

**PROPOSAL TUGAS AKHIR – IF184702**

**Pengenalan Video Peraga Menggunakan Transformer untuk Pembelajaran Bahasa Isyarat**

**Adam Hadi Prasetyo**

NRP 05111940000224

Dosen Pembimbing 1

**Dr. Eng. Nanik Suciati, S.Kom, M.Kom**

NIP 197104281994122001

Dosen Pembimbing 2

**Anny Yuniarti, S.Kom., M.Comp.Sc.**

NIP 198106222005012002

# HALAMAN JUDUL

**Program Studi S1 Teknik Informatika**

Departemen Teknik Informatika

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Surabaya

2022

# LEMBAR PENGESAHAN

**PENGENALAN VIDEO PERAGA MENGGUNAKAN TRANSFORMER  
 UNTUK PEMBELAJARAN BAHASA ISYARAT**

**PROPOSAL TUGAS AKHIR**

Diajukan untuk memenuhi salah satu syarat

Memperoleh gelar Sarjana Komputer pada

Program Studi S-1 Teknik Informatika

Departemen Teknik Informatika

Fakultas Teknologi Elektro dan Informatika Cerdas

Institut Teknologi Sepuluh Nopember

Oleh : **Adam Hadi Prasetyo**

NRP. 05111940000224

Disetujui oleh Tim Penguji Proposal Tugas Akhir:

|  |  |  |
| --- | --- | --- |
| 1. | Dr. Eng. Nanik Suciati, S.Kom, M.Kom | Pembimbing |
| 2. | Anny Yuniarti, S.Kom., M.Comp.Sc. | Ko-pembimbing |
| 3. | Nama dan gelar penguji | Penguji |
| 4. | Nama dan gelar penguji | Penguji |
| 5. | Nama dan gelar penguji | Penguji |

**SURABAYA**

**Desember, 2022**

**APPROVAL SHEET**

**RECOGNITION TO VIDEO DEMONSTRATIONS USING   
TRANSFORMERS FOR SIGN LANGUAGE LEARNING**

**FINAL PROJECT PROPOSAL**

Submitted to fulfill one of the requirements

for obtaining a degree Computer Science at

Undergraduate Study Program of Informatics Engineering

Department of Informatics

Faculty of Intelligent Electrical and Information Technology

Institut Teknologi Sepuluh Nopember

By: **Adam Hadi Prasetyo**

NRP. 05111940000224

Approved by Final Project Proposal Examiner Team:

|  |  |  |
| --- | --- | --- |
| 1. | Dr. Eng. Nanik Suciati, S.Kom, M.Kom | Advisor |
| 2. | Anny Yuniarti, S.Kom., M.Comp.Sc. | Co-Advisor |
| 3. | Name of Examiner and academic title | Examiner |
| 4. | Name of Examiner and academic title | Examiner |
| 5. | Name of Examiner and academic title | Examiner |

**SURABAYA**

**December, 2022**

**PENGENALAN VIDEO PERAGA MENGGUNAKAN TRANSFORMER   
UNTUK PEMBELAJARAN BAHASA ISYARAT**

**Nama Mahasiswa / NRP : Adam Hadi Prasetyo / 05111940000224**

**Departemen : Teknik Informatika FTEIC - ITS**

**Dosen Pembimbing :** **Dr. Eng. Nanik Suciati, S.Kom, M.Kom**

# ABSTRAK

**Abstrak**

Dalam kehidupan sehari-hari, komunikasi merupakan aktivitas manusia yang paling sering dilakukan. Komunikasi bertujuan untuk menyampaikan maksud dalam bentuk lisan, tulisan, gambar atau gestur. Komunikasi dengan gestur biasa disebut bahasa isyarat. Pada umumnya bahasa isyarat digunakan karena keterbatasan fisik yang tidak memungkinkan penggunaan komunikasi lisan. Standar Isyarat Bahasa Indonesia (SIBI) memberikan panduan dalam bentuk buku, namun pembelajaran dengan media buku kurang efektif karena tidak dapat memberikan contoh riil. Selain itu buku kurang praktis untuk dibawa. Bahasa isyarat dapat dikenali oleh komputer melalui pengenalan gambar (*image recognition*). Sebelum komputer mampu mengenali gambar yang diberikan, terdapat model pengenalan gambar yang harus dilatih dengan *machine learning*. Transformer adalah salah satu metode *machine learning* yang dapat digunakan dalam melatih model. Transformer merupakan implementasi mekanisme *self-attention* pada sebuah model, yang dapat memfokuskan bagian *input* yang diberikan. Mekanisme *attention* yang diberikan pada transformer memungkinkan model untuk mendapatkan konteks dari setiap bagian *input*.

**Kata kunci: Bahasa Isyarat, SIBI, *Image Recognition*, *Machine Learning*, Transformer, *Self-Attention***

**RECOGNITION TO VIDEO DEMONSTRATIONS USING   
TRANSFORMERS FOR SIGN LANGUAGE LEARNING**

**Student Name / NRP : Adam Hadi Prasetyo / 05111940000224**

**Department : Informatics Engineering FTEIC - ITS**

**Advisor : Dr. Eng. Nanik Suciati, S.Kom, M.Kom**

**Abstract**

In everyday life, communication is the most frequent human activity. Communication aims to convey intent in the form of spoken, written, pictures or gestures. Communication with gestures is known as sign language. In general, sign language is used due to physical limitations that do not allow the use of spoken communication. The Indonesian Language Sign Standards (SIBI) provide guidance in book form, but learning with book media is less effective because it cannot provide real examples. In addition, books are less practical to carry. Sign language can be recognized by a computer through image recognition. Before the computer is able to recognize a given image, there is an image recognition model that must be trained with machine learning. Transformer is one of the machine learning methods that can be used to train models. Transformer is the implementation of a self-attention mechanism in a model, which can focus on a given part of the input. The attention mechanism given to the transformer allows the model to get context from each part of the input.

