

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Di era globalisasi yang semakin maju, kemampuan berbahasa Inggris telah menjadi salah satu keterampilan fundamental yang krusial untuk dikuasai, bahkan sejak usia dini. Bagi siswa sekolah dasar di Indonesia, penguasaan bahasa Inggris bukan lagi sekadar nilai tambah, melainkan sebuah kompetensi esensial yang membuka gerbang literasi digital, akses informasi global, serta peluang di masa depan. Namun, berdasarkan laporan *EF English Proficiency Index (EF EPI)* tahun 2024, Indonesia menempati peringkat ke-80 dari 116 negara dengan kategori *low proficiency* [1]. Kemudian penelitian oleh Suryanto (2025) mengidentifikasi beberapa kendala utama dalam pengajaran bahasa Inggris di sekolah dasar Yogyakarta. Yaitu daya ingat materi yang rendah pada siswa, rentang konsentrasi yang terbatas, motivasi belajar yang minim, serta keterbatasan variasi metode mengajar yang dominan bersifat hafalan dan *drill* tanpa pendekatan interaktif [2]. Data ini menggarisbawahi urgensi untuk mencari solusi inovatif dalam meningkatkan kemampuan bahasa Inggris di Indonesia, terutama pada jenjang pendidikan dasar. Metode pengajaran konvensional yang cenderung berpusat pada hafalan dan pengulangan terkadang kurang efektif dalam menciptakan lingkungan belajar yang interaktif dan menarik.

Di sisi lain perkembangan pesat dalam ranah kecerdasan buatan (*Artificial Intelligence/AI*), khususnya pada *computer vision*, telah membuka potensi transformatif dalam mengatasi berbagai tantangan saat ini. Salah satu inovasi

paling signifikan dalam *computer vision* adalah teknologi deteksi objek, yang bisa mengklasifikasi dan lokalisasi objek secara otomatis dalam gambar maupun video. Di antara berbagai arsitektur yang ada, model *YOLO (You Only Look Once)* telah menunjukkan performa yang unggul dalam hal kecepatan dan akurasi deteksi objek. Dengan hadirnya *YOLO11* sebagai versi terkini yang memperkenalkan *blok C3k2* serta modul *C2PSA*, yang meningkatkan kemampuan serta efisiensi dalam mendeteksi objek yang kecil dan bersamaan [3] Menjadikannya pilihan optimal untuk implementasi dalam konteks aplikasi edukasi khususnya deteksi objek sehari-hari untuk siswa sekolah dasar, ini sesuai dengan karakteristik kognitif anak-anak yang menyukai eksplorasi visual.

Berbagai penelitian telah mengeksplorasi integrasi deteksi objek dan interaksi multimodal. Alahmadi *et al.* (2023) mengembangkan sistem berbasis *YOLOv4_ResNet101* yang mampu mendeteksi objek dan mengubah hasil deteksi menjadi suara untuk membantu penyandang tunanetra dengan akurasi mencapai 96,34% menggunakan *dataset MS COCO* [4]. Lilhare *et al.* (2021) mengimplementasikan *YOLOv3* pada *Raspberry Pi* yang terhubung ke kamera dan menggunakan *Google TTS* untuk menyuarakan objek yang terdeteksi, khususnya untuk meningkatkan mobilitas pengguna tunanetra [5]. Penelitian oleh Putri *et al.* (2024) juga mengintegrasikan *YOLOv5s* dan *TTS* dalam aplikasi edukasi anak usia dini, dengan akurasi deteksi sebesar 85% terhadap objek seperti abjad dan buah-buahan [6]. Selanjutnya, Swathi *et al.* (2024) menggunakan *YOLOv7* dan *Google TTS* untuk membangun sistem yang mendeteksi objek secara cepat dan akurat, kemudian menyuarakan hasil deteksi, yang tidak hanya

meningkatkan aksesibilitas visual tetapi juga mendukung pengguna dalam mengenali objek di berbagai kondisi lingkungan [7]. Penelitian -penelitian ini menunjukkan bahwa integrasi *YOLO* dan *TTS* sangat potensial diterapkan dalam berbagai aplikasi berbasis deteksi visual dengan umpan balik suara.

Dengan adanya potensi tersebut, penelitian ini bertujuan untuk mengimplementasikan *YOLO11* dalam bentuk aplikasi berbasis *web* untuk deteksi objek interaktif, khusus dalam konteks benda-benda sehari-hari. Hasil deteksi akan disajikan dalam bentuk teks dwi-bahasa Inggris-Indonesia dan dilengkapi dengan fitur audio untuk pelafalan bahasa Inggris serta fitur koreksi pengucapan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat dirumuskan beberapa permasalahan sebagai berikut : Bagaimana merancang dan membangun aplikasi berbasis *web* untuk deteksi objek dengan *YOLO11* yang dapat mendeteksi benda sehari-hari dan mengeluarkan hasil dalam bentuk teks Inggris-Indonesia serta audio berbahasa Inggris dan fitur koreksi pengucapan ?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, maka tujuan dari penelitian ini adalah : merancang dan membangun aplikasi berbasis *web* untuk deteksi objek sehari-hari dengan *YOLO11* dengan keluaran teks Inggris- Indonesia dan audio berbahasa Inggris dan fitur koreksi pengucapan.

