BAB II

TINJAUAN PUSTAKA

Untuk mendeteksi gerakan bahasa isyarat, teknologi pembelajaran mesin telah menjadi subjek penelitian yang semakin penting, karena teknologi ini dapat memungkinkan komunitas yang bergantung pada bahasa isyarat berinteraksi dengan lebih mudah dan inklusif. Tinjauan pustaka ini akan membahas literatur terkait untuk mendeteksi bahasa isyarat dalam waktu nyata menggunakan model *You Only Look Once* sebagai pembanding serta penambah wawasan pengetahuan dalam penelitian yang akan dikerjakan kali ini.

Penelitian yang dilakukan pada tahun 2021 oleh Daniels *et al.* [7] berfokus pada model YOLO untuk membangun sistem pengenalan bahasa isyarat Indonesia. Tujuan penelitian ini adalah membuat sistem yang dapat memproses input video secara *real time* dan menerjemahkan bahasa isyarat ke dalam teks. Penelitian ini menggunakan CNN yang memiliki model YOLO. Peneliti mengubah jumlah kelas dan channel untuk mempermudah pengenalan bahasa isyarat Indonesia. Peneliti menggunakan BISINDO untuk mengumpulkan data, yang terdiri dari 4.547 gambar yang menunjukkan 24 kelas bahasa isyarat statis. Pada penelitian ini yang menggunakan data gambar, sistem beroperasi dengan sangat baik dengan 100% presisi, *recall*, akurasi, dan skor F1. Pada eksperimen yang menggunakan data video, sistem beroperasi dengan presisi 77,14%, *recall* 93,1%, akurasi 72,97%, dan skor F1 84,38% pada kecepatan 8 FPS. Hasil penelitian menunjukkan bahwa metode YOLO dapat mengenali dan memproses bahasa isyarat Indonesia dalam data gambar dan video secara *real time*. Dengan menerjemahkan bahasa isyarat ke dalam teks, sistem ini diharapkan dapat membantu masyarakat umum berkomunikasi dengan penyandang disabilitas.

Penelitian selanjutnya ditulis oleh Arifah *et al.* [8] memiliki tujuan untuk membuat model pintar yang dapat mengidentifikasi teks yang terdiri dari bahasa isyarat. Ini dilakukan karena beberapa orang, terutama penyandang tunarungu, tidak dapat berkomunikasi dengan bahasa lisan. Mereka biasanya berbicara melalui bahasa isyarat. Penelitian ini menggunakan model YOLO untuk mengidentifikasi objek tangan dan CNN untuk mengklasifikasi jenis gerakan tangan yang berbeda.

Dataset BISINDO digunakan. Hasil uji coba menunjukkan bahwa model cerdas yang dikembangkan dalam penelitian ini dapat dengan cukup akurat mengidentifikasi bahasa isyarat sebagai teks dengan akurasi sebesar 89% dalam mendeteksi dan mengklasifikasi gerakan tangan pada video percakapan bahasa isyarat Indonesia yang menggunakan metode YOLO dan CNN.

Penelitian selanjutnya ditulis oleh Bankar *et al.* [9] memiliki tujuan untuk membangun sebuah model yang dapat mengenali bahasa isyarat secara *real time* dengan menggunakan algoritma pembelajaran mendalam model YOLO. Penelitian ini disebut untuk meningkatkan akurasi, penelitian ini menggunakan *dataset* bahasa isyarat *Roboflow* dan menambahkan 422 gambar buatan sendiri. Mereka juga membandingkan kinerja model CNN dan model YOLO v5. *Preprocessing* data dilakukan dengan mengubah gambar menjadi piksel. Selanjutnya, *dataset* dibagi menjadi data latih, validasi, dan uji. Dengan data yang telah disiapkan, model dilatih dan diuji. Kemudian model deteksi bahasa terbaik digunakan. Hasil penelitian menunjukkan bahwa model YOLO v5 yang disarankan memiliki akurasi 88,4%, presisi 76,6%, dan *recall* 81,2%, lebih tinggi dari model CNN yang hanya mencapai 52,98%. Selain itu, model YOLO v5 mendeteksi bahasa isyarat secara *real time* lebih cepat dan lebih akurat daripada model CNN.