**Keywords: Sign Language, SIBI, Image Recognition, Machine Learning, Transformer, Self-Attention**

# DAFTAR ISI

[HALAMAN JUDUL i](#_Toc121511240)

[LEMBAR PENGESAHAN ii](#_Toc121511241)

[ABSTRAK iv](#_Toc121511242)

[DAFTAR ISI vi](#_Toc121511243)

[DAFTAR GAMBAR/GRAFIK/DIAGRAM viii](#_Toc121511244)

[DAFTAR TABEL ix](#_Toc121511245)

[DAFTAR SIMBOL (jika ada) x](#_Toc121511246)

[DAFTAR SINGKATAN (jika ada) xi](#_Toc121511247)

[BAB I PENDAHULUAN 1](#_Toc121511248)

[1.1 Latar belakang 1](#_Toc121511249)

[1.2 Rumusan Masalah 2](#_Toc121511250)

[1.3 Batasan Masalah 2](#_Toc121511251)

[1.4 Tujuan 2](#_Toc121511252)

[1.5 Manfaat 2](#_Toc121511253)

[BAB II TINJAUAN PUSTAKA 3](#_Toc121511254)

[2.1 Penelitian Terkait 3](#_Toc121511255)

[2.2 Dasar Teori 3](#_Toc121511256)

[2.2.1 Deep Learning 4](#_Toc121511257)

[2.2.2 CNN 4](#_Toc121511258)

[2.2.3 RNN 5](#_Toc121511259)

[2.2.4 Transformer 6](#_Toc121511260)

[2.2.5 Tensorflow 7](#_Toc121511261)

[2.2.6 MediaPipe 7](#_Toc121511262)

[BAB III METODOLOGI 9](#_Toc121511263)

[3.1 Perancangan Aplikasi Pembelajaran 9](#_Toc121511264)

[3.1.1 Perancangan Model Pengenalan Bahasa Isyarat Menggunakan Transformer 10](#_Toc121511265)

[3.1.2 Perancangan Model Pembelajaran Bahasa Isyarat 10](#_Toc121511266)

[3.1.3 Dataset 10](#_Toc121511267)

[3.2 Peralatan pendukung 10](#_Toc121511268)

[3.3 Rencana Implementasi dan Uji Coba 10](#_Toc121511269)

[3.3.1 Antarmuka Aplikasi 10](#_Toc121511270)

[3.3.2 Pseudocode 11](#_Toc121511271)

[3.3.2.1 Cropping Video 11](#_Toc121511272)

[3.3.2.2 Ekstraksi Fitur 11](#_Toc121511273)

[3.3.2.3 Train dan Evaluasi Model 12](#_Toc121511274)

[3.3.2.4 Aplikasi Pembelajaran 12](#_Toc121511275)

[JADWAL KEGIATAN 13](#_Toc121511276)

[DAFTAR PUSTAKA 14](#_Toc121511277)

[LAMPIRAN-LAMPIRAN ATAU APPENDIKS (jika ada) 16](#_Toc121511278)

# DAFTAR GAMBAR/GRAFIK/DIAGRAM

[Gambar 2.1 Arsitektur Deep Learning dengan Multilayer Neural Network 4](#_Toc121515019)

[Gambar 2.2 Arsitektur CNN VGG-16 5](#_Toc121515020)

[Gambar 2.3 Arsitektur Umum RNN 6](#_Toc121515021)

[Gambar 2.4 Arsitektur Transformer 7](#_Toc121515022)

[Gambar 2.5 Google TensorFlow 7](#_Toc121515023)

[Gambar 2.6 MediaPipe Holistic Solution 8](#_Toc121515024)

[Gambar 3.1 Diagram Alir Pembuatan Model 9](#_Toc121515025)

[Gambar 3.2 Diagram Alir Aplikasi Pembelajaran 9](#_Toc121515026)

[Gambar 3.3 Antarmuka Modul Evaluasi Aplikasi Pembelajaran 11](#_Toc121515027)

# DAFTAR TABEL

[Tabel 2.1 Penelitian Terkait 3](#_Toc121341103)

[Tabel 3.1Spesifikasi Peralatan Pendukung 6](#_Toc121341104)

# DAFTAR SIMBOL (jika ada)

# DAFTAR SINGKATAN (jika ada)

# BAB I PENDAHULUAN

## Latar belakang

Manusia merupakan makhluk ciptaan tuhan yang berakal. Dalam kehidupan sehari-hari, manusia berkomunikasi satu sama lain untuk menyampaikan informasi. Komunikasi umumnya disampaikan melalui lisan, tulisan, serta visual. Akan tetapi, manusia tidak sempurna, beberapa memiliki kekurangan yang menghambat untuk berkomunikasi dengan sesama. Untuk mengatasi hambatan tersebut diciptakan metode lain untuk menggantikan cara komunikasi umum seperti dengan menggunakan bahasa isyarat. Masyarakat pada umumnya kesulitan dalam menggunakan bahasa isyarat karena tidak terbiasa. Untuk mengatasi permasalahan tersebut diperlukan adanya media pembelajaran bahasa isyarat.

Bahasa isyarat adalah bahasa yang menggunakan gestur anggota badan – pada umumnya tangan – untuk menyampaikan maksud. Bahasa isyarat sendiri memiliki standar yang berbeda, sesuai dengan bahasa daerah sekitarnya. Di Indonesia, standar baku bahasa isyarat tercantum dalam Standar Isyarat Bahasa Indonesia (SIBI). Pembakuan bahasa isyarat bertujuan untuk mempermudah penggunaan bahasa isyarat sehingga makna yang disampaikan tepat. SIBI memberikan panduan dan petunjuk sistematis dalam menggunakan bahasa isyarat yang baku dalam bentuk buku. Namun pembelajaran dengan media buku kurang efektif karena gambar pada buku tidak dapat memberikan contoh riil seperti praktek langsung. Selain itu buku kurang praktis untuk dibawa karena memiliki halaman tebal.

Untuk itu diperlukan sebuah sarana aplikasi yang dapat membantu dan mewadahi pembelajaran bahasa isyarat bagi masyarakat umum atau penyandang tunarungu. Diperlukan juga pembelajaran yang interaktif sehingga peserta belajar lebih mudah dalam memahami. Salah satu sarana yang memenuhi syarat tersebut adalah gim, yang menyediakan interaksi dengan peserta ajar dan memberikan umpan balik secara *real-time*.

