**LAPORAN**

**** **IMPLEMENTASI METODE SSD DAN LSTM UNTUK DETEKSI REAL-TIME DALAM KLASIFIKASI BAHASA ISYARAT TANGAN BERBASIS CITRA DIGITAL**

**RISET INFORMATIKA D081**

**Dosen pengampu :**

Dr. Basuki Rahmat, S.Si., M.T.

**Dibuat Oleh:**

21081010258 Satria Farras Athallansyah

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UPN “VETERAN” JAWA TIMUR**

**2024**

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang**

Bahasa isyarat adalah salah satu bentuk komunikasi yang digunakan oleh individu dengan gangguan pendengaran atau bicara. Bahasa ini mengandalkan kombinasi gerakan tangan, ekspresi wajah, dan postur tubuh untuk menyampaikan informasi. Meskipun bahasa isyarat telah diakui sebagai alat komunikasi resmi di berbagai negara, tingkat pemahaman masyarakat umum terhadap bahasa ini masih terbatas. Akibatnya, individu yang bergantung pada bahasa isyarat sering menghadapi kendala dalam berkomunikasi dengan komunitas luas.

Dalam beberapa tahun terakhir, kemajuan teknologi kecerdasan buatan (Artificial Intelligence/AI) telah membuka peluang baru untuk menjembatani kesenjangan ini. Salah satu pendekatan yang menjanjikan adalah penggunaan teknologi pengenalan bahasa isyarat berbasis citra digital. Teknologi ini dapat memanfaatkan video atau gambar untuk mendeteksi, mengenali, dan menerjemahkan gerakan tangan ke dalam teks atau suara yang dapat dipahami oleh masyarakat umum.

Namun, tantangan utama dalam pengembangan sistem pengenalan bahasa isyarat terletak pada kebutuhan untuk menangani gerakan tangan yang kompleks dan dinamis secara real-time. Setiap gerakan tangan dalam bahasa isyarat memiliki ciri khas tertentu yang melibatkan elemen temporal (urutan waktu) dan spasial (posisi dan orientasi dalam ruang). Oleh karena itu, diperlukan metode yang mampu mengolah data citra secara efisien sambil mempertimbangkan hubungan temporal antar-frame video.

Metode *Single Shot Multibox Detector* (SSD) adalah salah satu algoritme deep learning yang telah terbukti efektif dalam deteksi objek real-time. SSD mampu mendeteksi objek dengan berbagai ukuran dalam gambar atau video dengan kecepatan tinggi, sehingga cocok untuk digunakan dalam mendeteksi posisi tangan pada video bahasa isyarat. Di sisi lain, *Long Short-Term Memory* (LSTM) merupakan varian jaringan saraf tiruan (recurrent neural network) yang dirancang untuk menangkap pola temporal dalam data sekuensial. Dengan mengintegrasikan kedua metode ini, dimungkinkan untuk menciptakan sistem yang dapat mendeteksi posisi tangan secara real-time menggunakan SSD dan menganalisis urutan gerakan menggunakan LSTM untuk mengklasifikasikan bahasa isyarat.

Penelitian tentang integrasi metode SSD dan LSTM masih relatif baru, tetapi potensi penerapannya sangat luas. Di bidang komunikasi, sistem ini dapat digunakan untuk menerjemahkan bahasa isyarat secara langsung, membantu individu dengan gangguan pendengaran berinteraksi dengan masyarakat umum. Di bidang pendidikan, teknologi ini dapat menjadi alat pembelajaran bagi siswa dan guru untuk mempelajari bahasa isyarat secara efektif. Selain itu, di bidang pelayanan publik, sistem pengenalan bahasa isyarat dapat diterapkan dalam layanan informasi, seperti bank, rumah sakit, dan transportasi umum.