Penelitian yang dilakukan oleh Permana dan Sutopo [10] memiliki tujuan untuk membuat *platform* praktis pengenalan abjad Sistem Isyarat Bahasa Indonesia (SIBI) sebagai alat yang efektif untuk belajar bahasa isyarat. Dalam penelitian ini, model YOLOv5 dan CNN digunakan untuk mendeteksi gerakan bahasa isyarat SIBI. Pengujian dilakukan dalam dua tahap. Tahap pertama melibatkan pengujian antarmuka aplikasi yang berhasil menampilkan enam halaman antarmuka pengguna. Tahap kedua melibatkan membandingkan 26 kelas aktual untuk mendeteksi gerakan bahasa isyarat SIBI. Beberapa kelas seperti kelas 'A', 'E', 'D', dan 'J' memiliki tingkat akurasi deteksi yang rendah, menurut hasil penelitian. Ini mungkin karena gerakan yang hampir sama atau perubahan posisi yang lebih kompleks. Secara keseluruhan, hasil deteksi waktu nyata yang dilakukan oleh kamera ponsel menunjukkan akurasi sebesar 77%. Peneliti berharap bahwa *platform* ini akan membantu teman tuli dan masyarakat umum berkomunikasi dengan lebih mudah dan lancar.

Penelitian yang dilakukan oleh Prabhakar *et al.* [11] memiliki tujuan utama untuk membuat sistem yang dapat menerjemahkan analisis gerakan isyarat bahasa, mengidentifikasi, membuat deskripsi teks dalam bahasa Inggris, dan mengonversikan teks menjadi ucapan. Ini dilakukan untuk membantu orang tuli-bisu dan orang yang dapat berbicara berbicara. Para peneliti menggunakan algoritma seperti CNN, FRCNN (*Faster-Convolutional Neural Networks*), YOLO dan *MediaPipe* untuk mencapai tujuan tersebut. Semua algoritma ini digunakan untuk mendeteksi dan mengenali gerakan isyarat tangan yang ditangkap oleh kamera, lalu menghasilkan teks bahasa Inggris dan mengonversikannya menjadi ucapan. Hasil penelitian menunjukkan bahwa, meskipun algoritma CNN, FRCNN, dan YOLO dapat mengenali gerakan isyarat tangan, *MediaPipe* paling cocok untuk memenuhi semua persyaratan proyek. *MediaPipe* juga terbukti memiliki akurasi yang tinggi dan dapat melakukan konversi.

Penelitian yang dilakukan oleh Patil *et al.* [12] memiliki tujuan utama untuk membuat sistem yang dapat menerjemahkan isyarat ke teks bahasa Inggris secara *real time*. Permasalahan dari penelitian ini disebabkan oleh masalah yang sering dihadapi oleh orang dengan disabilitas tutur (*signer*) ketika mereka berinteraksi dengan orang yang tidak memahami bahasa isyarat (*non-signer*). Komunikasi menjadi sulit jika ada penerjemah bahasa isyarat yang terbatas, terutama di bidang medis, hukum, pendidikan, dan pelatihan. Para ilmuwan menggunakan metode *deep learning*, terutama CNN, untuk mengidentifikasi gerakan tangan sebagai bahasa isyarat dalam penelitian mereka. Untuk membedakan dan melacak gerakan tangan dalam video masukan, mereka juga menggunakan metode pelacakan objek dan segmentasi gambar. Sebuah sistem yang menerjemahkan isyarat ke teks bahasa Inggris secara *real time* diciptakan sebagai hasil dari penelitian ini. Diharapkan bahwa sistem ini akan membuat komunikasi menjadi lebih mudah dan lancar bagi penyandang disabilitas tutur dalam berbagai aspek kehidupan karena akan membantu mengurangi jarak antara orang yang menandatangani dan orang yang tidak menandatangani.

