

BAB I

PENDAHULUAN

1.1. Latar Belakang

Manusia dilahirkan dengan kelebihan dan kekurangan, bahkan ada yang mempunyai keterbatasan dalam berkomunikasi. Bentuk komunikasi lainnya adalah bahasa isyarat. Penggunaan bahasa isyarat dapat menimbulkan permasalahan lain, yaitu ketika masyarakat belum pernah memahami atau mempelajari bahasa isyarat [1]. Bahasa isyarat merupakan salah satu bentuk komunikasi non-verbal yang menjadi alternatif bahasa lisan bagi penyandang disabilitas yang menghalangi mereka untuk berkomunikasi melalui bahasa lisan [2].

Ketidakmampuan berbicara adalah suatu gangguan atau disabilitas yang menyebabkan seseorang tidak mampu dalam berbicara yang biasanya disebut tunarungu. Lalu, ada juga gangguan pendengaran yang mempengaruhi kemampuan dalam mendengarkan dan biasanya disebut tunawicara. Orang-orang tersebut termasuk disabilitas dalam berkomunikasi menggunakan media komunikasi, seperti bahasa isyarat. Data WHO menunjukkan bahwa pada tahun 2021, sekitar 1.3 miliar orang akan mengalami disabilitas di seluruh dunia. Angka ini setara dengan jumlah penduduk yang berjumlah 16% dari total populasi global [3].

Berdasarkan data WHO, jumlah penyandang disabilitas yang besar tersebut tentunya akan sulit berkomunikasi. Dalam hal ini, dibutuhkan seorang yang paham agar kedua pihak dapat saling mengerti. Namun, minat untuk memahami bahasa isyarat masih kurang sehingga menyebabkan banyak orang yang belum memahaminya. Untuk mengatasi permasalahan tersebut, diperlukan suatu sistem yang dapat mengenali huruf tulisan tangan. Hal ini memungkinkan sistem tersebut dapat memfasilitasi pembelajaran komunikasi melalui bahasa isyarat tangan, salah satunya adalah Bahasa Isyarat Amerika (*American Sign Language – ASL*).

Thomas Hopkins Gallaudet dan Laurent Clerc menciptakan ASL pada tahun 1817. Gallaudet membangun sekolah bahasa isyarat di Hartford, Connecticut, dan Clerc menjadi guru pertama bahasa isyarat di Amerika. ASL awalnya digunakan sebagai bahasa isyarat untuk berinteraksi dengan tunarungu lain di sekolah [4]. Dengan adanya ASL, membuat bahasa isyarat menjadi berkembang dan menjadi acuan dalam landasan bahasa isyarat di negara lain. Penelitian ini memilih ASL dibandingkan dengan Sistem Isyarat Bahasa Indonesia (SIBI) sebagai implementasi karena penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam tentang bagaimana bahasa isyarat berkembang sesuai dengan kebutuhan budaya dan bahasa masyarakat setempat. Selain itu, penelitian dengan ASL juga penting untuk meningkatkan sistem isyarat lokal, khususnya di Indonesia, dengan tujuan untuk mengembangkan teknologi dan sistem yang lebih baik dalam mendukung komunikasi penyandang tunarungu. Terlebih lagi, penelitian ini memiliki potensi untuk mendorong perkembangan aplikasi atau teknologi yang memudahkan komunikasi antarpemuter bahasa isyarat dari negara yang berbeda, membuka peluang untuk menciptakan sistem lintas bahasa isyarat yang lebih inklusif dan praktis.

Salah satu solusi permasalahan yang diberikan adalah melakukan deteksi gestur tangan Bahasa Isyarat Amerika untuk menerjemahkan artinya. Dalam beberapa penelitian, pengenalan gestur tangan untuk menerjemahkan bahasa isyarat dengan algoritma *matching learning Naive Bayes* mencapai akurasi 96% [5]. Berbagai metode dapat digunakan dalam pendeteksian teks atau pendeteksian objek secara umum. Salah satu metode tersebut adalah Faster R-CNN. Faster R-CNN merupakan salah satu arsitektur CNN berbasis region yang digunakan dalam pendeteksian objek, yang dikembangkan dari Fast R-CNN dengan memperkenalkan *Region Proposal Network (RPN)* untuk menggantikan mekanisme *Selective Search* [6]. Faster R-CNN memiliki beberapa kelebihan signifikan dibandingkan dengan model deteksi objek pendahulu lainnya, seperti CNN, Fast R-CNN, dan Mask R-CNN. CNN adalah model dasar untuk pengenalan gambar,