1.4 Batasan Penelitian

Untuk memastikan bahwa penelitian ini tetap fokus dan tidak terlalu luas, beberapa batasan berikut telah ditetapkan:

1. Model yang digunakan untuk proses *fine-tuning* adalah *yolo11m-obb.pt*, yaitu model pra-pelatihan (*pre-trained*) yang akan disesuaikan menggunakan *dataset custom*. *Dataset* ini dikumpulkan dari berbagai sumber seperti situs *Kaggle*, *Roboflow*, internet, maupun secara manual.
2. Objek yang akan di deteksi adalah 25 objek umum yang bisa di deteksi dengan kamera secara langsung, berikut 30 objek tersebut : sendok, garpu, piring, gelas, cangkir, mangkuk, gantungan baju, gunting, gunting kuku, sisir rambut, botol air mineral, pengisi daya ponsel, spidol, pulpen, pensil, penghapus karet, rautan pensil, penggaris , koin, uang kertas rupiah, pisang, salak, tomat, timun dan terong ungu.
3. Aplikasi yang dikembangkan berbasis *web* untuk perangkat yang mendukung kamera, sehingga aplikasi dapat diakses melalui peramban tanpa memerlukan instalasi tambahan.
4. Hasil dari deteksi akan disajikan dalam bentuk teks Inggris-Indonesia dan audio dalam bahasa Inggris serta fitur koreksi pengucapan tanpa menyediakan fitur terjemahan bahasa lainnya.
5. Penelitian ini terutama berorientasi pada implementasi sistem, tanpa melibatkan langsung pengujian pada siswa atau pengguna akhir, seperti anak-anak di sekolah.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Bagi anak-anak: Memberikan media pembelajaran bahasa Inggris yang interaktif dan menyenangkan melalui teknologi deteksi objek dengan objek di dunia nyata.
2. Bagi peneliti: Menjadi acuan dalam studi lanjut mengenai penerapan model *YOLO* dalam bidang edukasi interaktif.

1.6 Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri dari enam bagian utama, sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan pada penelitian ini. Teori-teori *computer vision*, deteksi objek, *YOLO11*, aspek edukasi dalam sistem interaktif anak seperti peran *SLA*, teks, audio dan *ui*, *unified modeling language* dan metode *waterfall*.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi tahapan-tahapan yang dilakukan dalam penelitian ini, termasuk metode pengumpulan data, metode pengembangan aplikasi, serta alat dan teknik yang digunakan untuk implementasi dan pengujian.

BAB 4 ANALISA DAN PERANCANGAN

Bab ini berisi analisa dan perancangan implementasi. Dalam bab ini, akan dijelaskan proses analisa kebutuhan, perancangan sistem, serta desain antarmuka pengguna.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dari hasil analisa dan perancangan yang telah dilakukan. Selain itu, bab ini juga akan menjelaskan proses pengujian aplikasi, termasuk metode pengujian, hasil pengujian, dan evaluasi terhadap aplikasi yang telah dikembangkan.

BAB 6 PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran-saran untuk pengembangan aplikasi atau penelitian selanjutnya. Bab ini juga akan merangkum kontribusi penelitian dan manfaat aplikasi yang telah dikembangkan.

BAB 2

LANDASAN TEORI

2.1 *Computer Vision* Dan Deteksi Objek

2.1.1 Konsep Dasar *Computer Vision*

Computer Vision merupakan bagian dari teknologi kecerdasan buatan yang berkaitan dengan data visual. Teknologi ini memungkinkan komputer untuk menangkap, melihat, dan memahami lingkungan sekitar melalui gambar atau video. Dengan *computer vision*, mesin dapat mendeteksi, memahami, dan merespons informasi visual dengan cara yang mirip dengan penglihatan manusia [8].

Dalam prakteknya, pemahaman mengenai konteks adalah faktor penting dalam berbagai tugas di bidang *computer vision*. Informasi yang berkaitan dengan konteks memiliki fungsi yang sangat berarti dalam banyak tugas *computer vision*, seperti pengenalan objek, pengawasan pergerakan dalam video, klasifikasi gambar, dan lain-lain. Terkadang, mengenali objek atau aksi tanpa konteks bisa jadi sangat sulit, dan informasi konteks dapat memberikan bantuan yang besar untuk memperjelas pemahaman tentang suatu adegan atau kejadian [9].

Informasi kontekstual sangat penting dalam *computer vision*, karena memberikan data yang berguna untuk menjelaskan dan menganalisis gambar *visual*. Dengan menggunakan informasi kontekstual yang terdapat di dalam gambar atau video, kita bisa meningkatkan akurasi dan kemampuan detektor objek [10].