Penelitian yang dilakukan oleh Mohamed dan Hussein [13], memiliki tujuan utama yaitu membangun sistem komunikasi yang menggunakan teknologi pengenalan gerakan tangan untuk penyandang tuna rungu dan wicara. Para peneliti

mencapai tujuan ini dengan menggunakan metode CNN, yang dianggap lebih baik daripada pendekatan sebelumnya dalam mendeteksi, melokalisasi, dan mengidentifikasi gerakan tangan manusia. Arkitektur *neural network* DARKNET-50 digunakan. Penelitian dilakukan dalam dua tahap utama. Pertama, menemukan area tangan manusia yang berbeda dari gambar. Kemudian, menggunakan kamus alfabet untuk mengklasifikasikan gambar tangan yang ditemukan. Hasil penelitian menunjukkan bahwa penggunaan jaringan saraf konvolusi buatan, terutama YOLO V-2 dan DARKNET-19, sangat sesuai dari segi akurasi, kecepatan eksekusi, dan generalitas dalam memproses data input. Ini memungkinkan penggunaan program tanpa menentukan bentuk, kecepatan, warna tangan, atau ekspresi fisik lainnya.

Penelitian yang dilakukan oleh Minh [14], bertujuan untuk memberikan informasi tentang bahasa isyarat dan budaya yang ada di sekitarnya, serta untuk melihat apa yang kurang dalam penelitian sebelumnya. Dengan menggunakan metode penelitian konstruktif, penelitian ini akan memberikan kontribusi yang berbeda. Untuk memberikan pembaca wawasan yang dapat diandalkan, Bui Hien Minh mengumpulkan informasi tentang bahasa isyarat dan komunitas tuli-bisu dari buku dan penelitian yang relevan. Selain itu, untuk menyediakan informasi terkini dan akurat, dia juga mengumpulkan pengetahuan teknis dari artikel ilmiah dan sumber internet. Untuk memastikan keragaman data yang digunakan untuk melatih model YOLO, secara manual dikumpulkan dari video YouTube. Bui Hien Minh membuat sebuah aplikasi *web* untuk menunjukkan hasil penelitiannya. Tujuan aplikasi ini adalah untuk mengevaluasi seberapa efektif model YOLO dalam menerjemahkan abjad bahasa isyarat Amerika. Meskipun YOLOv3 sangat baik untuk mendeteksi objek statis dan sangat cepat, penelitian ini menunjukkan bahwa versi ini bukan solusi terbaik untuk bahasa isyarat.

Penelitian yang dilakukan oleh Ahmadi *et al.* [15] memiliki tujuan utama untuk mengembangkan sistem pengenalan bahasa isyarat Arab (ArSL) yang tepat yang menggunakan pendekatan *deep learning*. Dalam penelitian ini, metode *deep learning* digunakan dengan model deteksi objek YOLO. Untuk meningkatkan akurasi pengenalan ArSL, mereka juga menggabungkan teknik ekstraksi fitur seperti blok perhatian pribadi, modul perhatian saluran, modul perhatian spasial, dan modul *cross-convolution*. Hasil penelitian menunjukkan bahwa model yang

diusulkan mampu mencapai akurasi pengenalan ArSL sebesar 98,9%. Selain itu, model ini memiliki tingkat presisi yang tinggi, dengan skor mAP@0.5 0,9909 dan mAP@0.5:0,95, jauh di atas metode modern lainnya. Ini menunjukkan kemampuan model ini untuk menemukan dan mengklasifikasikan tanda-tanda ArSL yang kompleks. Diharapkan model ini akan membantu meningkatkan inklusi sosial orang tuli di wilayah Arab dan menawarkan metode komunikasi yang berbeda untuk menghubungkan orang.

