objek adalah menentukan di mana objek berada dalam gambar yang diberikan (lokalisasi objek) dan kategori apa yang dimiliki oleh setiap objek (klasifikasi objek). Oleh karena itu, pipeline model deteksi objek tradisional dapat dibagi menjadi tiga tahap utama: *Informative Region Selection*, *Feature Extraction*, dan klasifikasi.

*Informative Region Selection* adalah tahap pertama dalam pipeline tradisional object detection model. Pada tahap ini, dilakukan pemilihan wilayah informatif dalam gambar yang kemungkinan mengandung objek. Pemilihan wilayah ini dilakukan untuk mengurangi jumlah wilayah yang harus dianalisis pada tahap berikutnya, sehingga dapat menghemat waktu dan sumber daya komputasi. Pada pendekatan tradisional, pemilihan wilayah informatif dilakukan dengan menggunakan metode sliding window, yaitu dengan memindahkan jendela persegi panjang dengan ukuran yang sama pada seluruh gambar. Namun, metode ini memiliki kelemahan karena menghasilkan banyak wilayah yang tidak relevan dan memakan waktu yang lama.

*Feature Extraction* adalah tahap kedua dalam pipeline tradisional object detection model. Pada tahap ini, dilakukan ekstraksi fitur dari wilayah informatif yang telah dipilih pada tahap sebelumnya. Fitur-fitur ini digunakan untuk menghasilkan representasi yang lebih sederhana dan mudah diolah dari gambar. Pada pendekatan tradisional, fitur-fitur ini dihasilkan dengan menggunakan metode ekstraksi fitur manual seperti *Histogram of Oriented Gradients* (HOG) [4] atau *Local Binary Patterns* (LBP). Metode HOG mengukur distribusi gradien orientasi pada wilayah gambar, sedangkan metode LBP mengukur pola tekstur pada wilayah gambar.

Terdapat beberapa pendekatan dalam melakukan deteksi objek, di antaranya:

1. Pendekatan berbasis region proposal: Pendekatan ini memerlukan tahap awal untuk menghasilkan daftar proposal lokasi objek dalam gambar, kemudian dilakukan klasifikasi pada setiap proposal untuk menentukan apakah proposal tersebut berisi objek atau bukan. Contoh dari pendekatan ini adalah R-CNN [5].
2. Pendekatan berbasis single-shot: Pendekatan ini melakukan deteksi objek dengan satu kali proses feedforward pada jaringan neural. Pendekatan ini lebih cepat daripada pendekatan berbasis region proposal. Contoh dari pendekatan ini adalah YOLO [6].
3. Pendekatan berbasis two-stage: Pendekatan ini memerlukan dua tahap, yaitu tahap proposal dan tahap klasifikasi. Pada tahap proposal, daftar proposal lokasi objek dihasilkan, kemudian pada tahap klasifikasi, setiap proposal diklasifikasikan untuk menentukan apakah proposal tersebut berisi objek atau bukan. Contoh dari pendekatan ini adalah Faster R-CNN [7].

Penelitian terkait deteksi objek tidak dapat dipisahkan dari perkembangan metode CNN sebagai dasar dari pengolahan citra dengan pendekatan *Deel Learning*. Pengembangan CCNN dimulai pada periode *Origin* (akhir 1980-an hingga 1990-an), CNN pertama yang populer adalah LeNet-5 yang dikembangkan pada tahun 1998[8] . CNN pada saat itu masih dalam tahap pengembangan dan belum banyak digunakan. Tujuan utama dari pengembangan CNN pada periode ini adalah untuk mendeteksi pola pada gambar dan mengenali karakter optik . CNN pada periode ini masih memiliki keterbatasan dalam hal pemrosesan waktu yang tinggi dan belum sepenuhnya dipahami cara kerjanya.

CCN selanjutnya berkembang pada periode *Stagnation* (awal 2000-an), pengembangan CNN mengalami hambatan karena belum sepenuhnya dipahami cara kerjanya dan belum ada dataset gambar yang beragam seperti Google's Open Images atau Microsoft's COCO.

Selain itu, CNN pada periode ini masih terbatas pada pengenalan karakter optik (OCR) dan membutuhkan waktu komputasi yang tinggi, sehingga meningkatkan biaya operasional. Pada periode ini, model pembelajaran mesin lain seperti Support Vector Machine (SVM) menunjukkan hasil yang lebih baik daripada CNN.

Pada periode Revival (2006-2011), pengembangan CNN mengalami kemajuan signifikan. Dalam sebuah penelitian menunjukkan bahwa menggunakan algoritma max-pooling untuk ekstraksi fitur daripada algoritma sub-sampling yang digunakan sebelumnya menghasilkan peningkatan yang signifikan [9]. Selain itu, pada periode ini, dataset gambar yang lebih beragam mulai tersedia, seperti ImageNet, yang memungkinkan pengembangan CNN untuk deteksi objek yang lebih baik. Pada periode ini, CNN mulai digunakan secara luas dalam berbagai aplikasi, termasuk pengenalan wajah, deteksi objek, dan pengenalan suara.

Pada periode Rise (2012-2013), pengembangan CNN mengalami kemajuan yang signifikan dalam hal akurasi dan kecepatan pemrosesan. Pada tahun 2012, AlexNet, sebuah arsitektur CNN yang sangat dalam, memenangkan kompetisi ImageNet dengan margin yang besar, menunjukkan bahwa CNN dapat mengungguli model pembelajaran mesin lainnya dalam tugas pengenalan gambar [10]. Selain itu, pada periode ini, penggunaan unit aktivasi ReLU (Rectified Linear Unit) dalam CNN menjadi populer, yang memungkinkan CNN untuk beroperasi lebih cepat dan menghasilkan hasil yang lebih baik. Pada periode ini, CNN mulai digunakan secara luas dalam berbagai aplikasi, termasuk pengenalan wajah, deteksi objek, dan pengenalan suara.