2.1.2 Deteksi Objek

Deteksi objek secara *real-time* adalah hal penting dalam *computer vision* untuk mengenali dan menentukan lokasi objek dalam gambar atau video. Teknologi ini banyak digunakan dalam keamanan, mobil tanpa pengemudi dengan kecepatan dan efisiensi sebagai faktor kunci [11].

Kemajuan dalam *deep learning* telah mengarah pada perkembangan penting, khususnya dengan algoritma *YOLO (You Only Look Once)*. *YOLO* terkenal karena dapat mendeteksi objek dengan cepat dan akurat. Versi terbarunya terus diperbaiki untuk efisiensi, membuatnya cocok untuk aplikasi *real-time* seperti pemantauan lalu lintas dan lingkungan [12].

Salah satu versi terbaru, *YOLO11*, dibuat dengan tujuan utama untuk meningkatkan efisiensi dalam mendeteksi objek di perangkat *edge* dan dalam lingkungan dengan sumber daya komputer terbatas. Model ini mengintegrasikan modul arsitektur baru seperti *C3k2* dan *C2PSA*, yang secara signifikan memperbaiki kemampuan dalam ekstraksi fitur dan ketepatan deteksi. Penelitian oleh Liu *et al.* (2024) menunjukkan bahwa *YOLO11* dapat meningkatkan akurasi *mAP* sebesar 1,7% dan mengurangi latensi inferensi hingga 7,6% jika dibandingkan dengan versi sebelumnya, sehingga menjadikannya pilihan yang lebih baik untuk kebutuhan deteksi objek *real-time* yang memerlukan kecepatan dan ketepatan tinggi [13].

2.1.3 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah bagian dari algoritma *deep learning* yang dikembangkan dari *Multilayer Perceptron (MLP)* dan dirancang

untuk mengolah data dua dimensi seperti citra gambar. CNN termasuk ke dalam jaringan syaraf tiruan dalam kategori *deep neural network* karena menggunakan banyak lapisan tersembunyi untuk mengekstraksi fitur secara bertahap dari input [14].

Convolutional Neural Networks (CNN) merupakan arsitektur jaringan saraf yang dirancang untuk memproses data *grid* seperti gambar melalui operasi konvolusi, yang mengekstrak fitur lokal seperti tepi dan tekstur secara hierarkis. Operasi ini menggunakan kernel kecil yang meluncur di atas input untuk menghitung *dot product*, memanfaatkan parameter sharing dan *sparse connectivity* guna mengurangi kompleksitas komputasi serta mencapai invariansi translasi. Lapisan konvolusi diikuti aktivasi *non-linear* seperti *ReLU* dan *pooling* untuk mempertahankan fitur penting sambil mengurangi dimensi spasial [15].

CNN menyediakan fondasi kuat bagi deteksi objek seperti *YOLO* melalui ekstraksi fitur multiskala dari data grid, di mana konvolusi MIMO memproses *multiple channel input* untuk menghasilkan representasi invariante terhadap deformasi geometris. Proses forward propagation melibatkan konvolusi, aktivasi, pooling, dan *fully-connected layer* untuk prediksi, sementara backward propagation mengoptimalkan *kernel* via *gradient descent*. Pendekatan ini memastikan efisiensi pada tugas *vision* seperti *bounding box prediction* di *YOLO* [16].

2.2 *YOLO11 (You Only Look Once Versi 11)*

Sejak diperkenalkan pada tahun 2016, algoritma *You Only Look Once (YOLO)* telah mengalami perkembangan signifikan dari *YOLOv1* hingga

YOLOv11. Versi awal, *YOLOv1*, memperkenalkan pendekatan deteksi objek satu tahap (*single-stage*) yang memungkinkan kecepatan inferensi tinggi. Pada *YOLOv2*, ditambahkan *anchor boxes* dan metode pelatihan *multi-scale* untuk meningkatkan ketepatan deteksi. Kemudian, *YOLOv3* memperkuat akurasi dengan arsitektur *Darknet-53* dan prediksi pada tiga skala berbeda. *YOLOv4* menyempurnakan model dengan *CSPDarknet-53*, penambahan *Feature Pyramid Networks (FPN)* dan *Path Aggregation Network (PAN)*, serta penggunaan fungsi kehilangan *Complete IoU (CIoU)*. Inovasi dilanjutkan oleh *YOLOv5*, yang dirilis oleh *Ultralytics* dengan basis *PyTorch* dan fitur *Dynamic Label Assignment (DLA)*. Pada *YOLOv6*, diperkenalkan *EfficientRep* sebagai *backbone* dan *decoupled head* untuk meningkatkan presisi. *YOLOv7* hadir dengan *E-ELAN* dan *RepConv* untuk efisiensi dan performa yang lebih baik. Kemudian, *YOLOv8* menambahkan dukungan *multitask* seperti *instance segmentation* dan *pose estimation*. *YOLOv9* memperkenalkan *Programmable Gradient Information (PGI)* dan *GELAN* untuk mengatasi *information bottleneck*. *YOLOv10* menggunakan arsitektur *One-to-Many* dan *One-to-One head* dalam proses pelatihan dan inferensi. Terakhir, *YOLOv11* mengintegrasikan blok *C3k2* dan modul *C2PSA* untuk memperkuat perhatian spasial dan meningkatkan fleksibilitas lintas tugas [17].