Penelitian yang dilakukan oleh Dima dan Ahmed [16] bertujuan untuk membantu masyarakat yang menggunakan bahasa verbal dan masyarakat yang menggunakan bahasa isyarat, yang sering mengalami kesulitan untuk memahami makna gerakan isyarat yang mereka gunakan. Dalam penelitian mereka, Dima dan Ahmed menggunakan algoritma YOLOv5, yang merupakan algoritma deteksi objek yang cepat dan efektif berbasis CNN. Untuk melatih dan memeriksa model yang mereka buat, mereka menggunakan *dataset* MU *HandImages* ASL. Hasil penelitian menunjukkan bahwa model YOLOv5 yang dikembangkan oleh Dima dan Ahmed memiliki kinerja yang cukup baik, dengan tingkat presisi sebesar 95%, tingkat *recall* sebesar 97%, map@0.5 sebesar 98%, dan map@0.5:0.95 sebesar 98%. Hasil ini dianggap cukup untuk mengidentifikasi gerakan isyarat dalam waktu nyata. Selain itu, model YOLOv5 yang digunakan sangat ringan, dengan memori hanya 167 MB, membuatnya dapat digunakan pada perangkat *mobile*.

Penelitian yang dilakukan oleh Asri, *et al.* [17] memiliki tujuan utama untuk mengembangkan sebuah model yang dapat mendeteksi bahasa isyarat Malaysia (*Malaysian Sign Language*) secara *real time* menggunakan metode CNN. Model yang digunakan, YOLOv3, diharapkan algoritma ini akan membantu orang normal berkomunikasi dengan orang penyandang disabilitas pendengaran yang menggunakan bahasa isyarat. Untuk melakukan ini, para peneliti mengumpulkan gambar bahasa isyarat dari berbagai sumber *web* dan video bahasa isyarat yang direkam. Kemudian, mereka melabeli gambar sebagai huruf atau gerakan. Sebelum melatih dan menguji sistem menggunakan kerangka kerja Darknet, data gambar diproses sebelum diproses. Hasil penelitian menunjukkan bahwa sistem yang disarankan mencapai akurasi 63% pada 7000 iterasi dengan saturasi pembelajaran (*overfitting*). Meskipun model ini masih belum ideal, para peneliti mengatakan

bahwa setelah diuji secara efektif, model ini akan diintegrasikan dengan *platform* lain seperti aplikasi telepon.

Penelitian yang dilakukan oleh Bhavadharshini, *et al.* [18] memiliki tujuan untuk mengembangkan sistem penerjemah bahasa isyarat Amerika (*American Sign Language*) secara *real time*. Dengan menerjemahkan bahasa isyarat secara otomatis, sistem ini diharapkan dapat membantu orang dengar dan tuna rungu berkomunikasi satu sama lain. Dua tahap utama dalam metodologi penelitian ini adalah Tahap Pelatihan dan Tahap Deteksi. Pada tahap pelatihan, proses pengumpulan data, *preprocessing* gerakan isyarat, dan pemantauan pergerakan tangan dilakukan dengan menggunakan algoritma kombinasi. Hasil penelitian menunjukkan bahwa sistem penerjemah bahasa isyarat yang menggunakan CNN dan model YOLO dapat mengenali dan menerjemahkan gerakan isyarat secara *real time*. Dengan akurasi dan kecepatan pemrosesan yang tinggi, sistem ini dapat membantu meningkatkan komunikasi antara orang dengar dan tuna rungu.

Penelitian yang dilakukan oleh AL-Shaheen [19] bertujuan untuk melatih sebuah model yang dapat mendeteksi dan mengenali gerakan dan isyarat tangan dan kemudian menerjemahkannya menjadi huruf, angka, dan kata-kata dengan menggunakan model YOLO melalui gambar atau video dalam waktu nyata (*real time*). Para peneliti menggunakan metode YOLO, yang mendeteksi dan mengenali objek menggunakan CNN. YOLO sangat akurat dan cepat. Hasil penelitian menunjukkan bahwa model yang dikembangkan memiliki akurasi yang luar biasa dengan MAP (*Mean Average Precision*) sebesar 98,01%, rata-rata kehilangan sebesar 1,3, *recall* sebesar 0,96, dan F1 sebesar 0,96. Namun, akurasi videonya sama dengan akurasi gambar, dengan kecepatan 28,9 *frame* per detik. Secara keseluruhan, penelitian ini menghasilkan sebuah model yang dapat mendeteksi dan mengenali bahasa isyarat Amerika dengan sangat akurat dan cepat, bahkan dalam waktu nyata, menggunakan model YOLO.