Pada periode Architectural Innovations (2014-2020), terjadi banyak inovasi dalam arsitektur CNN. Salah satu arsitektur yang terkenal dan banyak digunakan adalah VGG, yang dikembangkan pada tahun 2014 [11]. Arsitektur VGG memiliki banyak lapisan dan telah terbukti sangat efektif dalam tugas-tugas pengenalan gambar. Selain itu, pada periode ini, arsitektur CNN yang lebih kompleks dan lebih dalam mulai dikembangkan, seperti ResNet, Inception, dan DenseNet, yang semuanya memiliki ratusan atau bahkan ribuan lapisan [12]. Pada periode ini, CNN juga mulai digunakan dalam berbagai aplikasi baru, seperti mobil otonom, pengenalan bahasa alami, dan pengenalan tulisan tangan.

### 2.1. Metode Eksperimen

Penelitian ini dimulai dengan tahap pertama, yaitu pengambilan gambar menggunakan kamera. Proses ini dilakukan dengan untuk mendapatkan dataset yang representatif dan mencakup variasi bahasa isyarat huruf Bisindo. Kualitas gambar sangat diperhatikan agar dapat mendukung keakuratan proses deteksi objek pada tahap selanjutnya.

Tahap kedua setelah mendapatkan dataset gambar, langkah berikutnya adalah anotasi gambar dengan menandai area bahasa isyarat pada setiap gambar. Anotasi ini menjadi langkah krusial yang nantinya menjadi dasar untuk melatih model deteksi objek, karena memastikan bahwa model dapat mengenali dan memahami posisi serta konteks dari bahasa isyarat huruf Bisindo dalam berbagai situasi dan kondisi.

Tahap ketiga adalah proses pemodelan dilakukan dengan menggunakan metode SSD-Mobilenet V1 dan V2. Kedua model tersebut dipilih karena kombinasi kecepatan dan akurasi deteksi objeknya. SSD (Single Shot Multibox Detector) digunakan untuk mendeteksi objek pada gambar secara real-time, sementara Mobilenet V1 dan V2 memberikan efisiensi komputasi yang diperlukan untuk aplikasi deteksi objek pada perangkat dengan sumber daya terbatas.

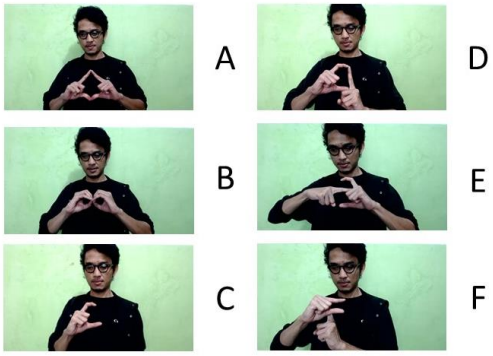
Setelah pemodelan selesai, dilanjutkan tahap terkahir evaluasi data dilakukan untuk mengukur performa model. Metrik evaluasi yang digunakan meliputi intersection over union (IoU), precision, recall, dan mean Average Precision (mAP). IoU memberikan informasi sejauh mana hasil deteksi model bersesuaian dengan anotasi yang sebenarnya, sedangkan precision dan recall memberikan gambaran keakuratan dan kelengkapan deteksi. mAP kemudian memberikan nilai keseluruhan dari performa model dalam

mengenali bahasa isyarat huruf Bisindo, menjadi tolak ukur utama dalam mengevaluasi efektivitas model yang telah dikembangkan.

### 2.2. Pengambilan Gambar

Prose pengambilan gambar dilakukan dibantu dengan peraga bahasa isyarat untuk semua huruf kecuali huruf J dan R. Kedua huruf tersebut tidak dikutkan dalam pengambilan gambar karena dalam standar BISINDO, kedua hutuf tersebut adalah sebuah gerakah tangan. Pada penelitian ini pendekatan yang digunakan adalah deteksi objek dengan batasan single frame (tidak bisa membaca gerakan).

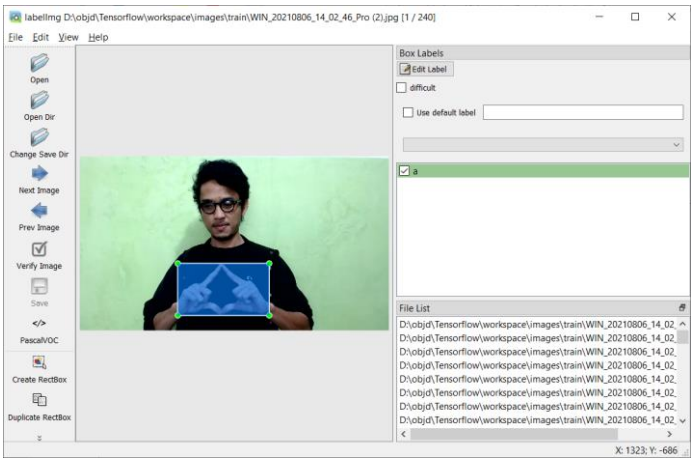
Gambar yang diambil untuk setiap class adalah 15 gambar dimana, 10 gambar nantinya dipakai untuk data latih, sedangkan 5 gambar dipakai untuk data uji. Dengan jumlah huruf sebanyak 24, maka gambar yang dikumpulkan adalah 360 gambar. Peraga Menggunakan baju berwarna hitam dan latar belakang foto berwarna hijau.