Hingga saat ini sudah ada beberapa aplikasi yang dikembangkan untuk mewadahi pembelajaran bahasa isyarat. Salah satunya adalah Aplikasi Bahasa Isyarat untuk Tuna Rungu yang menggunakan platform android. Aplikasi ini menyediakan materi dan evaluasi pembelajaran bahasa isyarat dengan media ponsel. Adapun materi yang disajikan dalam bentuk gambar dan video. Untuk kategori materi bahasa isyarat yang tersedia antara lain abjad, angka, benda-benda umum, serta kelompok kata. Modul evaluasi pada aplikasi berupa tebak gambar dan tebak video, yang mana pengguna aplikasi akan memilih gambar atau video yang sesuai dengan soal yang diberikan, atau sebaliknya (Mandarani et al, 2020).

Aplikasi Pembelajaran Bahasa Isyarat (PemBais), adalah aplikasi pembelajaran bahasa isyarat serupa yang memberikan materi gambar dan video penggunaan bahasa isyarat juga telah dikembangkan. Materi pada PemBais memiliki beberapa kategori seperti kosa kata, angka, abjad, serta video gerakan bahasa isyarat. Aplikasi ini tidak menyajikan modul evaluasi bagi pengguna (Pradikja et al, 2019).

Dari contoh yang diberikan, aplikasi pembelajaran bahasa isyarat dapat menyampaikan pembelajaran melalui gambar dan video. Namun pembelajaran bahasa isyarat tersebut belum dapat memberikan evaluasi bagi pengguna tentang ketepatan penggunaan bahasa isyarat. Salah satu cara untuk mengevaluasi ketepatan tersebut adalah dengan memberikan aplikasi pembelajaran kemampuan untuk mengenali bahasa isyarat.

Dalam mengenali bahasa isyarat, solusi yang dapat digunakan adalah *deep learning*. *Convolutional Neural Network* (CNN) merupakan salah satu metode *deep learning* yang umum digunakan ketika *dataset* berupa gambar statis (Suharjito et al, 2018). Jika *dataset* berupa video atau urutan gambar yang saling terhubung, *Recurrent Neural Network* (RNN) dapat menghasilkan akurasi yang lebih baik (Lee et al, 2021).

Pada pengembangan aplikasi pembelajaran bahasa isyarat ini, metode *deep learning* yang akan digunakan untuk mengenali bahasa isyarat adalah arsitektur transformer. Arsitektur tersebut memiliki kelebihan dalam mengolah data sekuensial seperti RNN. Namun, transformer dapat mengolah semua *input* dalam satu waktu yang bersamaan.

## Rumusan Masalah

Berikut adalah beberapa hal yang menjadi rumusan masalah dalam tugas akhir ini:

1. Bagaimana rancangan model pengenalan bahasa isyarat dengan arsitektur transformer?
2. Bagaimana rancangan aplikasi pembelajaran bahasa isyarat yang memiliki modul latihan dan evaluasi?
3. Bagaimana rancangan modul latihan pada aplikasi pembelajaran bahasa isyarat dengan mempertimbangkan kesesuaian gestur tangan?
4. Bagaimana rancangan modul evaluasi?
5. Bagaimana cara mengevaluasi aplikasi pembelajaran bahasa isyarat?

.

## Batasan Masalah

Batas permasalahan yang dibahas dalam tugas akhir sebagai berikut:

1. Aplikasi yang akan dibuat dijalankan pada laptop atau komputer dengan sistem operasi Windows.

2. Target pengguna dari aplikasi pembelajaran ini adalah masyarakat umum dan penyandang tunarungu umur 10 hingga 35 tahun.

3. Game Engine yang digunakan untuk aplikasi pembelajaran yaitu Unity dengan bahasa pemrograman C#.

4. Modul evaluasi mandiri hanya menerima masukan isyarat per kata, bukan kalimat utuh.

## Tujuan

Tujuan dari pembuatan tugas akhir ini adalah mengembangkan model pengenalan bahasa isyarat untuk pembelajaran bahasa isyarat dengan menerapkan arsitektur transformer.

## Manfaat

Manfaat dari hasil pembuatan tugas akhir ini antara lain:

1. Memberikan sarana praktek penggunaan bahasa isyarat.

2. Memberikan pengalaman baru dalam belajar bahasa isyarat bagi pengguna.

3. Mengasah kemampuan pengguna dalam menggunakan dan mengenal bahasa isyarat.

# BAB II TINJAUAN PUSTAKA

## Penelitian Terkait

Dalam pengenalan bahasa isyarat atau *sign language recognition* (SLR) alat penangkap *input* yang biasa digunakan adalah kamera. Selain kamera, ada cara lain untuk mendapatkan *input* seperti dengan menggunakan *accelerometer*, sarung tangan yang dilengkapi sensor gerak, serta Microsoft Kinect. Dengan penggunaan kamera sederhana seperti kamera ponsel atau *webcam,* *input* yang dihasilkan berupa video yang memiliki *frame rate* 30 FPS. Sedangkan dengan penggunaan Kinect, dapat memberikan data *color, depth dan skeleton* (Suharjito et al, 2017).

Pada saat pemrosesan data, metode umum yang digunakan adalah *Hidden Markov Model* (HMM) yang digabungkan dengan algoritma lain seperti *Fuzzy K-Means* untuk melakukan klasifikasi bentuk tangan (Suharjito et al, 2017). Selain HMM, *deep learning* juga digunakan dalam pemrosesan data. *Convolutional Neural Network* merupakan salah satu jenis neural network yang umum digunakan pada data citra. CNN terdiri dari *neuron* yang memiliki bobot, bias, dan fungsi aktivasi. Secara umum, CNN terdiri dari: dua bagian, yaitu *feature extraction layer* dan *fully connected layer* (Daniels et al, 2021).

Dalam merancang tugas akhir, diperlukan adanya referensi untuk membantu pengerjaan. Tabel 2.1 berisikan penelitian yang terkait dengan tugas akhir. Adapun isi dari tabel seperti nama penelitian, dataset dan spesifikasinya, serta kelebihan dan kekurangan dari penelitian tersebut.

