



Segmentasi Objek Berbasis Gambar *Thermal* Menggunakan *Deep Learning* (*Pre-Trained Resnet 152*)

(*Object Segmentation Based On Thermal Images Using Deep Learning* (*Pre-Trained Resnet 152*))

Isra Fanliv Noviely¹, Arief Suryadi Satyawan² Heni Puspita³

¹ Teknik Elektro, Universitas Nurtanio

E-mail: israfanliv17@gmail.com

² Badan Riset dan Inovasi Nasional, Universitas Nurtanio

E-mail: arief.suryadi@akane.waseda.jp, henipuspita043@gmail.com

³AAU

Abstract— *The latest technological developments in the field of Artificial Intelligence have very rapid capabilities and are able to produce systems that facilitate human activities, especially in the field of transportation, especially driving cars or autonomous electric cars. Artificial Intelligence technology itself is able to support success for object detection by detecting objects using semantic segmentation. Neural Network and Image processing are methods used to detect objects semantically as input signal processing in the form of images, and the FLIR thermal camera is used as input from the vehicle. The deep learning method uses a Fully Convolutional Network (FCN) with a Residual Network (ResNet) architectural model as its feature extraction. ResNet is an architectural model from FCN that works from this architectural model not to decline even though the architecture is getting deeper, so it can help humans to drive more productively. The method used in this final project is automatic extraction using deep learning technology with Residual Neural Network 152 (ResNet) architecture. The performance of the semantic segmentation system was tested with 3040 image frames offline using 800 labeled data sets. This method has an extraction accuracy for autonomous vehicle function training reaching 96% with a resolution of 640x512 pixels. The performance of the segmentation system resulted in 18576 image frames in good category, 9333 image frames in sufficient category and 6121 image frames in poor category.*

Keywords— *Kendaraan Listrik Otonom, Deep Learning, Residual Network 152.*

Abstrak— Perkembangan teknologi dalam bidang *Artificial Intelligence* terkini memiliki kemampuan yang sangat pesat serta mampu menghasilkan sistem yang mempermudah kegiatan manusia, terutama dalam bidang transportasi khususnya mobil berkendara atau mobil listrik otonom. Teknologi *Artificial Intelligence* sendiri mampu mendukung kesuksesan untuk pendeteksi objek dengan cara mendeteksi objek menggunakan segmentasi semantik. *Neural Network* dan *Image processing* merupakan metode yang digunakan untuk mendeteksi objek secara semantik sebagai proses sinyal input yang berupa gambar, dan kamera *thermal FLIR* digunakan sebagai input dari kendaraan. Metode pembelajaran mendalam menggunakan *Fully Convolutional Network* (FCN) dengan model arsitektur *Residual Network* (ResNet) sebagai ekstraksi fiturnya. ResNet adalah model arsitektur dari FCN yang bekerja dari model arsitektur ini tidak menurun meskipun arsitekturnya semakin dalam, sehingga dapat membantu manusia dalam berkendara lebih produktif. Metode yang digunakan dalam tugas akhir adalah ekstraksi otomatis dengan menggunakan teknologi deep learning dengan arsitektur *Residual Neural Network 152* (ResNet). Kinerja sistem segmentasi semantik diuji dengan 3040 frame gambar secara *offline* menggunakan 800 data set yang sudah dilabeli. Metoda ini memiliki akurasi ekstraksi untuk pelatihan fungsi kendaraan otonom mencapai 96% dengan resolusi yang dipakai 640x512 pixel. Kinerja sistem

segmentasi menghasilkan 18576 *frame* gambar dengan kategori baik, 9333 *frame* gambar dengan kategori cukup dan 6121 *frame* gambar dengan kategori kurang baik.