YOLO11, atau *You Only Look Once versi 11*, adalah versi terbaru dari model *YOLO* yang dirancang untuk meningkatkan kecepatan dan efisiensi deteksi objek. Model ini menggabungkan inovasi dalam arsitektur seperti *C3k2*, *SPPF*, dan *C2PSA*, yang secara signifikan meningkatkan akurasi dan efisiensi dalam

proses deteksi. Model ini banyak digunakan di berbagai bidang, termasuk *edge computing* dan sistem *AI* modern [3].

Alif (2024) juga menyatakan bahwa *YOLO11* menunjukkan kinerja yang bagus dalam mendeteksi benda-benda kecil dan yang sebagian tertutupi, hal ini cocok untuk aplikasi pendidikan seperti pengenalan objek sehari-hari dengan cepat dan tepat untuk anak-anak di sekolah dasar [11].

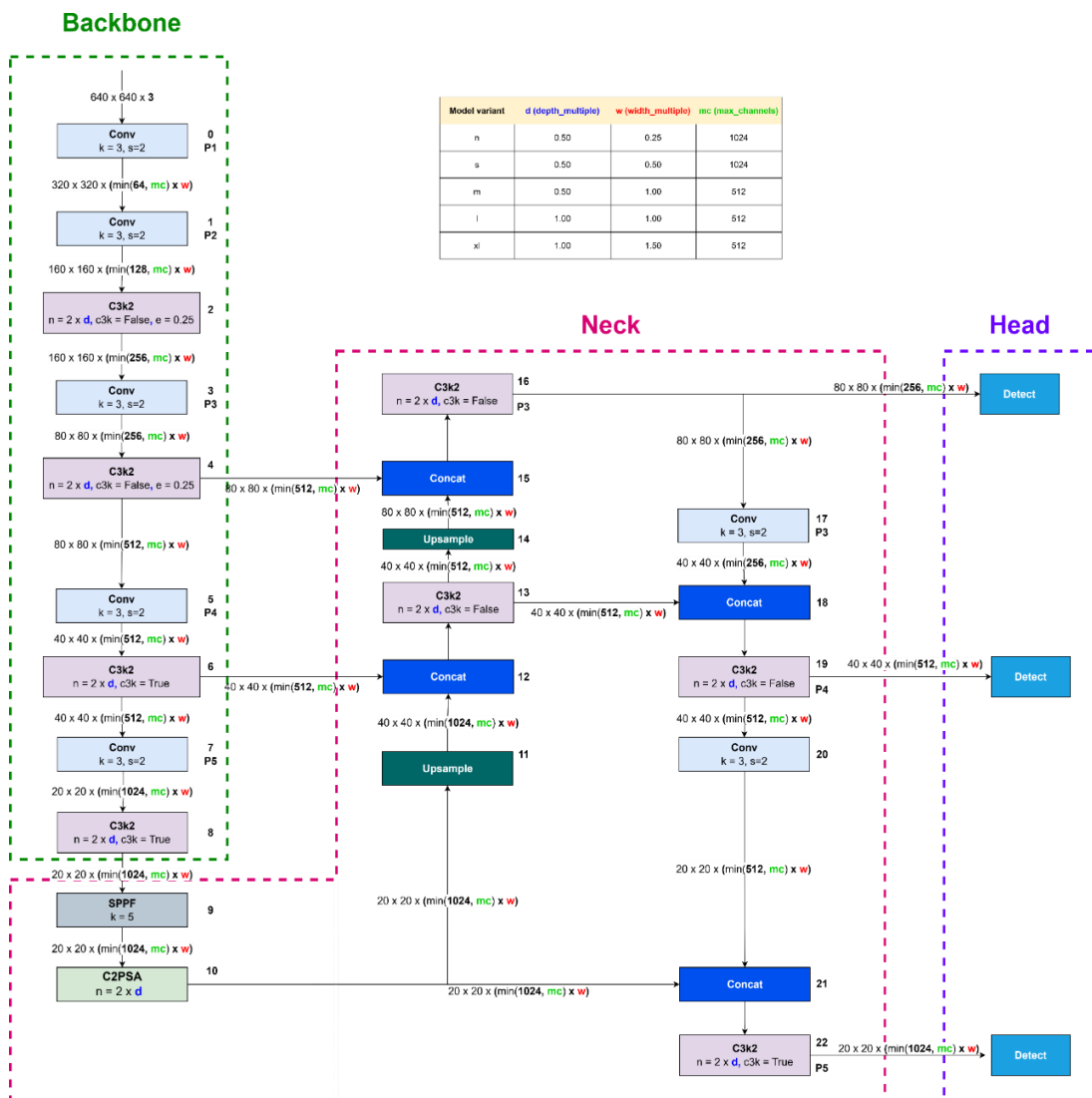
Ultralytics (2024) menjelaskan bahwa *YOLO11* adalah versi terbaru yang menggunakan lebih sedikit parameter dibandingkan versi sebelumnya seperti *YOLOv8*, namun tetap mempertahankan atau bahkan meningkatkan akurasi deteksi objek [18]. Hal ini membuatnya sangat cocok untuk sistem *lightweight*, seperti situs pembelajaran interaktif yang menyediakan keluaran teks dan audio dalam bahasa Inggris. Oleh karena itu, penggunaan *YOLO11* untuk mengembangkan situs deteksi objek bagi siswa sekolah dasar merupakan pilihan yang tepat dari segi performa dan efisiensi.