**Gambar 1.** Contoh gambar dan Label

### 2.3. Anotasi Data

Proses anotasi data deteksi objek digunakan perangkat lunak LabelImg [1]. LabelImg, sebuah alat annotasi gambar yang banyak dipakai oleh Tzutalin dan melibatkan kontribusi dari puluhan kontributor. Seiring berjalannya waktu, LabelImg secara resmi menjadi bagian dari komunitas Label Studio. Label Studio, yang kini menjadi pemiliki bagi LabelImg, menawarkan solusi pelabelan data yang sangat fleksibel. Ini tidak hanya mencakup gambar, tetapi juga teks, hiperlink, audio, video, dan data deret waktu. Dengan kata lain, Label Studio memberikan keandalan dan keunggulan sebagai alat pelabelan data opensource yang cukup komprehensif.




**Gambar 2.** Contoh Anotasi Data Huruf A

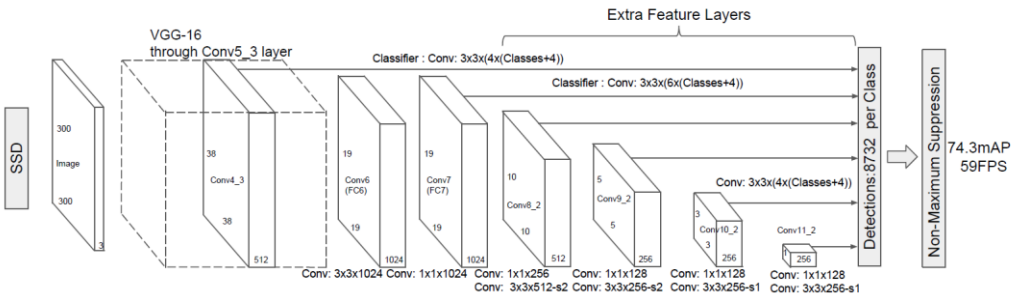
[1] https://github.com/HumanSignal/labelImg

Seluruh data gambar yang diambil, dianotasi seuai dengan labelnya masing-masing. Prosesnya dimulai dengan memilih folder tempat semua gambar. Selanjutnya annotator menandari area tangan yang menjadi area utama isyarat, lalu membutuhkan label class yang sesuai. Format data yang dipakai adalah PascalVOC *[13]*. Setelah dilakukan anotasi, selanjutnya data dipisahkan antara data latih dan data uji secara acak.

### 2.4. Pemodelan Dengan SSD Mobile Net

SSD adalah singkatan dari Single Shot MultiBox Detector, yaitu sebuah metode deteksi objek pada gambar menggunakan jaringan saraf dalam satu tahap *[1]*. Dengan SSD, dapat dengan mudah dan akurat mengidentifikasi objek dari berbagai ukuran dan bentuk, tanpa perlu proposal objek atau tahap resampling yang kompleks. SSD adalah salah satu teknik terbaru dalam bidang computer vision dan telah menunjukkan kinerja yang sangat baik dalam berbagai tugas deteksi objek.

SSD (Single Shot MultiBox Detector) bekerja dengan menggunakan jaringan saraf konvolusi untuk memproses gambar dan menghasilkan prediksi lokasi dan kelas objek pada gambar. Cara kerja SSD secara umum dimulai dengan ekstraksi fitur. SSD menggunakan jaringan saraf konvolusi untuk mengekstraksi fitur dari gambar input. Setiap layer pada jaringan saraf konvolusi menghasilkan fitur dengan resolusi yang berbeda. Langkah kedaua adalah prediksi lokasi dan kelas objek. SSD menggunakan beberapa layer konvolusi terakhir untuk memprediksi lokasi dan kelas objek pada gambar. Setiap layer konvolusi terakhir menghasilkan beberapa kotak pembatas (bounding box) dan skor kelas untuk setiap kotak pembatas. Langkah ketiga Non-maximum suppression. SSD menggunakan teknik non-maximum suppression untuk menghilangkan kotak pembatas yang tumpang tindih dan memilih kotak pembatas dengan skor kelas tertinggi sebagai hasil deteksi. Langkah terakhir Post-processing. SSD mel.kukan post-processing pada kotak pembatas yang dipilih untuk meningkatkan akurasi deteksi. Post-processing termasuk regresi kotak pembatas untuk meningkatkan presisi lokasi objek dan filtering kotak pembatas yang tidak memenuhi kriteria tertentu. Dengan cara kerja ini, SSD dapat menghasilkan deteksi objek yang akurat dan cepat dengan menggunakan satu jaringan saraf dalam satu tahap.




**Gambar 3.** Aksitektur SSD Mobilenet [1]

Arsitektur SSD dimulai dengan lapisan konvolusi 3x3, yang bertanggung jawab untuk mengekstraksi fitur-fitur lokal pada gambar input. Proses ini diikuti oleh lapisan konvolusi 3x3 Depthwise, yang memperdalam pemahaman terhadap fitur-fitur tersebut dengan melakukan konvolusi terhadap setiap saluran input secara terpisah. Selanjutnya, dilakukan konvolusi 1x1 untuk menggabungkan informasi dari seluruh saluran dan mereduksi dimensi. Proses ini diulang dengan konvolusi 3x3 Depthwise, diikuti oleh konvolusi 1x1, untuk terus meningkatkan kompleksitas representasi fitur.