Kata Kunci— *Kendaraan Listrik Otonom, Deep Learning, Residual Network 152.*

I. PENDAHULUAN

Teknologi informasi dan komunikasi saat ini yang berkembang cukup pesat khususnya di Indonesia memungkinkan penerapannya ke semua bidang kehidupan. Terutama pada pembangunan di masa industri 4.0. Teknologi pada saat ini diberbagai negara sudah mulai banyak dikembangkan. Kendaraan listrik otonom (KLO) merupakan salah satu kendaraan listrik tanpa campur tangan manusia dan kendaraan ini merupakan hasil perkembangan teknologi tersebut. Untuk mempermudah manusia dalam bertransportasi kendaraan listrik otonom mampu mengurangi kelalaian manusia dan memberi kemudahan khususnya bagi yang berusia lanjut dan anak-anak dalam perjalanan tanpa perlu mengemudi. Maka dari itu kendaraan listrik otonom membutuhkan sistem yang dapat mendeteksi dan membedakan objek menggunakan sensor kamera termal dengan metode segmentasi berbasis *deep learning* yang dilatih dengan dataset yang dibutuhkan.

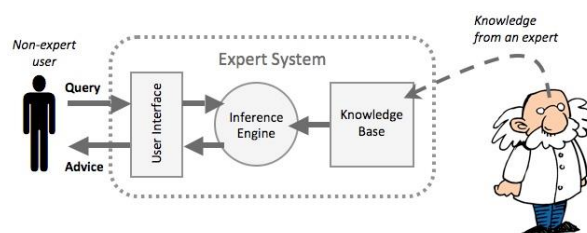
Pada tugas akhir ini, penulis akan melakukan penelitian yang mendukung system pendeteksian objek dengan melakukan segmentasi objek berbasis sensor kamera termal dengan jenis FLIR ADK. Sistem yang dikembangkan menggunakan komputer untuk memproses gambar yang diperoleh kamera termal dan konsep *deep learning* yang dikembangkan. Dan untuk menjalankan model aplikasinya digunakanlah Jetson AGX Xavier.

II. LANDASAN TEORI

A. Artificial Intelligence.

Menurut John McCarthy (1956) yang dikutip dari jurnal penelitian (Pannu, 2015) AI merupakan aplikasi yang mampu mengetahui dan memodelkan proses berpikir manusia dan mendesain mesin serta mampu menirukan aktifitas manusia. Terdapat dua hal bagian utama yang dibutuhkan dalam pembuatan aplikasi kecerdasan buatan yakni :

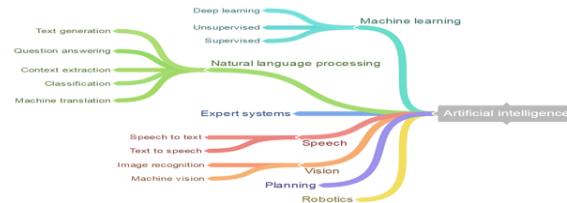
1. Basic Knowledge (Dasar Pengetahuan). Bagian ini terdiri dari fakta teori, pemikiran dan hubungan – hubungan lainnya
2. Inferensi Engine (Motor Inferensi) merupakan kemampuan menarik hasil berdasarkan pengalaman.



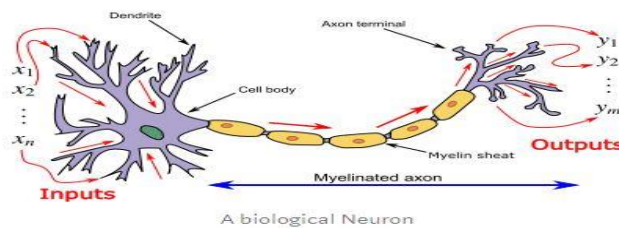
Gambar 2.1 Penerapan Konsep Kecerdasan Buatan

B. Artificial Neural Network (ANN).

Artificial Neural Network (ANN) merupakan suatu model komputasi paralel yang meniru fungsi dari sistem jaringan syaraf biologi otak manusia. Dalam otak manusia terdiri dari milyaran neuron yang saling berhubungan.

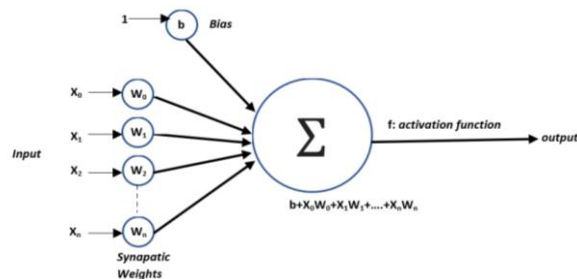
Gambar 2.2 Himpunan Bagian *Artificial Intelligence*

Synapses. Merupakan kumpulan neuron terdiri dari satu inti sel yang akan melakukan pemrosesan informasi, satu akson (*axon*) dan minimal satu dendrit.