2.2.1 Arsitektur YOLO11

Arsitektur *YOLO11* terdiri dari tiga bagian utama yaitu *backbone* untuk mengekstraksi fitur gambar pada berbagai resolusi menggunakan blok konvolusi dan *C3K2*, *neck* yang menggabungkan fitur-fitur ini melalui *SPF* (*Spatial Pyramid Pooling Fast*), *C2PSA* (*Cross Stage Partial with Position Sensitive Attention*) untuk meningkatkan representasi spasial, serta *upsample* dan *concat layer* untuk menyatukan peta fitur; dan *head* yang memprediksi kelas dan *bounding box* objek berdasarkan fitur yang telah ditingkatkan resolusinya, dengan blok deteksi khusus untuk objek kecil, sedang, dan besar. Model ini memiliki lima

varian (*nano, small, medium, large, extra large*) yang disesuaikan oleh parameter *depth multiple*, *width multiple*, dan *max channels* [19].

Selain itu, dokumen *Ultralytics* (2024) resmi telah mengungkapkan bahwa arsitektur ini membuat *YOLO11* lebih fleksibel untuk berbagai tugas dalam *computer vision*, termasuk dalam deteksi objek, segmentasi dan pemantauan langsung [18]. Berikut Gambar Arsitektur *YOLO11* :



Gambar 2. 1 Arsitektur *YOLO11*. [19]

2.2.2 Keunggulan Untuk Deteksi *Real-Time*

Evaluasi menyeluruh terhadap *YOLOv5-v11* pada berbagai dataset mengungkap bahwa seluruh varian *YOLOv11* dari nano hingga large menawarkan kombinasi terbaik antara akurasi tinggi, kecepatan, dan efisiensi komputasi [20]. Pada pengujian deteksi objek alat di instalasi listrik, *YOLOv11* mencatat mAP tertinggi sebesar 57,2%, mengungguli *YOLOv10* dan pendahulunya [21]. Selain itu, untuk aplikasi agrikultur seperti pendeteksian dan penghitungan buah di lingkungan kebun, varian *YOLOv11n* mengungguli model pesaing dengan inferensi tercepat hanya 2,4 ms, ditambah akurasi *mAP50* mencapai 0,933 [22].

2.3 Implementasi *YOLO11*

Implementasi *YOLO11* menampilkan performa menjanjikan yang sangat baik pada berbagai bidang berdasarkan deteksi objek secara *real-time*. *YOLO11* berhasil dalam mendeteksi perilaku kecurangan saat ujian dengan akurasi yang tinggi dalam studi oleh Seung Jin Kim (2025), serta mencapai *mAP 0.991* dalam mengklasifikasikan enam perilaku berbeda, termasuklah menyontek serta berinteraksi selama ujian [23].

Penerapan lain juga terdapat pada klasifikasi citra sampah daur ulang untuk mendukung pengelolaan lingkungan yang lebih efisien. Dengan menggunakan dataset berisi 400 gambar dari empat kategori utama plastik, kaca, kardus, dan logam—model dilatih selama lima *epoch* dan mencapai akurasi tinggi sebesar 98,75%, *precision* 98,8%, *recall* 98,75%, serta *F1-score* 98,7% (25+source). Hasil ini menunjukkan kemampuan *YOLO11* dalam mengenali jenis sampah secara cepat dan akurat, serta kemampuannya untuk melakukan

generalisasi terhadap data baru, menjadikannya potensial untuk digunakan dalam sistem otomasi pemilahan sampah atau aplikasi edukatif berbasis citra [24].

Pendekatan serupa telah digunakan pula untuk pengawasan perilaku siswa di ruang kelas, seperti ditunjukkan oleh Chen *et al.* (2023) dengan performa deteksi meningkat hingga 4.2% melalui penggabungan modul multi-skala dan perhatian spasial [25]. Ketiga penelitian ini menegaskan fleksibilitas dan efektivitas *YOLO11* dalam berbagai konteks aplikasi, dari pendidikan hingga keamanan akademik dan edukasi sosial.

2.4 Teknologi Yang Akan Digunakan

2.4.1 Teknologi *Frontend*

Frontend adalah bagian dari antarmuka pengguna yang berfungsi untuk menunjukkan tampilan, mendapat masukan, dan menampilkan hasil deteksi dari bagian backend. Teknologi yang digunakan di *frontend* aplikasi ini mencakup:

1. *HTML5* dan *CSS3*

Keduanya digunakan untuk membuat struktur dasar halaman *web* serta merancang antarmuka yang responsif, mudah digunakan untuk anak-anak, dan menarik secara *visual*. *HTML5* memberikan elemen-elemen semantik yang ditingkatkan dan mendukung multimedia secara langsung, sementara *CSS3* digunakan untuk mempercantik elemen-elemen tersebut [26] [27].