Rangkaian konvolusi 3x3 Depthwise dan 1x1 terus dilakukan secara berulang, memungkinkan model untuk mengekstraksi hierarki fitur yang semakin kompleks. Proses ini memberikan fleksibilitas yang diperlukan untuk mengenali pola-pola yang lebih

abstrak dan kontekstual dalam data citra. Terakhir, setelah serangkaian konvolusi, dilakukan lapisan Average Pooling untuk merata-ratakan hasil dari konvolusi-konvolusi sebelumnya, menghasilkan representasi fitur yang global. Selanjutnya, Fully Connected layer digunakan untuk menyatukan informasi dari seluruh fitur dan Softmax diaplikasikan pada lapisan terakhir untuk menghasilkan probabilitas kelas output. Dengan demikian, arsitektur ini memberikan pendekatan yang efektif dalam ekstraksi fitur pada citra, memungkinkan pengenalan pola dengan tingkat akurasi yang tinggi.

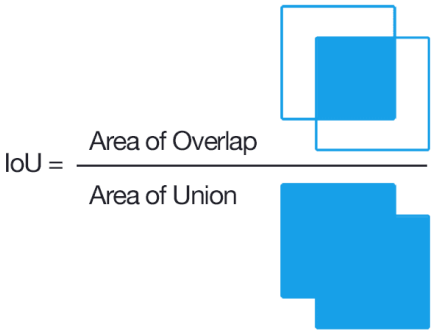
SSD memiliki beberapa keunggulan dibandingkan dengan metode deteksi objek lainnya, antara lain:

1. Kecepatan: SSD lebih cepat daripada metode deteksi objek lainnya seperti Faster R-CNN dan YOLO. Hal ini karena SSD tidak memerlukan tahap proposal objek dan tahap resampling yang kompleks.
2. Akurasi: SSD memiliki akurasi yang sangat baik dalam mendeteksi objek, bahkan sebanding dengan teknik yang lebih lambat yang melakukan proposal objek dan pooling eksplisit.
3. Kemudahan pelatihan: SSD mudah dilatih dan memiliki arsitektur yang sederhana. Hal ini memungkinkan pengguna untuk dengan mudah melatih model deteksi objek mereka sendiri.
4. Kemampuan mengatasi objek dengan ukuran dan bentuk yang berbeda: SSD dapat dengan mudah mengatasi objek dengan ukuran dan bentuk yang berbeda karena menghasilkan prediksi dari berbagai skala dan aspek rasio.
5. Integrasi yang mudah: SSD mudah diintegrasikan ke dalam sistem yang memerlukan komponen deteksi objek.

Dalam penelitian ini, dilakukan variasi model deteksi objek untuk mengoptimalkan kinerja sistem. Tiga model yang diuji adalah SSD MobileNet V1 FPN 640x640 *[14]*, SSD MobileNet V2 FPNLite 320x320, dan SSD MobileNet V2 FPNLite 640x640. Model pertama, SSD MobileNet V1 FPN 640x640, dipilih karena memiliki kemampuan untuk menangkap fitur-fitur kompleks pada gambar dengan resolusi tinggi, yang sesuai dengan kebutuhan deteksi objek pada dataset bahasa isyarat huruf Bisindo. Sementara itu, model kedua, SSD MobileNet V2 FPNLite 320x320, memberikan fokus pada efisiensi komputasi, memungkinkan deteksi objek yang cepat tanpa mengorbankan akurasi. Model ketiga, SSD MobileNet V2 FPNLite 640x640, merupakan kombinasi dari keunggulan kedua model sebelumnya, dengan resolusi tinggi dan efisiensi komputasi yang seimbang. Dengan menguji ketiga model ini, penelitian ini bertujuan untuk mengevaluasi kinerja masing-masing model dalam konteks deteksi objek bahasa isyarat huruf Bisindo, mempertimbangkan trade-off antara akurasi dan kecepatan komputasi.

### 2.5. Metode Evaluasi

Intersection over Union adalah pengukuran yang digunakan untuk mengukur seberapa banyak area yang tumpang tindih antara bounding box hasil prediksi dengan bounding box yang sebenarnya (ground-truth) [15]. IOU dihitung dengan membagi luas area tumpang tindih antara kedua bounding box dengan luas area gabungan dari keduanya. IOU biasanya digunakan untuk menentukan apakah suatu deteksi dianggap benar atau salah dalam evaluasi algoritma deteksi objek. Semakin tinggi nilai IOU, semakin baik deteksi objek yang dilakukan.




**Gambar 4.** Interaction over Union

Precision adalah kemampuan suatu model untuk mengidentifikasi objek yang relevan atau penting saja [15]. Precision dihitung dengan membagi jumlah deteksi objek yang benar positif dengan jumlah total deteksi positif yang dilakukan oleh model. Precision mengukur seberapa akurat model dalam mengklasifikasikan objek sebagai positif, yaitu objek yang relevan atau penting. Semakin tinggi nilai precision, semakin sedikit objek yang salah diklasifikasikan sebagai positif.

$$P = \frac{TP}{TP+FP} = \frac{TP}{all\ detection} \quad (1)$$

Recall adalah kemampuan suatu model untuk menemukan semua objek yang relevan atau penting [15]. Recall dihitung dengan membagi jumlah deteksi objek yang benar positif dengan jumlah total objek yang sebenarnya ada dalam gambar atau video. Recall mengukur seberapa baik model dalam menemukan semua objek yang relevan atau penting, tanpa memperhatikan apakah model mengklasifikasikan objek tersebut dengan benar atau salah. Semakin tinggi nilai recall, semakin banyak objek yang berhasil ditemukan oleh model.

$$R = \frac{TP}{TP+FN} = \frac{TP}{all\ ground\ truths} \quad (2)$$

mAP atau mean Average Precision adalah metrik evaluasi yang digunakan untuk mengukur akurasi suatu model deteksi objek pada seluruh kelas objek yang ada dalam dataset [15]. mAP dihitung dengan mengambil rata-rata dari nilai AP (Average Precision) pada setiap kelas objek. AP sendiri dihitung dengan menghitung luas area di bawah kurva Precision-Recall (P-R) pada setiap kelas objek. mAP memberikan gambaran keseluruhan tentang kinerja model deteksi objek pada seluruh kelas objek yang ada dalam dataset. Semakin tinggi nilai mAP, semakin baik kinerja model deteksi objek tersebut. mAP sering digunakan sebagai metrik evaluasi pada kompetisi deteksi objek.