Melihat potensi tersebut, penelitian ini bertujuan untuk mengembangkan sistem deteksi dan klasifikasi bahasa isyarat berbasis citra digital dengan memanfaatkan metode SSD dan LSTM. Sistem ini dirancang untuk bekerja secara real-time, sehingga dapat digunakan dalam aplikasi praktis sehari-hari. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam pengembangan teknologi berbasis AI, tetapi juga memiliki dampak sosial yang signifikan dalam meningkatkan inklusivitas komunikasi.

Di samping itu, pendekatan ini juga relevan dengan tren teknologi terkini yang berfokus pada pengembangan sistem AI yang lebih cerdas, efisien, dan mampu berinteraksi secara alami dengan manusia. Dengan menggunakan dataset yang representatif dan pendekatan yang terstruktur, penelitian ini diharapkan mampu menghasilkan sistem pengenalan bahasa isyarat yang akurat, cepat, dan dapat diandalkan untuk digunakan di berbagai situasi.

**1.2 Rumusan Masalah**

1. Bagaimana metode SSD dapat digunakan untuk mendeteksi posisi tangan secara real-time?
2. Bagaimana metode LSTM dapat memanfaatkan data temporal dari gerakan tangan untuk mengklasifikasikan bahasa isyarat?
3. Bagaimana kinerja kombinasi metode SSD dan LSTM dalam mendeteksi dan mengklasifikasikan bahasa isyarat tangan secara real-time?

**1.3 Tujuan Penelitian**

1. Mengimplementasikan metode SSD untuk deteksi posisi tangan secara cepat dan akurat.
2. Menggunakan metode LSTM untuk klasifikasi gerakan bahasa isyarat berdasarkan data temporal.
3. Mengevaluasi kinerja gabungan metode SSD dan LSTM dalam aplikasi real-time.

**1.4 Manfaat Penelitian**

* Memberikan solusi teknologi bagi individu dengan keterbatasan pendengaran.
* Memperluas penelitian tentang integrasi SSD dan LSTM untuk aplikasi berbasis citra digital.
* Mendorong perkembangan sistem klasifikasi berbasis AI dalam bidang inklusi sosial.

**BAB II**

**TINJAUAN PUSTAKA**

**2.1 Bahasa Isyarat dan Kebutuhan Teknologi Pendukung**

Bahasa isyarat adalah media komunikasi visual yang menggunakan gerakan tangan, ekspresi wajah, dan posisi tubuh. Setiap bahasa isyarat memiliki sistem tata bahasa yang unik dan bervariasi antar negara. Sebagai contoh, American Sign Language (ASL) memiliki struktur yang berbeda dari Bahasa Isyarat Indonesia (BISINDO). Meskipun penting sebagai sarana komunikasi, keterbatasan pemahaman masyarakat luas terhadap bahasa isyarat menjadi tantangan besar dalam interaksi sosial.

Seiring perkembangan teknologi, berbagai upaya dilakukan untuk menciptakan alat bantu komunikasi bagi pengguna bahasa isyarat. Teknologi pengenalan bahasa isyarat berbasis citra digital bertujuan untuk menerjemahkan gerakan tangan menjadi teks atau suara yang dapat dipahami oleh masyarakat umum. Hal ini dapat dilakukan dengan mengintegrasikan teknik pengolahan citra, pembelajaran mesin, dan kecerdasan buatan.

Beberapa teknologi yang telah dikembangkan sebelumnya meliputi penggunaan perangkat keras seperti sarung tangan sensorik untuk merekam gerakan tangan. Namun, pendekatan ini memiliki keterbatasan, termasuk ketidakpraktisan penggunaan dan biaya produksi yang tinggi. Oleh karena itu, pendekatan berbasis citra digital menjadi pilihan yang lebih fleksibel dan efisien, karena hanya membutuhkan kamera untuk menangkap gerakan tangan