Gambar 2.3. Jaringan Saraf Manusia

Fungsi sistem saraf yang ditunjukkan pada Gambar 2.3. diupayakan oleh sinyal yang memasuki perikaryon/badan sel (*cell body*) melalui dendrit. Sinyal tersebut kemudian diproses di dalam badan sel berdasarkan fungsi tertentu (*proses total*). Ketika sinyal yang diproses melebihi ambang batas tertentu, sinyal tersebut menciptakan neuron dan melanjutkan sinyal tersebut. Di sisi lain, jika sinyal jatuh di bawah ambang batas, itu akan diblokir. Sinyal yang ditransmisikan kemudian melewati sinapsis untuk mencapai akson dan akhirnya ke neuron lain.

Gambar 2.4 Model ANN (*Perceptron*) yang Diinspirasi Jaringan Saraf Tiruan

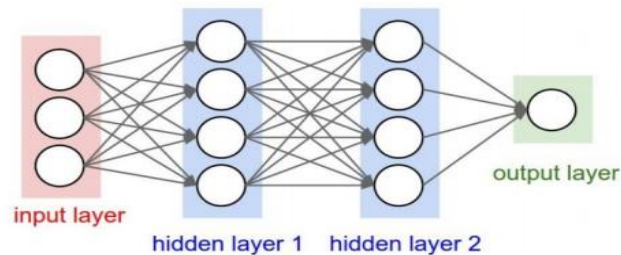
Model JST yang ditunjukkan pada Gambar 2.4. adalah kombinasi linier dari bentuk berikut:

1. $\{x_0, x_1, x_2, x_3, \dots, x_n\}$ mewakili jumlah *input* (*independent variables*) menuju *node*.
2. $\{w_0, w_1, w_2, w_3, \dots, w_n\}$ adalah bobot setiap *input*, dan menunjukkan kekuatan setiap *node*.
3. b digunakan sebagai *bias* yang dapat menambah atau mengurangi nilai keluaran dari *activation function* (fungsi aktivasi).
4. $x_1 \cdot w_1 + x_2 \cdot w_2 + x_3 \cdot w_3 \dots x_n \cdot w_n = \sum x_i \cdot w_i$ adalah fungsi dengan fungsi aktivasi yang menggambarkan kombinasi linier.
5. Jika jenis fungsi aktivasi yang berbeda (seperti *sigmode* atau ReLU) digunakan, model matematika akan disesuaikan.

C. *Convolutional Neural Network (CNN)*.

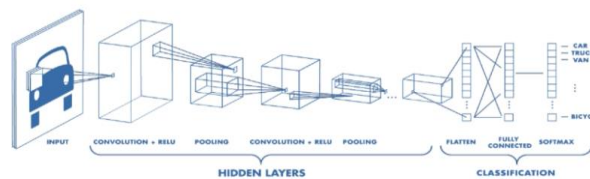
Convolutional Neural Network adalah salah satu metode *Deep Learning* (DL) yang dapat digunakan untuk mendeteksi dan mengenali sebuah objek pada sebuah citra digital (gambar). *Stride* dan *padding* terdapat di dalam *convolutional neural network*. *Stride* adalah jumlah piksel yang digunakan untuk menggeser filter secara horizontal atau vertikal. Dimana, dalam hal ini *stride* memindahkan filter satu piksel pada setiap langkah untuk menghitung *output* konvolusi berikutnya. Sedangkan *padding* adalah modifikasi yang umum digunakan yang memungkinkan ukuran *input* disesuaikan dengan kebutuhan.

Secara keseluruhan, *Convolutional Neural Network* (CNN) tidak jauh berbeda dari neural network biasa. CNN terdiri dari neuron dengan memiliki *weight*, bias dan *activation function*. *Convolutional layer* juga terdiri dari neuron yang diatur untuk membentuk sebuah filter panjang dan tinggi (pixels).