$$mAP = \frac{1}{N}\sum_{i=i}^{N} AP_i \quad (3)$$

## 3. Hasil Dan Pembahasan

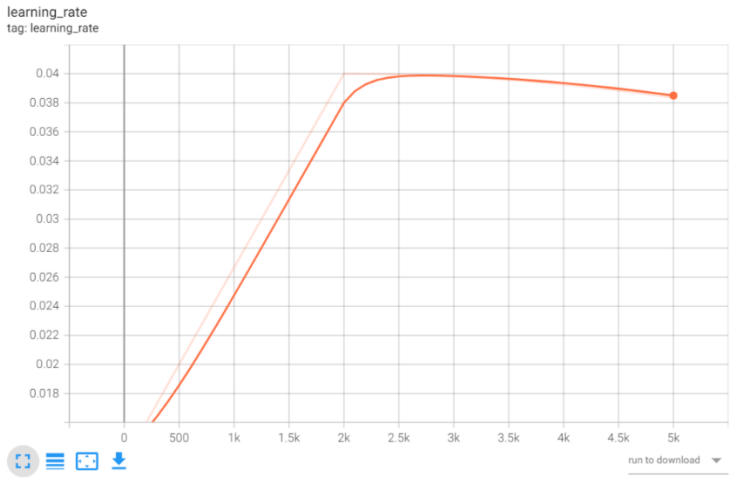
Eksperimen pada penelitian ini dimulai dari proses training data lalu mengevaluasi model yang dihasilkan. Pada bagian di atas telah dijelaskan mengenai data dan model yang dipakai. Di bagian ini, dijelaskan hasil dari proses pelatihan data dan evaluasi model. Evaluasi dilakukan terhadap semua model untuk mendapatkan model terbaik. Setelah itu, model terbaik akan dilihat kemampuannya dalam melakukan klasifikasi setiap class.

Proses pelatihan data menjadi model digunakan libary bantuan Tensorflow Object Detection dengan bahasa pemrograman python *[16]*. Proses pelatihan menggunakan GPU agar waktu yang digunakan menjadi lebih cepat. Dengan berbagai variasi model dan jumlah iterasi makan akan mendapatkan berbagai variasi hasil yang dianalisi pada bagian ini.

Aspek yang diperkatikan dalam eksperimen antar lain adalah model, besaran gambar dan epoch. Model yang dibandingkan adalah SSD MobileNet V1 FPN, SSD MobileNet V2 FPNLite. Besar gambar yang dibandingkan adalah 320x320 dan 640x640. Sedangkan jumlah epoch, atau iterasi pelatihan dibandingkan adalah 1000, 2000, 3000, 4000 dan 5000.

### 3.1. Proses Training

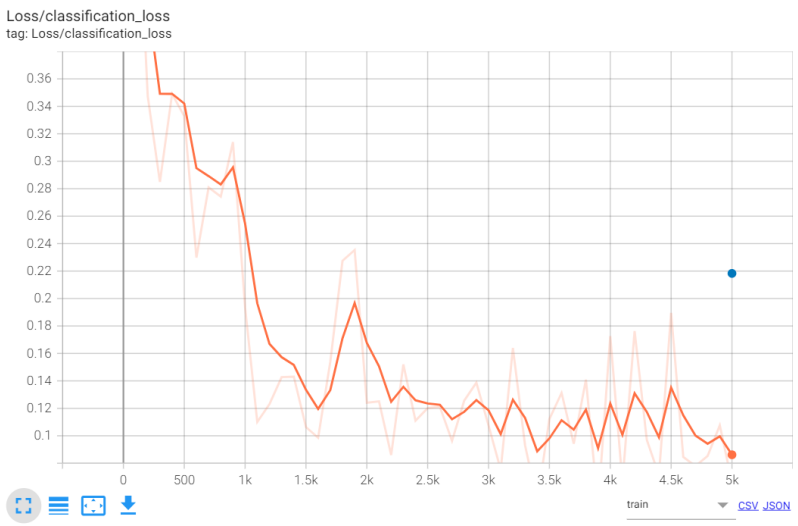
Model masin learning dengan neural network, memiliki komponen learning rate yang bekerja untuk menentukan laju percepatan. Dalam penelitian ini, learning rate disetting secara adaptif sehingga nilainya berubah pada setiap iterasi. Proses ini agar proses pelatihan berjalan dengan baik.




**Gambar 5.** Loss Classification

Gambar 5, menunjukan prose perubahan learning rate untuk setiap iterasinya dari iterasi awal hingga iterasi ke 5000. Dari grafik dapat dibaca bahwa learning rate terus meningkat nilainya hingga di kisaran 0.04 pada iterasi ke 2500. Selanjutnya learning rate cenderung stabil, meskipun menurut tapi tidak terlalu drastis.