Penelitian yang dilakukan oleh Alaftekin, [20] bertujuan untuk membuat model deteksi objek yang lebih efektif untuk mengidentifikasi isyarat bahasa Turki. Untuk meningkatkan kinerja dan deteksi *real time*, tim peneliti menggunakan algoritma YOLOv4-CSP yang berbasis CNN. CSPNet ditambahkan ke leher YOLOv4 asli untuk meningkatkan kinerja jaringan. Selanjutnya, untuk

mendapatkan deteksi objek yang lebih efisien dalam bahasa isyarat Turki, model deteksi objek baru diusulkan dengan mengoptimalkan model YOLOv4-CSP. Hasil penelitian menunjukkan bahwa penelitian ini telah berhasil mengembangkan model YOLOv4-CSP yang lebih efisien untuk mendeteksi angka dalam bahasa isyarat Turki. Dalam waktu 9,8 milidetik, metode yang diusulkan memperoleh presisi 98,95%, recall 98,15%, skor F1 98,55, dan mAP 99,49%. Algoritma ini juga terbukti mengungguli algoritma lain dalam kinerja *real time* dan akurasi prediksi isyarat tangan, terlepas dari latar belakang.

Penelitian yang dilakukan oleh Tyagi, [21] bertujuan untuk menggunakan model YOLO untuk mendeteksi bahasa isyarat Amerika (*American Sign Language*) dan membandingkan berbagai model YOLO dengan model yang disesuaikan untuk mengenali bahasa isyarat. Selain itu, model YOLO sendiri digunakan untuk klasifikasi dan deteksi objek dalam penelitian ini. Untuk pelatihan dan pengujian, *dataset* bahasa isyarat Amerika digunakan untuk menjalankan eksperimen dengan model yang disesuaikan. Hasil penelitian menunjukkan bahwa YOLOv8 menunjukkan hasil yang lebih baik dalam hal presisi dan mAP (*mean Average Precision*) dibandingkan dengan versi YOLO lainnya. Sementara itu, YOLOv7 menunjukkan nilai recall yang lebih tinggi selama pengujian. Menurut penelitian ini, model kustom yang diusulkan mencapai presisi 95%, recall 97%, dan mAP 96% dalam pengenalan gerakan tangan secara *real time*. Dengan menggunakan model YOLO yang efektif dan akurat, penelitian ini sangat membantu dalam pengembangan sistem pendeteksi bahasa isyarat Amerika dan dapat menjadi dasar untuk pengembangan aplikasi terkait bahasa isyarat di masa depan.

Sehingga dapat disimpulkan dari penelitian-penelitian sebelumnya memiliki kelebihan dan juga perbedaan dari penelitian yang dilakukan. Penelitian sebelumnya memiliki kelebihan dari segi *metrics evaluation* yang beberapa diantaranya sudah memiliki nilai yang baik tetapi tidak dijelaskan lebih lanjut untuk kondisi yang mendukung tinggi rendahnya akurasi itu sendiri dan berdasarkan hasil juga didapati penggunaan model YOLO dianggap sudah tepat pada pendeteksian gambar. Namun, terdapat perbedaan atau pembaharuan dari penelitian sebelumnya yang akan dikembangkan pada penelitian kali ini yaitu menggunakan versi model YOLO terbaru yaitu model YOLOv8, penggunaan *dataset* yang lebih banyak dan

menambahkan kondisi pencahayaan yang menjadi salah satu faktor tinggi rendahnya akurasi.