Tabel 2.1 Penelitian Terkait

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | **Nama** | **Dataset** | **Kelebihan** | **Kekurangan** |
| 1 | Sign Language Recognition Using Modified Convolutional Neural Network Model (Suharjito et al, 2018) | LSA64, 10 *vocabulary,*  10 pemeraga,  5 kali percobaan per *vocabulary* (500 video) | Metode pengolahan data dan pengujian dijelaskan dengan detail.  Hasil akurasi model yang tinggi dengan sedikit *class.* | Akurasi model menurun dengan *class* yang banyak. |
| 2 | American sign Language Recognition and Training Method with Recurrent  Neural Network  (Lee et al, 2021) | 26 data alfabet, 100 sampel per alfabet (2600 data) | Penggunaan RNN untuk 26 alfabet memiliki performa rata-rata yang lebih tinggi dibandingkan dengan metode lain dalam *cross validation* | Menggunakan *Leap Motion Controller* untuk mendeteksi gestur. Aplikasi yang dikembangkan menggunakan pendekatan *image processing*. |

## Dasar Teori

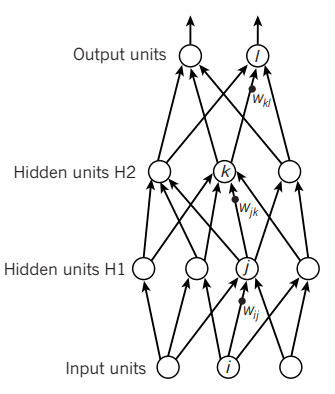
Dasar teori merupakan landasan utama dalam merancang tugas akhir. Dasar teori berisikan penjelasan lebih lanjut mengenai teori, alat, serta *library* yang akan digunakan. Pada sub bab 2.2 akan menjelaskan mengenai *deep learning*, *Convolutional Neural Network, Recurrent Neural Netowork*, Transformer, TensorFlow, serta MediaPipe.

## Deep Learning

*Deep learning* atau *deep structured learning* adalah bagian dari metode *machine learning* dengan basis *neural network* (jaringan saraf) buatan. *Deep learning* memungkinkan model yang terdiri dari beberapa lapisan pemrosesan untuk mempelajari representasi data dengan berbagai tingkat abstraksi. Metode ini menemukan struktur rumit dalam *data set* besar dengan menggunakan algoritma *backpropagation* untuk menunjukkan bagaimana mesin harus mengubah parameter yang digunakan untuk menghitung representasi di setiap lapisan dari representasi di lapisan sebelumnya (LeCun et al, 2015).

Metode deep learning adalah metode *representation-learning* dengan beberapa tingkat representasi, diperoleh dengan menyusun modul non-linier sederhana yang masing-masing mengubah representasi pada satu tingkat (dimulai dengan *input* mentah) menjadi representasi pada tingkat yang lebih tinggi, dengan sedikit tambahan abstrak. Dengan komposisi transformasi yang cukup, fungsi yang sangat kompleks bisa dipelajari. Untuk tugas klasifikasi, lapisan representasi yang lebih tinggi memperkuat aspek *input* yang penting untuk membedakan dan menekan variasi yang tidak relevan. Sebagai contoh, sebuah gambar muncul dalam bentuk *array* nilai *pixel*, dan fitur yang dipelajari di bagian pertama lapisan representasi biasanya mewakili ada atau tidak adanya tepi pada orientasi dan lokasi tertentu dalam gambar. Kedua lapisan biasanya mendeteksi motif dengan melihat pengaturan tertentu dari tepi, terlepas dari variasi kecil dalam posisi tepi. Ketiga lapisan dapat merakit motif menjadi kombinasi yang lebih besar yang sesuai ke bagian objek yang sudah dikenal, dan lapisan berikutnya akan mendeteksi objek sebagai kombinasi dari bagian-bagian ini (LeCun et al, 2015).

Arsitektur *deep learning* seperti *deep neural network* (DNN), *recurrent neural network* (RNN), *convolutional neural* network (CNN), dan *transformer* telah diterapkan di berbagai bidang seperti citra computer, pengenalan suara, serta *natural language processing* (NLP).

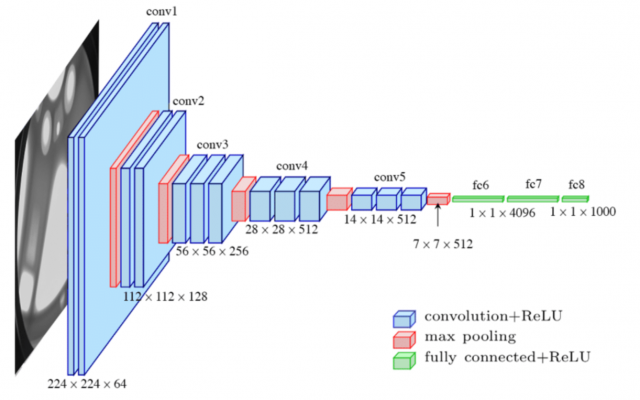


Gambar 2.1 Arsitektur Deep Learning dengan Multilayer Neural Network (LeCun et al, 2015)

## CNN

CNN (*Convolutional Neural Network* atau *ConvNet*) adalah salah satu arsitektur *deep learning*. CNN merupakan versi regular dari *multilayer perceptron* yang merupakan jaringan yang terhubung penuh. Pada *multilayer perceptron* setiap *neuron* dalam satu lapisan terhubung dengan semua *neuron* yang ada di lapisan berikutnya (Valueva et al, 2020).