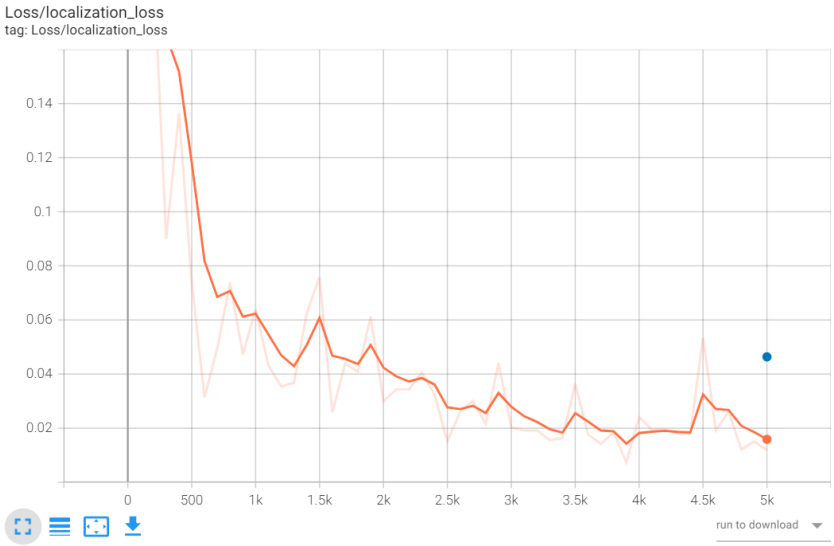
Dalam proses pelatihan, dapat diperhatikan pergerakan dari nilai loss. Nilai loss yang pertama dilihat dalah loss classification. Loss ini mengukur seberapa baik model mengklasifikasikan objek yang terdapat dalam satu gambar. Dalam implementasinya, loss classification mengukur perbedaan antara kelas yang diprediksi oleh model dan kelas yang sebenarnya dari objek dalam bounding box. Tujuan dari mengevaluasi loss ini adalah untuk mengoptimalkan prediksi setiap kelas.




**Gambar 6.** Loss Classification

Pada gambar di atas, dapat dilihat bahwa loss classification semakin lama-semakin menurun. Penurunan paling besar dari mulai awal iterasi hingga iterasi ke 1500. Setelah iterasi tersebut, loss ini cenderung naik turun namun dengan progresi yang tetap membaik. Hal ini indikatar bahwa proses mengoptimalkan klasifikasi setiap class berjalan dengan baik.

Dalam proses training juga dilihat loss localization. Loss ini mengukur seberapa baik model menentukan posisi bounding box untuk setiap objek. Model deteksi objek menghasilkan bounding box yang terdeteksi. Yang diukur oleh komponen ini adalah perbedaan bounding box yang diprediksi oleh model dengan koordinat sebenarnya. Tujuan dari loss ini untuk mengoptimalkan prediksi lokalisasi agar bounding box yagn dihasilkan semakin mendeteksi sebenarnya.

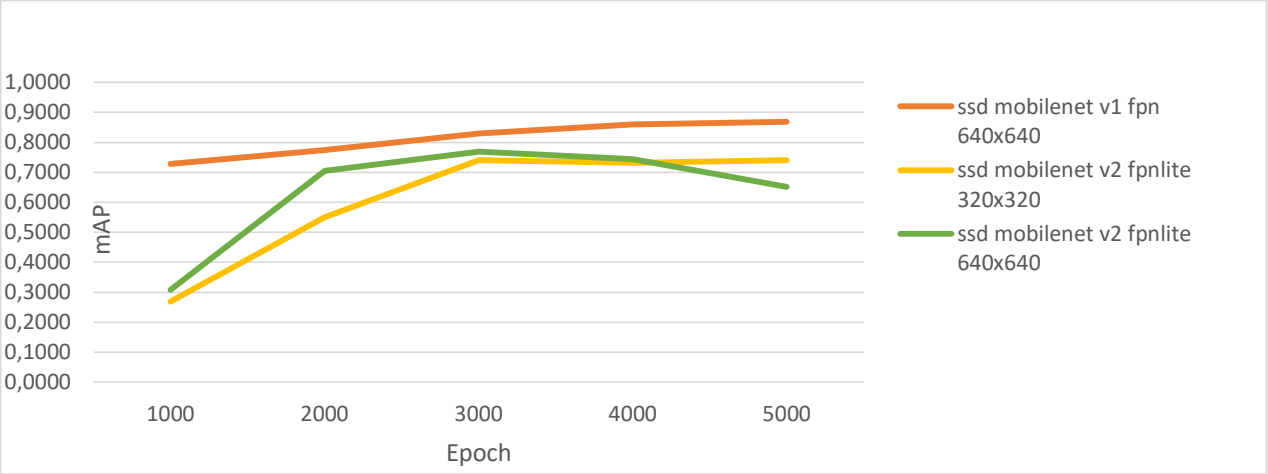



**Gambar 7.** Loss Regularization

Pada gambar di atas, dapat dilihat bahwa loss regularization terus menurus seiring dengan jumlah iterasi pelatihan. Dari mulai iterasi pertama hingga iterasi ke 5000, nilai loss terus menurun, meskipun pada iterai akhir tidak terlalu ekstrim menurun. Gambaran ini menunjukan proses training berjalan dengan baik.

### 3.2. Perbandingan Performa Model

SSD Mobilenet sebagai model utama yang dipakai, digunakan dengan beberapa variasi. Variasi pertama adalah versi 1 dan versi 2. Selain itu, digunakan variasi bentuk arsitektur *Feature Pyramid Networks* (FPN) dan FPNlite. Perfromasi untuk model, variasi dan area gambar yang dipakai, dilatih dalam semua batasan epoch untuk mendapatkan model, variasi dan epoch terbaik.




**Gambar 8.** Perbanding mAP Setiap Model

Dari Gambar 8, dapat di lihat bahwa model dengan performa terbaik adalah SSD Mobilnet V1 PFN 640x640. Performa model ini stabil menjadi yang paling baik dalam seluruh epoch pelatihan. Model ini lebih mengguli model lain terlihat dari penggunaan FPN dibandingkan dengan FPNlite. Indikasi ini karena model lain yang menggunakan area yang sama 640x640, tidak menunjukan hasul yang konsisten. Begitupun juga dengan versi yang digunakan, meskipin versi1, tapi hasilnya masih lebih baik dari versi 2.