2. *JavaScript (Vanilla)*

Digunakan untuk mengelola interaksi dengan pengguna, seperti mengambil foto dari kamera, mengupload *file*, dan mengirim gambar ke

backend melalui *fetch API*. Selain itu, *JavaScript* juga berfungsi untuk menggambar hasil deteksi objek di elemen *canvas* [28].

3. *Canvas API*

Bounding box dan label objek yang terdeteksi dapat digambar menggunakan *canvas*. *API* ini memungkinkan pengaturan grafis 2D dengan cara yang dinamis di dalam peramban [29].

4. *Web Speech API*

API ini berfungsi sebagai alat untuk pelafalan objek dalam bahasa Inggris. Ini dimanfaatkan untuk membantu anak-anak belajar kosakata melalui cara yang interaktif [30].

2.4.2 **Teknologi Backend**

Backend adalah bagian yang mengatur pemrosesan data yang berasal dari *frontend*. Tugasnya termasuk menerima gambar, melakukan deteksi objek dengan menggunakan model pembelajaran mesin, dan mengirimkan hasilnya kembali dalam forma *JSON*. Beberapa teknologi yang akan dipakai meliputi:

1. *Flask*

Adalah sebuah *framework web* kecil yang dibuat dengan *Python*, yang berfungsi untuk membuat titik akhir *API* seperti *index* dan *detect*. *Flask* memiliki ukuran kecil, sangat fleksibel, dan ideal untuk keperluan prototipe maupun produksi [31].

2. *Ultralytics YOLO11*

Model *YOLO11*, atau *You Only Look Once* versi 11, digunakan untuk mengenali objek dalam gambar. Model ini memberikan kecepatan dan tingkat ketepatan tinggi serta memungkinkan inferensi secara *real-time*. Terdapat beberapa variasi model *pre-training* yang lebih ringan seperti *yolo11n.pt*, yang ideal untuk perangkat dengan kekurangan daya komputasi [32].

3. *OpenCV*

Ini digunakan untuk mengubah gambar dan mengubah data citra ke dalam format yang cocok agar bisa diproses oleh model *YOLO11* [33].

4. *NumPy*

Dipakai untuk mengolah *array* gambar yang berada dalam format angka sebelum memasukkannya ke dalam model deteksi [34].

5. *Pillow (PIL Fork)*

Pillow digunakan untuk membuka dan memproses berkas gambar, termasuk gambar yang diambil dari permintaan *HTTP* dengan bantuan *library* lain seperti *requests*, dan dapat dikonversi menjadi *array NumPy* untuk proses deteksi lebih lanjut [35].

2.5 Aspek Edukasi dalam Sistem Interaktif Anak Usia Sekolah Dasar

2.5.1 Bahasa Inggris sebagai Bahasa Kedua (*Second Language Acquisition*)

Bahasa Inggris sebagai bahasa kedua (*Second Language Acquisition- SLA*) merupakan proses kompleks yang dipengaruhi oleh berbagai faktor seperti usia, lingkungan, dan media pembelajaran. Kusumawardani dan Anjaniputra (2024) melalui studi sistematisnya menekankan bahwa pembelajaran bahasa Inggris pada

anak-anak di Indonesia membutuhkan strategi dan media yang sesuai agar akuisisi berjalan efektif, terutama di konteks pembelajaran dasar [36]

Rohmah (2016) menunjukkan bahwa selain aktivitas sekolah, media interaktif seperti permainan video *online* dapat memperkuat perolehan bahasa karena melibatkan keterlibatan emosional dan linguistik secara simultan [37]. Sementara itu, studi kasus oleh Ilma *et al.* (2024) mengungkap bahwa anak usia enam tahun yang belajar bahasa Inggris mengalami kesulitan dalam struktur sintaksis, seperti penggunaan kata kerja bantu dan bentuk jamak, sehingga konsistensi dan pendekatan berbasis konteks sangat dibutuhkan [38]. Ketiga studi ini menunjukkan bahwa *SLA* pada anak dapat ditingkatkan melalui pendekatan multimodal yang mencakup aktivitas nyata, media digital, dan lingkungan yang mendukung pembelajaran.

2.5.2 Peran Teks dalam Pembelajaran Anak

Teks memiliki peran penting dalam pembelajaran anak sekolah dasar karena mampu meningkatkan kemampuan literasi seperti membaca, menulis, dan berbicara. Rahmawati *et al.* (2019) menunjukkan bahwa pembelajaran berbasis teks secara signifikan meningkatkan hasil belajar siswa kelas V melalui pendekatan saintifik dan metode yang variatif [39].