**2.2 Metode Single Shot Multibox Detector (SSD)**

**2.2.1 Konsep Dasar SSD**

SSD adalah algoritme deteksi objek yang dirancang untuk memproses data secara real-time dengan efisiensi tinggi. Berbeda dengan metode deteksi objek tradisional yang memisahkan proses seleksi wilayah dan klasifikasi objek, SSD menggabungkan keduanya dalam satu langkah dengan menggunakan *Convolutional Neural Network* (CNN). Proses ini melibatkan:

1. **Ekstraksi Fitur**: SSD menggunakan jaringan CNN untuk mengekstrak fitur dari gambar input.
2. **Prediksi Bounding Box dan Kategori**: SSD membuat prediksi langsung untuk posisi (*bounding box*) dan kelas objek pada setiap skala fitur.

SSD menggunakan arsitektur multiskala yang memungkinkan deteksi objek pada berbagai ukuran, sehingga cocok untuk mendeteksi gerakan tangan dalam video yang memiliki variasi posisi dan skala

**2.2.2 Keunggulan SSD**

* Efisiensi dalam pemrosesan real-time.
* Deteksi multiscale yang memungkinkan identifikasi objek kecil dan besar.

**2.2.3 Penerapan SSD pada Pendeteksian Bahasa Isyarat**

Dalam pengenalan bahasa isyarat, SSD digunakan untuk mendeteksi posisi tangan dalam setiap frame video. Proses ini melibatkan pelatihan model pada dataset yang berisi gambar tangan dalam berbagai posisi dan latar belakang. Output dari SSD berupa koordinat bounding box yang menentukan lokasi tangan

**2.3 Metode Long Short-Term Memory (LSTM)**

**2.3.1 Konsep Dasar LSTM**

LSTM adalah varian dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi masalah *vanishing gradient* dalam pelatihan data sekuensial. LSTM menggunakan tiga gerbang utama:

1. **Input Gate**: Mengontrol informasi baru yang akan ditambahkan ke sel memori.
2. **Forget Gate**: Menentukan informasi lama yang harus dihapus.
3. **Output Gate**: Mengatur informasi yang akan digunakan pada waktu tertentu

**2.3.2 Keunggulan LSTM**

* Kemampuan untuk mempelajari pola temporal.
* Efektif dalam analisis data sekuensial seperti video.

**2.3.3 Penerapan LSTM pada Klasifikasi Bahasa Isyarat**

LSTM digunakan untuk menganalisis data urutan gerakan tangan yang telah dideteksi oleh SSD. Input berupa urutan koordinat tangan dari beberapa frame video, sementara output berupa label kelas yang mewakili kata atau frasa dalam bahasa isyarat.

**2.4 Penelitian Terkait**

1. Studi A: Penggunaan SSD dalam deteksi objek secara real-time pada aplikasi pengawasan.
2. Studi B: Implementasi LSTM untuk analisis video dalam aplikasi olahraga.
3. Studi C: Integrasi SSD dan LSTM untuk pengenalan aktivitas manusia berbasis video.

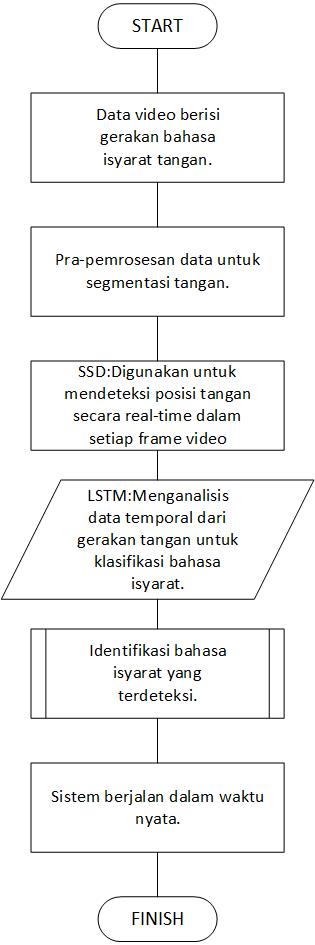
**2.5 Gap Penelitian**

Penelitian sebelumnya cenderung berfokus pada aplikasi individu SSD atau LSTM dalam berbagai domain. Namun, integrasi kedua metode ini untuk mendeteksi dan mengklasifikasikan bahasa isyarat secara real-time masih jarang dieksplorasi. Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan sistem yang mampu memanfaatkan keunggulan SSD dalam deteksi objek dan LSTM dalam analisis temporal untuk menciptakan solusi yang cepat, akurat, dan aplikatif