Sebuah *convolutional neural network* terdiri atas *input layer, hidden layer,* dan *output layer*. Pada *feed-forward neural network*, setiap lapisan tengah disebut sebagai lapisan tersembunyi (*hidden layer*) karena *input* dan *output* ditutupi oleh fungsi aktivasi dan konvolusi akhir. Dalam *convolutional neural network,*  lapisan tersembunyi berisi lapisan yang melakukan konvolusi. Umumnya, lapisan ini mencakup lapisan yang melakukan *dot product* dari kernel konvolusi dengan *input matrix* lapisan. Saat kernel konvolusi berjalan di sepanjang *input matrix* pada lapisan, operasi konvolusi menghasilkan peta fitur, yang kemudian berkontribusi pada *input layer* berikutnya. Hal ini diikuti oleh lapisan lain seperti *pooling layer*, *fully connected layer*, dan *normalization layer.*



Gambar 2.2 Arsitektur CNN VGG-16 (Le, 2017)

## RNN

*Recurrent Neural Network* adalah bagian dari *deep learning* yang memiliki *self-connected hidden layer*. Salah satu kelebihan dari RNN adalah "memori" atau ingatan dari *input* di masa lalu tersimpan dalam *network internal state*. Dari "memori" tersebut RNN dapat memanggil atau mendapatkan konteks *input* masa lalu. Kelebihan lain yang dimiliki RNN, tingkat perubahan (*rate of change*) dari *internal state* dapat diatur melalui *recurrent weights*. Tingkat perubahan yang modular tersebut memberikan kekokohan terhadap distorsi lokal dari *input* (Graves et al, 2009).

RNN merupakan jaringan node mirip neuron yang diatur ke dalam "lapisan" yang berurutan. Setiap node dalam lapisan tertentu terhubung dengan koneksi terarah ke setiap node lain di lapisan berikutnya. Setiap node (neuron) memiliki aktivasi bernilai riil yang bervariasi waktu. Setiap koneksi (sinaps) memiliki bobot bernilai riil yang dapat dimodifikasi. Node adalah node input (menerima data dari luar jaringan), node output (menghasilkan hasil), atau node tersembunyi (yang mengubah data dalam perjalanan dari input ke output).

Graphical user interface, application

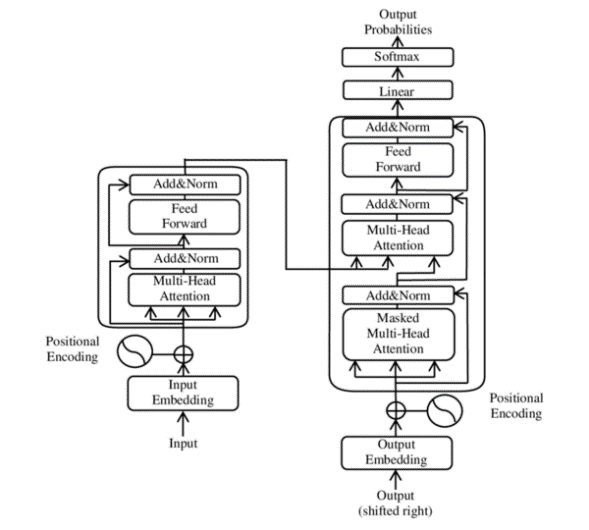
Description automatically generated

Gambar 2.3 Arsitektur Umum RNN Terkompresi (kiri) dan Tidak Terkompresi (kanan)

## Transformer

Transformer merupakan implementasi mekanisme *self-attention* pada sebuah model *deep learning*, yang dapat memfokuskan bagian *input* yang diberikan. Model transformer memiliki kesamaan dengan *Recurrent Neural Network* (RNN) yang didesain untuk memproses *input* data sekuensial. Perbedaan yang dimiliki oleh transformer dengan model RNN adalah model transformer dapat memproses semua *input* secara bersamaan. Mekanisme *attention* yang diberikan pada transformer memungkinkan model untuk mendapatkan konteks dari setiap bagian *input*. Sebagai contoh, ketika model diberikan *input* berupa sebuah kalimat bahasa alami, transformer akan memproses seluruh kalimat secara utuh (Vaswan, 2017).

Transformer terdiri dari dua komponen utama: satu set *encoder* yang terhubung dan satu set *decoder* yang terhubung. Fungsi setiap *encoder* adalah memproses vektor inputnya untuk menghasilkan *encoding*, yang berisi informasi tentang bagian-bagian input yang relevan satu sama lain. *Encoding* ini melewati set pengkodean yang dihasilkan ke *encoder* berikutnya sebagai *input*. Setiap *decoder* melakukan yang sebaliknya, mengambil semua pengkodean dan memprosesnya, menggunakan informasi kontekstual yang digabungkan untuk menghasilkan sekuens *output*. Untuk mencapai hal ini, setiap *encoder* dan *decoder* menggunakan mekanisme *self-attention*, yang untuk setiap input, menimbang relevansi setiap input dan menarik informasi yang sesuai saat menghasilkan output. Setiap *decoder* juga memiliki mekanisme *attention* tambahan yang menarik informasi dari *output decoder* sebelumnya, sebelum *decoder* mengambil informasi dari pengkodean. Baik *encoder* dan *decoder* memiliki *feed-forward neural network* akhir untuk pemrosesan *output* tambahan, dan juga berisi koneksi residual dan langkah normalisasi lapisan (Vaswan, 2017).