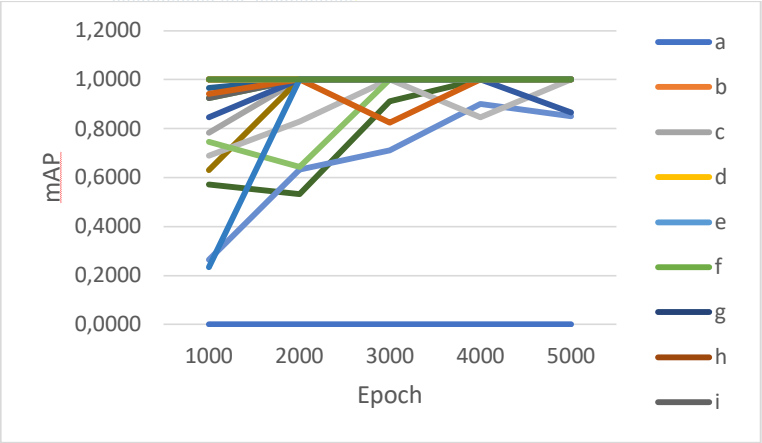
Pada tabel di bawah ini, dapat dilihat nilai performa mAP untuk setiap model dalam semua epoch. Model SSD Mobilenet v1 FPN 640x640 memiliki skor terbaik untuk setiap epochnya. Pada setiap iterasi pun stabil terus naik nilainya. Dari angka-angka ini dapat disimpulkan bahwa model ini adalah yang terbaik di bandingkan dengan yang lainnya

**Tabel 1.** Perbandingan mAP setiap model pada setiap iterasi

| Model | Epoch | | | | |
|---|---|---|---|---|---|
| | 1000 | 2000 | 3000 | 4000 | 5000 |
| Ssd mobilenet v1 fpn 640x640 | 0.7280 | 0.7745 | 0.8293 | 0.8597 | 0.8687 |
| Ssd mobilenet v2 fpnlite 320x320 | 0.2687 | 0.5500 | 0.7397 | 0.7308 | 0.7407 |
| Ssd mobilenet v2 fpnlite 640x640 | 0.3088 | 0.7055 | 0.769 | 0.7448 | 0.6520 |

### 3.3. Performa Klasifikasi

Ssd Mobilenet v1 FPN 640x640 terpilih sebagai model dengan variasi terbaik, selanjutkan akan dianalisis performasi model ini lebih dalam untuk setiap *class*. Analisis data untuk setiap class dilakukan pada model SSD Mobilenet V1 FPN 640 x 640 pada keseluruhan epoch.




**Gambar 9.** Evaluasi Setiap Class

Pada Gambar 9, dapat dilihat bahwa performa model klasifikasi setiap class cenderung memiliki kenaikan pada setiap epochnya. Performas terbaik terdapat pada class huruf F sedang performas terburuk terdapat pada performa huruf A. Klasifikasi pada class huruf F dari awal epoch hingga epoch terbanyak, nilainya cenderung stabil pada performa yang baik. Sedangkan untuk klasifkasi pada class huruf A, dari awal hingga epoch terbanyak pun nilainya tidak cenderung membaik.

**Tabel 2. Hasil Evaluasi mAP setiap Class**

| CLASS | Average of AP@0.5IOU |
|---|---|
| a | 0.0004 |
| b | 1.0000 |
| c | 0.9565 |
| d | 1.0000 |
| e | 1.0000 |
| f | 0.9850 |
| g | 1.0000 |
| h | 1.0000 |
| i | 0.9850 |
| k | 0.9263 |
| l | 0.9933 |
| m | 0.8032 |
| n | 0.6718 |
| o | 1.0000 |
| p | 0.8726 |
| q | 1.0000 |
| s | 1.0000 |
| t | 0.8780 |
| u | 0.9423 |
| v | 0.9537 |
| w | 1.0000 |
| x | 1.0000 |
| y | 0.8470 |
| X | 1.0000 |

Pada Tabel 2, terlihat performa pengukuran untuk setiap bahasa isyarat pada model terbaik. Dari data tersebut terlihat bahwa class Huruf A merupakan kelas yang memiliki nilai performa paling buruk di banding dengan yang lain. Karena data tersebut, maka perlu dilihat data evaluasi untuk class tersebut.

Data Eval Class A (mudah diprediksi) Data Eval Class A (sulid diprediksi)

**Gambar 10.** Perbandingan Gambar yang Mudah dan Sulit Diprediksi

Dengan melihat data uji pada class huruf A, maka dapat dilihat perbedaan data yang mudah diprediksi dengan data yang sulit diprediksi. Data yang mudah diprediksi memiliki posisi yang lebih umum yaitu pada di depan dada peraga bahasa isyarat. Sedangkan untuk data yang sulit diprediksi, posisinya berada di depan muka. Selain dari posisi, variasi dari latar pun mempengaruhi kemampuan untuk melakukan klasifikasi area. Area yang latarnya seragam, cenderung lebih mudah diprediksi dibanding dengan latar yang berkontur.

## 4. Kesimpulan

Penelitian ini telah menerapkan metode SSD Mobile Net untuk mendeteksi bahasa isyarat bisindo untuk kategori Huruf. Model terbaik yang ditunjukan dari hasil evaluasi adalah SSD Mobilenet V1 FPN 640 x 640. Model tersebut mengungkuli model lainnya pada epoch 5000. Pada analisis kemampuan mendeteksi bahasa isyarat setiap kelas, masih ditemukan kelemahan pada pendeteksian data dengan variasi warna dan kontur latar belakang.