**BAB II**

**TINJAUAN PUSTAKA**

#### **3.1 Pendekatan Penelitian**

Penelitian ini menggunakan pendekatan **eksperimen** berbasis **data sekunder** dengan fokus pada pengembangan dan pengujian sistem deteksi dan klasifikasi bahasa isyarat tangan secara real-time. Pendekatan ini melibatkan integrasi metode **Single Shot Multibox Detector (SSD)** untuk deteksi objek dan **Long Short-Term Memory (LSTM)** untuk analisis temporal data sekuensial. Setiap langkah penelitian dirancang untuk memastikan keberhasilan implementasi algoritme dalam mendeteksi posisi tangan dan mengklasifikasikan gerakan bahasa isyarat secara akurat dan efisien. Dan berikut mind map dari laporan ini

#### **3.2 Alur Penelitian**

Proses penelitian terdiri dari enam tahapan utama, yaitu:

1. **Pengumpulan Data**
   * **Jenis Data**: Dataset berupa gambar dan video yang memuat gerakan bahasa isyarat. Data ini mencakup berbagai pose tangan dengan latar belakang, pencahayaan, dan sudut pandang yang bervariasi.
   * **Dataset Publik**: Digunakan dataset seperti **American Sign Language (ASL)** dan **Bahasa Isyarat Indonesia (BISINDO)**, yang memuat ratusan kelas gerakan tangan.
   * **Pengelompokan Data**: Data dibagi menjadi dua kategori utama:
     + **Data Pelatihan**: 80% dataset digunakan untuk melatih model.
     + **Data Pengujian**: 20% dataset digunakan untuk mengukur kinerja model.
2. **Preprocessing Data**
   * **Normalisasi**: Semua piksel gambar dinormalisasi ke rentang nilai 0-1 untuk meningkatkan stabilitas pelatihan model.
   * **Resizing**: Gambar diubah ukurannya menjadi 300x300 piksel untuk kompatibilitas dengan SSD.
   * **Augmentasi Data**: Data pelatihan diperluas dengan teknik augmentasi seperti flipping, rotation, dan brightness adjustment untuk meningkatkan robusta model terhadap variasi kondisi nyata.
   * **Pelabelan Data**:
     + Untuk SSD: Bounding box ditambahkan untuk menandai lokasi tangan pada setiap frame.
     + Untuk LSTM: Urutan bounding box diurutkan secara temporal untuk mewakili gerakan.
3. **Pengembangan Model Deteksi dengan SSD**
   * **Arsitektur SSD**:
     + Backbone model: **MobileNetV2** untuk efisiensi komputasi.
     + Lapisan tambahan: Digunakan untuk menghasilkan prediksi multiskala.
   * **Hyperparameter**:
     + Learning rate: 0.001
     + Optimizer: Adam
     + Epochs: 50
   * **Pelatihan**:
     + Input berupa gambar yang berisi pose tangan.
     + Output berupa koordinat bounding box dan label kelas objek (tangan).
4. **Pengembangan Model Klasifikasi dengan LSTM**
   * **Arsitektur LSTM**:
     + Dua lapisan LSTM dengan 256 unit per lapisan.
     + Dropout rate: 20% untuk mengurangi overfitting.
     + Fully connected layer di akhir untuk mengklasifikasikan gerakan.
   * **Hyperparameter**:
     + Sequence length: 10 frame untuk setiap urutan gerakan.
     + Optimizer: RMSProp
   * **Pelatihan**:
     + Input berupa koordinat bounding box yang dihasilkan oleh SSD.
     + Output berupa label gerakan (kelas bahasa isyarat).
5. **Integrasi Model SSD dan LSTM**
   * SSD menghasilkan bounding box dan label posisi tangan dari setiap frame video.
   * Output SSD diproses sebagai input LSTM untuk analisis temporal dan klasifikasi gerakan.
   * Sistem mengintegrasikan kedua metode ini untuk menghasilkan prediksi bahasa isyarat secara real-time.
6. **Evaluasi Kinerja Sistem**
   * Kinerja sistem dievaluasi menggunakan beberapa metrik:
     + **Mean Average Precision (mAP)** untuk mengukur performa deteksi objek.
     + **Akurasi** untuk mengukur keberhasilan klasifikasi gerakan.
     + **Frame Per Second (FPS)** untuk menilai kecepatan sistem dalam skenario real-time.