Gambar 2.4 Arsitektur Transformer

## TensorFlow

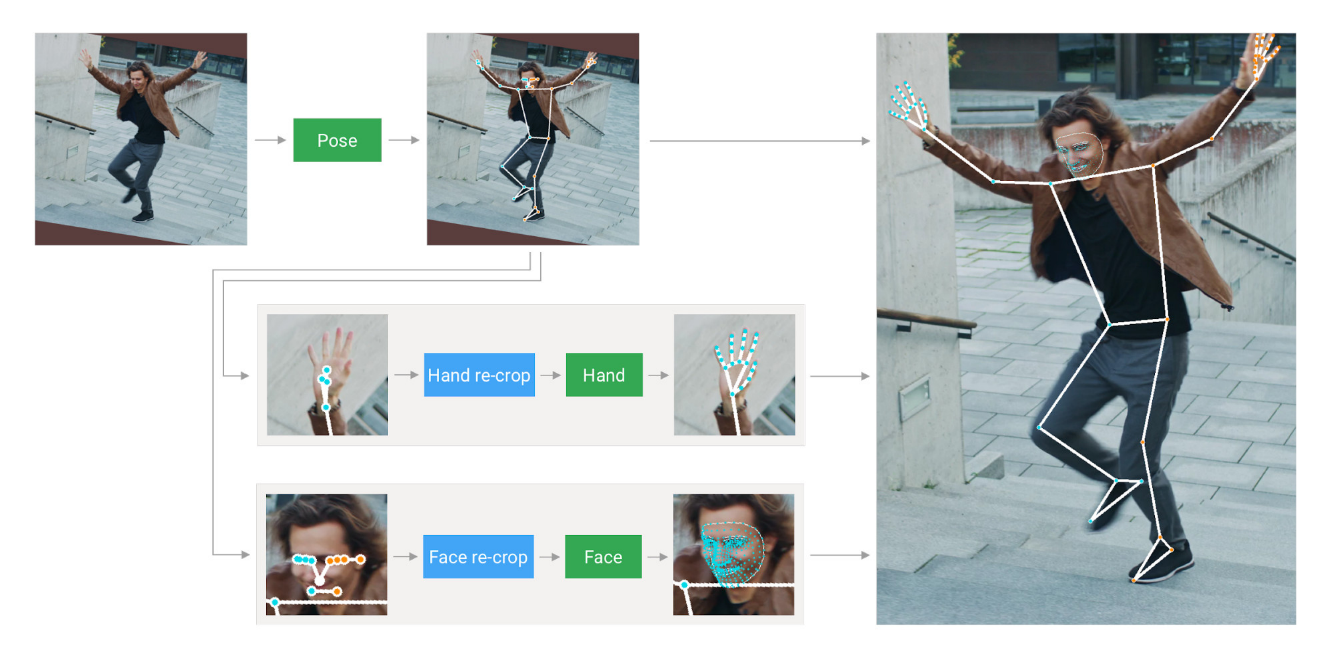


Gambar 2.5 Google TensorFlow

TensorFlow merupakan *library* yang difokuskan untuk mengerjakan kegiatan yang berhubungan dengan *machine learning* dan *artificial intelligence*. Pada awalnya, TensorFlow dikembangkan oleh tim Google Brain untuk keperluan riset dan produksi internal Google. Pada tahun 2017, versi 1.0.0 TensorFlow dirilis untuk umum. *Library* ini tersedia dalam beberapa bahasa pemrograman seperti Python, C++, JavaScript, dan Java. Dalam penggunaannya, TensorFlow menyediakan API yang menggunakan Keras. API tersebut memungkinkan pengguna untuk membuat model *machine learning* sesuai dengan kebutuhan. TensorFlow dapat memanfaatkan beberapa CPU dan GPU yang tersedia pada computer untuk menjalankan *model implementation*.

## MediaPipe

MediaPipe adalah *framework* *open-source* yang digunakan untuk mendeteksi objek secara *real-time*. MediaPipe menyediakan solusi untuk berbagai masalah deteksi objek seperti *tracking*, *segmentation*, serta *object detection*. Salah satu solusi MediaPipe yakni Holistic, memungkinkan pengguna untuk menggabungkan beberapa model *tracking* menjadi kesatuan yang utuh seperti kerangka. MediaPipe Holistic dapat digunakan untuk memprediksi pose, dengan batasan resolusi video *input* 256x256 pixel.



Gambar 2.6 MediaPipe Holistic Solution

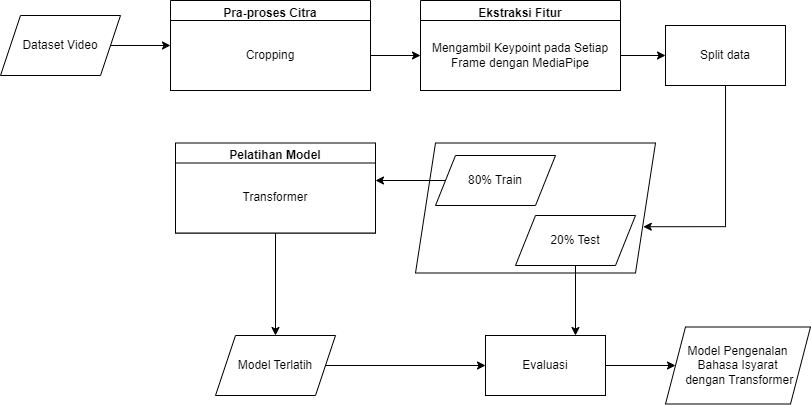
# BAB III METODOLOGI

## Perancangan Aplikasi Pembelajaran

Pada sub bab 3.1, akan membahas mengenai perancangan pembuatan model, pembuatan aplikasi, serta *dataset* yang akan digunakan. Adapun diagram alir (*flowchart*) sebagai visualisasi proses.

## Perancangan Model Pengenalan Bahasa Isyarat Menggunakan Transformer

Pada Gambar 3.1 menunjukkan rancangan pembuatan model pengenalan bahasa isyarat dengan transformer. Tahap pertama yang dilakukan adalah melakukan *preprocessing* pada *dataset* yang digunakan dengan menggunakan metode *cropping*. *Preprocess* tersebut dilakukan untuk mengurangi waktu *training* model dengan memfokuskan area pengenalan pada *region of interest*. Setelah *preprocess* akan dilakukan ekstraksi fitur dengan mengambil *keypoint* setiap *frame* menggunakan MediaPipe. Kemudian dilanjutkan *split* data dengan rasio *train : test* sebesar 80:20. Data akan dilatih menggunakan transformer, lalu dievaluasi. Setelah evaluasi, model akan digunakan pada aplikasi pembelajaran.



Gambar 3.1 Diagram Alir Pembuatan Model

## Perancangan Model Pembelajaran Bahasa Isyarat

Pada Gambar 3.2 menunjukkan rancangan aplikasi pembelajaran bahasa isyarat. Saat memulai aplikasi, akan terdapat dua pilihan modul, Belajar dan Evaluasi. Pada modul belajar terdapat video yang memperagakan penggunaan bahasa isyarat dengan Standar Isyarat Bahasa Indonesia (SIBI). Pada modul evaluasi, terdapat soal kosakata bahasa isyarat yang dapat diselesaikan. Untuk melakukan pengecekan ketepatan dalam penggunaan, aplikasi akan melakukan *tracking* gestur menggunakan kamera. Gestur akan dievaluasi dengan model yang telah dilatih menggunakan transformer. Jika hasil evaluasi adalah benar, maka soal kosakata akan dilanjutkan dengan soal berikutnya. Bila evaluasi adalah salah, soal tidak akan dilanjutkan hingga penggunaan bahasa isyarat dinilai benar oleh sistem.