Farida dan Nugraheni (2023) juga membuktikan bahwa membaca teks karangan dapat meningkatkan kemampuan berbicara siswa kelas III, termasuk kepercayaan diri dan kelancaran dalam menyampaikan pendapat [40]. Selain itu, Angely dan Suriani (2024) menemukan bahwa penggunaan model *Discovery Learning* dalam menulis teks eksplanasi membantu siswa memahami struktur teks

dan meningkatkan kualitas tulisan mereka [41]. Temuan ini menegaskan bahwa integrasi teks dalam pembelajaran mampu menunjang perkembangan keterampilan bahasa secara menyeluruh pada siswa sekolah dasar.

2.5.3 Peran Audio dalam Pembelajaran Interaktif

Penggunaan audio dalam pembelajaran interaktif terbukti efektif dalam meningkatkan motivasi dan hasil belajar siswa sekolah dasar. Media audio-visual yang menggabungkan elemen suara dan gambar mampu menarik perhatian, meningkatkan pemahaman konsep, serta memperkuat daya ingat siswa terhadap materi yang disampaikan.

Penelitian oleh Sofiana *et al.* (2023) menunjukkan adanya peningkatan motivasi belajar siswa IPA setelah diterapkannya media audio-visual [42]. Sementara Farhiza *et al.* (2023) menemukan bahwa skor motivasi belajar siswa yang menggunakan media audio-visual lebih tinggi dibandingkan kelompok kontrol [43]. Selain itu, Putra *et al.* (2022) juga membuktikan bahwa pengembangan media audio-visual mampu meningkatkan kemampuan menulis siswa secara signifikan [44].

2.5.4 Desain Antarmuka dan Pengalaman Pengguna (UI/UX) untuk Anak Sekolah Dasar

Desain antarmuka pengguna (UI) yang menyenangkan untuk anak dan pengalaman pengguna (UX) yang baik adalah elemen penting dalam menciptakan aplikasi pembelajaran bagi siswa di sekolah dasar. Penelitian yang dilakukan oleh Putri (2022) dan rekan-rekannya menyoroti betapa pentingnya elemen visual yang menarik, kemudahan dalam navigasi, dan interaksi yang mudah dipahami dalam

aplikasi pendidikan untuk anak-anak usia dini, dengan *prototype* yang dibuat menggunakan *Figma* [45].

Di sisi lain, penelitian oleh Sari dan tim (2022) menerapkan metode *Design Thinking* untuk merancang aplikasi pembelajaran bahasa Inggris bagi anak-anak di sekolah dasar, yang menghasilkan nilai pada *System Usability Scale* (*SUS*) mencapai 83,75, dan ini termasuk dalam kategori "*Excellent*" [46]. Selain itu, Fitra *et al.* (2023) menerapkan pendekatan *Human-Centered Design* dalam menciptakan permainan edukasi untuk menulis dalam bahasa Indonesia, yang dirancang khusus untuk memenuhi kebutuhan kognitif dan emosional anak-anak [47]. Hasil dari penelitian ini menunjukkan bahwa penggunaan prinsipal desain yang fokus pada anak dapat meningkatkan efektivitas dan kenyamanan dalam belajar secara digital.

2.7 Perancangan Sistem

2.7.1 Unified Modeling Language (UML)

Unified Modeling Language (UML) akan diterapkan untuk memodelkan sistem deteksi objek yang berbasis *YOLO11*, menghasilkan *output* berupa teks dan audio berbahasa Inggris dengan cara yang terstruktur dan sistematis. Model ini meliputi *diagram use case* yang menunjukkan hubungan antara pengguna dan sistem, diagram aktivitas (*Activity Diagram*) untuk menggambarkan proses dari *input* gambar ke keluaran suara, serta diagram urutan (*Sequence Diagram*) yang menjelaskan urutan interaksi antara komponen sistem seperti modul deteksi, penerjemah label menjadi teks, dan *text-to-speech*.

Metode ini membantu *programmer* memahami keseluruhan alur kerja sistem dan memudahkan proses pengembangan serta pemecahan masalah. Studi oleh Koç *et al.* mengindikasikan bahwa penerapan *UML* dapat memperbaiki pemahaman tim terhadap sistem berbasis *AI* secara visual [48]. Gadhi *et al.* (2023) menyatakan bahwa penggabungan deteksi objek dengan *TTS* memerlukan struktur modular yang bisa divisualisasikan dengan baik menggunakan *UML* [49]. Sementara itu, Bates *et al.* menunjukkan bahwa *UML* sangat efektif dalam pengembangan sistem *deep learning* dengan *output multimodal*, karena mampu mengilustrasikan hubungan dan proses antar komponen secara jelas [50].

2.7.2 Model Waterfall

Metode *waterfall* dalam pengembangan perangkat lunak telah sering digunakan dalam berbagai proyek sistem informasi karena pendekatannya yang terstruktur dan sistematis. Misalnya, studi oleh Tri Hartati *et al.* (2021) menggunakan model *waterfall* untuk merancang Sistem Informasi Penilaian Kinerja Pegawai (SI-PKP) di Direktorat Jenderal ILMATE Kementerian Perindustrian Jakarta, yang mencakup tahap analisis kebutuhan, desain, implementasi, pengujian, dan pemeliharaan [51].