#### **3.3 Arsitektur Sistem**

Sistem dirancang dalam tiga tahap utama:

1. **Input Video**  
   Video diambil melalui kamera dan diproses frame by frame.
2. **Deteksi Posisi Tangan dengan SSD**  
   SSD mendeteksi tangan pada setiap frame dan menghasilkan bounding box.
3. **Klasifikasi Gerakan dengan LSTM**  
   LSTM menerima bounding box sebagai input, memproses urutan temporal, dan menghasilkan prediksi kelas bahasa isyarat.

#### **3.4 Implementasi Metode**

##### 3.4.1 Implementasi SSD

* **Pretrained Model**: SSD menggunakan backbone MobileNetV2 yang telah dilatih sebelumnya pada dataset COCO, kemudian di-tuning dengan dataset bahasa isyarat.
* **Output Model**: Bounding box dengan confidence score untuk mendeteksi posisi tangan.

##### 3.4.2 Implementasi LSTM

* **Input Data**: Data sekuensial berupa bounding box tangan.
* **Output Model**: Prediksi kelas gerakan tangan, seperti huruf, angka, atau kata dalam bahasa isyarat.

#### **3.5 Perangkat Pengujian**

1. **Perangkat Keras**
   * GPU: NVIDIA RTX 3060 untuk mempercepat pelatihan dan inferensi model.
   * CPU: Intel Core i7 untuk pengolahan data umum.
   * RAM: 16GB untuk mendukung pemrosesan data besar.
2. **Perangkat Lunak**
   * **Framework**: TensorFlow 2.x, Keras.
   * **Library**: OpenCV untuk pemrosesan video, Matplotlib untuk visualisasi hasil.

#### **3.6 Validasi Sistem**

1. **Pengujian Akurasi**  
   Dilakukan pada dataset pengujian untuk menghitung mAP SSD dan akurasi klasifikasi LSTM.
2. **Pengujian Real-Time**  
   Mengukur FPS pada video dengan variasi pencahayaan, latar belakang, dan jumlah gerakan tangan.
3. **Pengujian Robustness**
   * Menguji sistem dengan noise pada gambar/video.
   * Mengukur ketahanan sistem terhadap perubahan lingkungan nyata.

#### **3.7 Analisis Data**

Data hasil pengujian dianalisis untuk:

1. Menilai keunggulan integrasi SSD dan LSTM dibandingkan metode lain.
2. Mengidentifikasi kelemahan sistem, seperti kesalahan pada kondisi latar belakang kompleks.
3. Memberikan rekomendasi untuk pengembangan lebih lanjut.

#### **3.8 Output yang Diharapkan**

1. **Sistem Deteksi dan Klasifikasi Real-Time**
   * Deteksi posisi tangan dengan bounding box.
   * Klasifikasi gerakan tangan menjadi label bahasa isyarat.
2. **Laporan Evaluasi Kinerja**
   * Tingkat akurasi.
   * Kecepatan pemrosesan (FPS).
   * Studi kasus penerapan dalam skenario nyata.