Diagram

Description automatically generated

Gambar 3.2 Diagram Alir Aplikasi Pembelajaran

## Dataset

Untuk melatih model pengenalan bahasa isyarat menggunakan transformer, diperlukan adanya *dataset*. Pada tugas akhir ini, *dataset* bahasa isyarat dengan Standar Isyarat Bahasa Indonesia yang digunakan berisi 120 *class* video berdurasi 4 detik dengan 6 variasi dalam setiap *class*. Setiap *class* merupakan kata dasar serta imbuhan yang sering digunakan sehari-hari. *Dataset* diambil dalam posisi *landscape* dengan resolusi 1920x1080 *pixel* dan *frame-rate* 30fps.

## Peralatan pendukung

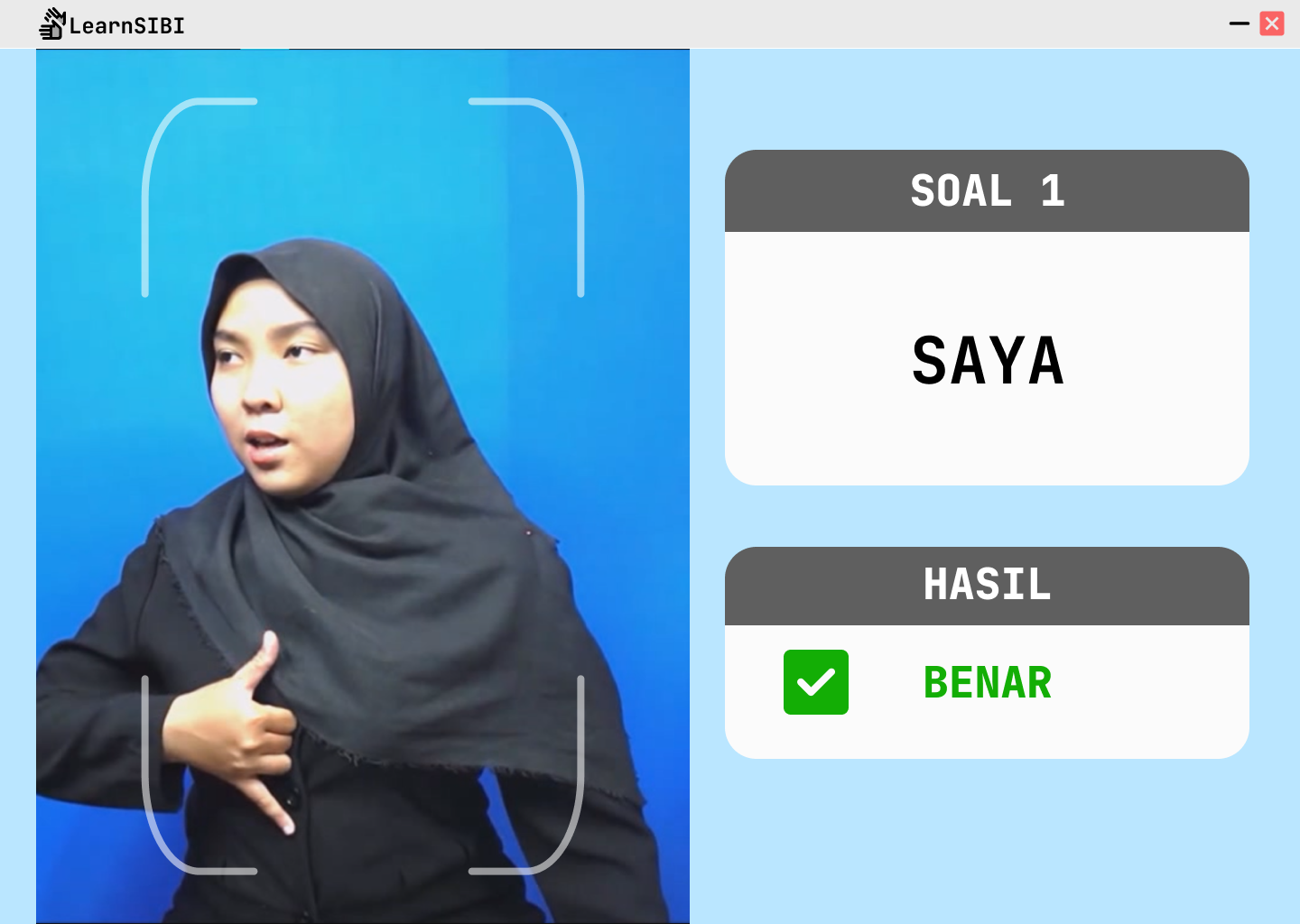
Dalam pengembangan aplikasi, diperlukan alat yang mendukung pengerjaannya. Adapun peralatan yang mendukung pengembangan aplikasi dengan rincian dan spesifikasi yang terdapat pada Tabel 3.1.

Tabel 3.1Spesifikasi Peralatan Pendukung

|  |  |  |
| --- | --- | --- |
| **No** | **Perangkat** | **Spesifikasi** |
| 1 | Sistem Operasi | Windows 10 Home 64-bit |
| 2 | RAM | 16GB |
| 3 | CPU | AMD Ryzen 5 4600H |
| 4 | GPU | NVIDA GeForce GTX 1650 Ti |
| 5 | Kamera | 20 MP, f/2.0, (wide), 1/3", 0.9µm 1080p@30fps |

## Rencana Implementasi dan Uji Coba

## Antarmuka Aplikasi



Gambar 3.3 Antarmuka Modul Evaluasi Aplikasi Pembelajaran

Gambar 3.3 merupakan antarmuka modul evaluasi pada aplikasi pembelajaran. *Tracking* gestur akan dilakukan melalui kamera, lalu penggunaan bahasa isyarat dievaluasi dengan model pengenalan bahasa isyarat. Jika penggunaan dinilai benar, maka akan menampilkan hasil “BENAR”.