Selanjutnya, Khairul Abidin *et al.* (2023) menerapkan metode *waterfall* dalam pembuatan sistem pendataan peternakan berbasis web di Kabupaten Kampar, dengan maksud untuk mempermudah pengelolaan data peternakan secara digital. [52]. Di samping itu, Pengembangan aplikasi manajemen keuangan menggunakan model *Waterfall* oleh Fernanda *et al.* (2025) berhasil menghasilkan aplikasi yang memiliki fitur lengkap, seperti *login*, *dashboard*, pencatatan

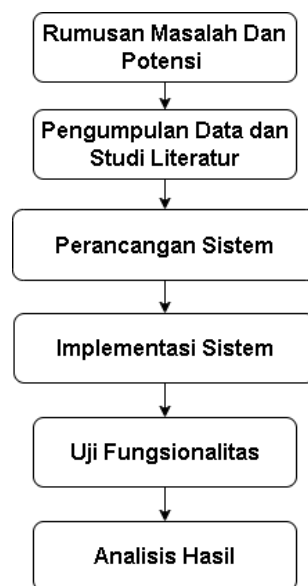
transaksi (pemasukan dan pengeluaran), pengelolaan investasi, laporan keuangan dan investasi, serta pengaturan profil. Pengujian sistem dilakukan menggunakan metode *blackbox* testing, yang menunjukkan bahwa seluruh fitur berjalan dengan baik dan sesuai spesifikasi [53].

Ketiga penelitian ini menunjukkan bahwa pendekatan *waterfall* efektif untuk pengembangan sistem perangkat lunak yang memerlukan proses terstruktur dan dokumentasi yang komprehensif.

BAB 3

METODOLOGI PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan yang saling berhubungan dan terintegrasi, setiap tahapan dijelaskan secara komprehensif dalam bagian metodologi ini, yang disusun secara sistematis dan terstruktur, untuk memberikan gambaran keseluruhan alur penelitian dari awal hingga akhir. Sebagaimana digambarkan dalam diagram alir tahapan penelitian berikut :



Gambar 3. 1 Tahapan Penelitian

3.1 Rumusan Masalah dan Potensi

Tahap ini bertujuan untuk mengidentifikasi permasalahan utama, yaitu kurangnya media pembelajaran bahasa Inggris yang interaktif berbasis teknologi untuk siswa sekolah dasar. Selain itu, dilakukan juga analisis potensi pemanfaatan

teknologi seperti deteksi objek dan *text-to-speech* sebagai solusi pembelajaran yang menarik dan edukatif.

3.2 Pengumpulan Data dan Studi Literatur

Pengumpulan data dilakukan melalui studi pustaka. Studi pustaka ini mencakup referensi dari jurnal, artikel ilmiah, dan dokumentasi teknis yang berkaitan dengan teknologi yang digunakan seperti *YOLO11*, *Flask*, dan *text-to-speech*. Selain itu, studi pustaka juga dilakukan untuk memahami karakteristik dan kebutuhan pembelajaran bahasa Inggris pada anak-anak sekolah dasar, yang menjadi dasar perancangan sistem ini.

3.3 Perancangan Sistem

Perancangan sistem meliputi pembuatan desain antarmuka pengguna (*UI*), pelatihan model dengan dataset kostum, arsitektur sistem, alur proses aplikasi menggunakan *Unified Modeling Language (UML)* digunakan sebagai bahasa pemodelan visual untuk memodelkan struktur dan perilaku sistem secara terstruktur dan sistematis. Perancangan ini bertujuan agar sistem mudah digunakan oleh target pengguna, yaitu anak-anak sekolah dasar.

3.4 Implementasi Sistem

Setelah perancangan selesai, dilakukan implementasi sistem menggunakan bahasa pemrograman *Python* dengan *framework Flask*. Model *YOLO11* digunakan untuk mendeteksi objek dari gambar kamera, lalu hasilnya ditampilkan sebagai teks Indonesia, tombol audio dengan tulisan Inggris sesuai nama objek yang di deteksi dan disuarakan menggunakan modul audio yang ada di sistem dan *text-to-speech* berbahasa Inggris serta tombol dengan fitur koreksi pengucapan.

3.5 Uji Fungsionalitas

Tahap ini dilakukan untuk memastikan bahwa sistem berjalan sesuai fungsi yang telah dirancang. Pengujian dilakukan terhadap proses deteksi objek, keluaran teks, dan hasil audio serta koreksi pengucapan. Tujuannya adalah mengevaluasi apakah sistem dapat digunakan untuk deteksi objek pada aplikasi web.

3.6 Analisis Hasil

Hasil pengujian dianalisis untuk menilai keberhasilan sistem dalam memenuhi kebutuhan pengguna. Analisis dilakukan terhadap keakuratan deteksi objek, keterbacaan teks, fungsi koreksi pengucapan serta kejelasan audio yang dihasilkan.