yang hanya mampu mengklasifikasikan gambar ke dalam kategori tertentu tanpa memberikan informasi lokasi objek dalam gambar. Faster R-CNN mengatasi hal ini dengan menambahkan kemampuan untuk mendeteksi objek serta menentukan posisi objek dalam gambar melalui *bounding boxes* [37]. Selanjutnya, dibandingkan dengan Fast R-CNN, Faster R-CNN menawarkan peningkatan besar dalam hal kecepatan dan efisiensi. Fast R-CNN menggunakan *Region of Interest Pooling (RoI Pooling)* untuk mengekstrak fitur dari area yang relevan, namun masih bergantung pada algoritma eksternal seperti *Selective Search* untuk menghasilkan *region proposals*. Sebaliknya, Faster R-CNN mengintegrasikan *Region Proposal Network (RPN)* yang secara otomatis menghasilkan *region proposals* langsung di dalam jaringan. Hal ini tidak hanya meningkatkan efisiensi, tetapi juga mempercepat proses deteksi objek secara signifikan [28]. Dibandingkan dengan Mask R-CNN, Faster R-CNN tetap unggul dalam hal kecepatan dan kesederhanaan, meskipun Mask R-CNN lebih lanjut mengembangkan Faster R-CNN dengan menambahkan kemampuan untuk menghasilkan *segmentation masks* untuk setiap objek yang terdeteksi. Mask R-CNN memerlukan perhitungan tambahan untuk menghasilkan segmentasi *pixel per pixel*, yang membuatnya lebih berat dan lebih lambat dibandingkan Faster R-CNN yang hanya menghasilkan *bounding boxes* dan klasifikasi objek. Oleh karena itu, Faster R-CNN lebih cocok untuk aplikasi deteksi objek yang tidak memerlukan segmentasi mendalam karena lebih efisien dan cepat [38]. Secara keseluruhan, Faster R-CNN menawarkan keseimbangan yang baik antara akurasi dan efisiensi dalam deteksi objek. Integrasi *Region Proposal Network* membuatnya lebih cepat dibandingkan Fast R-CNN, sementara kesederhanaannya membuatnya lebih ringan dibandingkan Mask R-CNN. Model Faster R-CNN ideal untuk aplikasi yang memerlukan deteksi objek yang cepat dan akurat, namun tidak memerlukan detail segmentasi tingkat tinggi.

Di penelitian lainnya, mengenali Bahasa Isyarat Amerika dengan metode Faster R-CNN *backbone DenseNet-41* didapatkan capaian akurasi

98,6%. Metode ini dapat mencapai akurasi 98.6% karena penggunaan blok padat yang efisien dalam mengekstrak *fitur* dari gambar, penggunaan fungsi transformasi non-linear untuk mengurangi kompleksitas komputasi yang diperlukan, dan pengurangan jumlah parameter yang disesuaikan [7]. Selain *backbone DenseNet-41* pada penelitian sebelumnya, banyak *backbone* Faster R-CNN, seperti *MobileNet* dan *ResNet*.

MobileNet adalah jaringan saraf tiruan ringan yang dikembangkan oleh *Google* untuk aplikasi pada perangkat mobile. Cara kerjanya adalah dengan menggantikan operasi konvolusi standar dengan konvolusi terpisah kedalaman untuk mengurangi jumlah parameter model [8]. *ResNet* adalah arsitektur jaringan saraf tiruan. *ResNet* ini menggunakan blok residual yang terdiri dari dua lapisan konvolusi dan koneksi *shortcut* untuk mengalirkan *output* blok sebelumnya ke blok berikutnya tanpa modifikasi. Koneksi *shortcut* ini membantu mengatasi masalah pelatihan yang dalam jaringan yang sangat dalam [9].

Untuk *Non-maximum Suppression* (NMS) itu adalah teknik deteksi objek yang populer, terutama dalam pengolahan gambar dan penglihatan komputer. Tujuan utama NMS adalah untuk mengurangi tumpang tindih yang dibuat oleh algoritma deteksi objek antar kotak pembatas (*bounding boxes*). Kriteria ini biasanya berupa angka probabilitas dan ukuran tumpang tindih [10].

Berdasarkan permasalahan dan metode yang sudah dijelaskan sebelumnya, penelitian ini akan membahas mengenai “ADAM DAN SGD PADA FASTER RCNN RESNET DAN MOBILENET UNTUK DETEKSI OBJEK GESTUR TANGAN BAHASA ISYARAT”.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan, maka dapat disimpulkan rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana penerapan *non-maximum suppression* terhadap adam dan SGD pada Faster R-CNN *ResNet* dan *MobileNet* untuk deteksi objek

gestur tangan bahasa isyarat?

2. Bagaimana evaluasi performa adam dan SGD pada Faster R-CNN *ResNet* dan *MobileNet* untuk deteksi objek gestur tangan bahasa isyarat?

1.3. Tujuan

Berdasarkan rumusan masalah diatas, maka tujuan dari penelitian ini adalah sebagai berikut.

1. Mengimplementasikan penerapan *non-maximum suppression* terhadap adam dan SGD pada Faster R-CNN *ResNet* dan *MobileNet* untuk deteksi objek gestur tangan bahasa isyarat.
2. Mengevaluasi performa adam dan SGD pada Faster R-CNN *ResNet* dan *MobileNet* untuk deteksi objek gestur tangan bahasa isyarat.

1.4. Manfaat

Adapun manfaat ini adalah sebagai berikut.

1. Menjadi jurnal acuan penelitian selanjutnya dalam membandingkan berbagai optimisasi, *backbone*, dan model *matching learning* atau *deep learning* terbaru.
2. Membantu pengguna dalam mengartikan gestur tangan pada Bahasa Isyarat Amerika.

1.5. Batasan Masalah

Adapun batasan masalah dari pelaksanaan dan pembahasan penelitian ini adalah sebagai berikut.

1. Gestur tangan yang akan dideteksi adalah yang merepresentasikan alfabet dari A sampai Z pada Bahasa Isyarat Amerika.
2. Data yang digunakan adalah data sekunder. File gambar (*image*) berformat “.jpg” dan anotasi setiap gambar (*image*) berformat “.xml”. Data tersebut diambil dari sumber website

www.universe.roboflow.com

3. Fokus utama metode yang digunakan dalam penelitian ini adalah membandingkan evaluasi Faster R-CNN *backbone ResNet-50* dan *MobileNet v3* dengan optimisasi adam dan SGD.