## Pseudocode

*Pseudocode* adalah representasi cara kerja sistem dalam bentuk *syntax* yang disederhanakan. Berikut merupakan *pseudocode* pada sistem :

## Cropping Video

READ video

while video.isOpened :

    Crop video.frame

WRITE video\_cropped

Pada *Cropping Video*, akan dilakukan pembacaan *file* video. Kemudian, akan dilakukan *loop* sebanyak *frame* yang ada pada video. Setiap *frame* yang ada pada video akan di*-crop­* dengan resolusi yang telah ditentukan. Setelah iterasi *loop* selesai, hasil akan disimpan.

## Ekstraksi Fitur

READ video\_cropped

while video\_cropped.isOpened :

    Detect landmark

    Feature Extract landmark

    WRITE keypoints\_data

Split train\_data, test\_data

Saat proses *cropping* selesai, video yang telah di-*crop* akan diekstrak *keypoint*-nya. Hasil ekstraksi lalu disimpan. Setelah semua video telah diekstrak, akan dilakukan *split* *dataset*.

## Train dan Evaluasi Model

Define Transformer

Import train\_data

Train train\_data

Export model

Predict test\_data to model

PRINT evaluate\_result

Tuning model

Export evaluated\_model

Pada proses *train* dan evaluasi model, akan dilakukan *training* dengan transformer. Kemudian, model yang telah selesai dilatih akan dievaluasi hasilnya. Bila hasil evaluasi belum sesuai, akan dilakukan *tuning* model. Jika *tuning* sudah memberikan hasil yang sesuai, model akan di-*export* untuk digunakan pada aplikasi pembelajaran.

## Aplikasi Pembelajaran

Import evaluated\_model

while camera.READ(video) :

    Detect frame\_landmark

    Extract frame\_landmark

    Predict frame\_landmark to evaluated\_model

    if frame\_landmark == evaluated\_model

        Cek\_Ketepatan = True

        PRINT 'Benar !'

        BREAK

    else PRINT 'Salah !'

Pada aplikasi pembelajaran, akan dilakukan *tracking* gestur melalui kamera. Objek yang terekam dalam kamera akan dideteksi *landmark*-nya. Selanjutnya *landmark* akan diekstrak dan dibandingkan dengan model yang telah dibuat. Jika hasil pengecekkan maka akan menampilkan hasil “Benar !”. Jika salah, maka akan di-*track* hingga menghasilkan nilai “Benar !”.

# JADWAL KEGIATAN

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | Nama Kegiatan | Minggu ke- | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | Studi Pustaka |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Perancangan peralatan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Survey Lapangan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Eksperimen |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Analisa |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Pengolahan data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Pelaporan kemajuan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Pembuatan absrak seminar |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Mengikuti seminar |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 | Penyusunan laporan Tugas Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# DAFTAR PUSTAKA

Mandarani, Putri. Putra, Yogi. (2020). “Aplikasi Bahasa Isyarat untuk Tuna Rungu Menggunakan Platform Android”. Institut Teknologi Padang. <https://teknoif.itp.ac.id/index.php/teknoif/article/view/40/780>.

Pradikja, Maharoni Hendra. Tolle, Herman. Brata, Kumang Candra. (2019). “Pengembangan Aplikasi Pembelajaran Bahasa Isyarat Berbasis Android Tablet”. Universitas Brawijaya. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1705/645>

C.K.M.Lee, Kam K.H.Ng, Chun-Hsien Chen, H.C.W.Lau, S.Y.Chung, Tiffany Tsoi. (2021). “American Sign Language Recognition and Training Method with Recurrent Neural Network”. https://www.sciencedirect.com/science/article/pii/S0957417420310745

Suharjito. Anderson, Ricky. Wiryana, Fanny. Meita Chandra Ariesta, Gede Putra

Kusuma. (2017). Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output.

<https://www.sciencedirect.com/science/article/pii/S1877050917320720>

Suharjito. Gunawan, Herman Thiracitta, Narada. Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. IEEE. DOI: 10.1109/INAPR.2018.8627014

Daniels, Steve. Suciati, Nanik. Fathichah, Chastine. (2021). Indonesian Sign Language Recognition using YOLO Method. Institut Teknologi Sepuluh Nopember.

<https://www.researchgate.net/publication/350082064_Indonesian_Sign_Language_Recognition_using_YOLO_Method/fulltext/608b697e458515d315e6bd95/Indonesian-Sign-Language-Recognition-using-YOLO-Method.pdf>

LeCun. Yann, Bengio. Yoshua, Hinton. Geoffrey (2015). "Deep Learning". Nature. 521 (7553): 436–444.

Valueva, M.V.; Nagornov, N.N.; Lyakhov, P.A.; Valuev, G.V.; Chervyakov, N.I. (2020). "Application of The Residue Number System to Reduce Hardware Costs of The Convolutional Neural Network Implementation". Mathematics and Computers in Simulation. Elsevier BV. 177: 232–243. doi:10.1016/j.matcom.2020.04.031

Graves. Alex, Liwicki. Marcus, Fernandez. Santiago, Bertolami. Roman, Bunke, Horst; Schmidhuber, Jürgen (2009). "A Novel Connectionist System for Improved Unconstrained Handwriting Recognition".

Vaswan. Ashish, Shazeer. Noam, Parmar. Niki, Uszkoreit. Jakob, Jones. Llion, Gomez. Aidan N., Kaiser. Łukasz, Polosukhin. Illia. (2017). “Attention is All You Need”. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>

Le, Khuyen. (2017). *An overview of VGG16 and NiN models*. <https://medium.com/mlearning-ai/an-overview-of-vgg16-and-nin-models-96e4bf398484>. Diakses pada 6 Desember 2022.

# LAMPIRAN-LAMPIRAN ATAU APPENDIKS (jika ada)