Pengembangan selanjutnya dapat dilakukan dengan menambah jumlah dataset yang digunakan. Selain itu perlu diperhatikan bentuk dari dataset yang tingkat keseragamannya tidak terlalu jauh namun tetap variatif. Model deteksi objek yang lain juga dapat digunakan sebagai pembanding guna mendapatkan hasil yang lebih baik.

## Ucapan Terima Kasih

Penelitian ini mendapatkan pendaan dari program hibah Riset Dikti aktegori Penelitian Dosen Pemula periode tahun 2023. Penelitian ini juga dilakukan atas kerjasama dengan PT. Inovasi Disabilitas Indonesia.

## Daftar Pustaka

[1] W. Liu *Dkk.*, “Ssd: Single Shot Multibox Detector,” Vol. 9905, 2016, Hlm. 21–37. Doi: 10.1007/978-3-319-46448-0_2.

[2] A. S. Nugraheni, A. P. Husain, Dan H. Unayah, “Optimalisasi Penggunaan Bahasa Isyarat Dengan Sibi Dan Bisindo Pada Mahasiswa Difabel Tunarungu Di Prodi Pgmi Uin Sunan Kalijaga,” *Jht*, Vol. 5, No. 1, Hlm. 28, Feb 2023, Doi: 10.24853/Holistika.5.1.28-33.

[3] Z.-Q. Zhao, P. Zheng, S.-T. Xu, Dan X. Wu, “Object Detection With Deep Learning: A Review,” *Ieee Trans. Neural Netw. Learning Syst.*, Vol. 30, No. 11, Hlm. 3212–3232, Nov 2019, Doi: 10.1109/Tnnls.2018.2876865.

[4] N. Dalal Dan B. Triggs, “Histograms Of Oriented Gradients For Human Detection,” Dalam *2005 Ieee Computer Society Conference On Computer Vision And Pattern Recognition (Cvpr’05)*, San Diego, Ca, Usa: Ieee, 2005, Hlm. 886–893. Doi: 10.1109/Cvpr.2005.177.

[5] R. Girshick, J. Donahue, T. Darrell, Dan J. Malik, "Rich Feature Hierarchies For Accurate Object Detection And Semantic Segmentation," Dalam *2014 Ieee Conference On Computer Vision And Pattern Recognition*, Columbus, Oh, Usa: Ieee, Jun 2014, Hlm. 580–587. Doi: 10.1109/Cvpr.2014.81.

[6] J. Redmon, S. Divvala, R. Girshick, Dan A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Dalam *2016 Ieee Conference On Computer Vision And Pattern Recognition (Cvpr)*, Las Vegas, Nv, Usa: Ieee, Jun 2016, Hlm. 779–788. Doi: 10.1109/Cvpr.2016.91.

[7] S. Ren, K. He, R. Girshick, Dan J. Sun, "Faster R-Cnn: Towards Real-Time Object Detection With Region Proposal Networks," *Ieee Trans. Pattern Anal. Mach. Intell.*, Vol. 39, No. 6, Hlm. 1137–1149, Jun 2017, Doi: 10.1109/Tpami.2016.2577031.

[8] Y. Lecun, "Gradient-Based Learning Applied To Document Recognition," *Proceedings Of The Ieee*, Vol. 86, No. 11, 1998.

[9] M. Ranzato, F. J. Huang, Y.-L. Boureau, Dan Y. Lecun, "Unsupervised Learning Of Invariant Feature Hierarchies With Applications To Object Recognition," Dalam *2007 Ieee Conference On Computer Vision And Pattern Recognition*, Minneapolis, Mn, Usa: Ieee, Jun 2007, Hlm. 1–8. Doi: 10.1109/Cvpr.2007.383157.

[10] A. Krizhevsky, I. Sutskever, Dan G. E. Hinton, "Imagenet Classification With Deep Convolutional Neural Networks," *Commun. Acm*, Vol. 60, No. 6, Hlm. 84–90, Mei 2017, Doi: 10.1145/3065386.

[11] K. Simonyan Dan A. Zisserman, "Very Deep Convolutional Networks For Large-Scale Image Recognition." Arxiv, 10 April 2015. Diakses: 1 Desember 2023. [Daring]. Tersedia Pada: Http://Arxiv.Org/Abs/1409.1556

[12] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, Dan V. Pattabiraman, "Comparative Analysis Of Deep Learning Image Detection Algorithms," *J Big Data*, Vol. 8, No. 1, Hlm. 66, Des 2021, Doi: 10.1186/S40537-021-00434-W.

[13] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, Dan A. Zisserman, "The Pascal Visual Object Classes (Voc) Challenge," *Int J Comput Vis*, Vol. 88, No. 2, Hlm. 303–338, Jun 2010, Doi: 10.1007/S11263-009-0275-4.

[14] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, Dan S. Belongie, "Feature Pyramid Networks For Object Detection," Dalam *2017 Ieee Conference On Computer Vision And Pattern Recognition (Cvpr)*, Honolulu, Hi: Ieee, Jul 2017, Hlm. 936–944. Doi: 10.1109/Cvpr.2017.106.

[15] R. Padilla, S. L. Netto, Dan E. A. B. Da Silva, "A Survey On Performance Metrics For Object-Detection Algorithms," Dalam *2020 International Conference On Systems, Signals And Image Processing (Iwssip)*, Niterói, Brazil: Ieee, Jul 2020, Hlm. 237–242. Doi: 10.1109/Iwssip48289.2020.9145130.

[16] H. Yoon, S.-H. Lee, Dan M. Park, "Tensorflow With User Friendly Graphical Framework For Object Detection Api".