Gambar. 2.5. Arsitektur Multi Layer Perceptron (MLP)

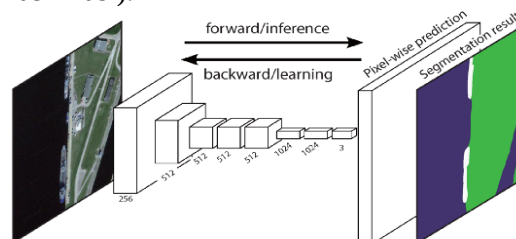
Pada dasarnya CNN merupakan arsitektur ANN yang memiliki sifat khusus, dapat dilatih, dan memiliki beberapa tahapan yang terdiri dari dari masukan (*input*) dan keluaran (*output*). *Input* dan *output* pada setiap tahap terdiri dari beberapa *array* yang digunakan untuk *feature map*.



Gambar 2.6 Struktur Dasar *Convolutional Neural Network*

D. Fully Convolutional Network (FCN).

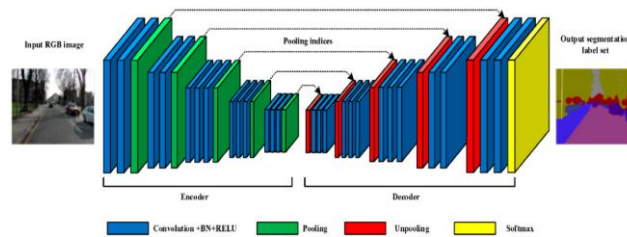
Fully Convolutional Network (FCN) adalah arsitektur yang digunakan terutama untuk segmentasi semantic. Proses pendekatan FCN yakni dengan mengadaptasi dan modifikasi model rekognisi CNN hingga dapat digunakan untuk proses segmentasi. Untuk melakukan modifikasi rekognisi model yakni dengan menghasilkan output berbentuk 1D (berupa probabilitas untuk setiap klasifikasi) untuk model segmentasi menghasilkan putput 3D (vector pprobabilitas untuk setiap klasifikasi).



Gambar 2.7 Struktur Dasar *Fully Convolutional Network*

Ada dua jenis arsitektur yang digunakan untuk segmentasi semantic, yaitu arsitektur *encoder-decoder*. *Encoder* biasanya adalah jaringan klasifikasi yang telah dilatih sebelumnya seperti VGG / ResNet, diikuti oleh jaringan *decoder*. Peran *decoder* adalah memproyeksikan

fitur gambar yang diperiksa oleh *enkoder* secara semantik ke dalam ruang piksel (resolusi lebih tinggi) untuk mendapatkan klasifikasi yang lebih padat.

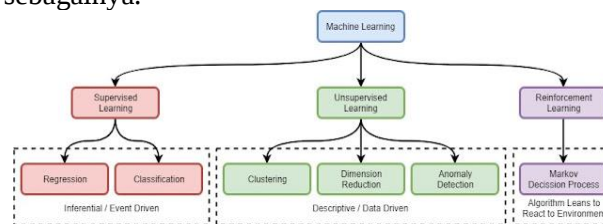


Gambar 2.8 Arsitektur *Fully Convolutional Network*

E. Machine Learning.

Menurut Arthur Samuel 1959 *machine learning* adalah suatu ilmu komputer yang mampu memberikan pembelajaran kepada komputer untuk mengetahui sesuatu tanpa pemrograman yang jelas. Ada dua aplikasi utama pada machine learning, yakni klasifikasi dan prediksi. Beberapa permasalahan yang mampu diselesaikan menggunakan machine learning adalah :

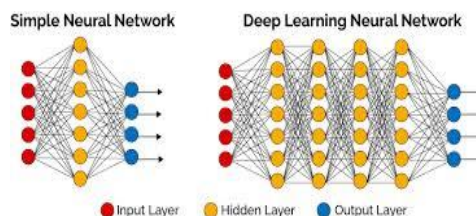
1. Classification
2. Deteksi Anomali
3. Estimasi Fungsi Massa Probabilitas; dan lain sebagainya.
4. Mesin Tranlasi
5. Regression
6. Transkripsi



Gambar 2.9 Beberapa Jenis Algoritma Pembelajaran Pada *Machine Learning*

F. Deep Learning.

Deep Learning adalah salah satu cabang dari ilmu pembelajaran mesin (*Machine Learning*) yang terdiri algoritma pemodelan abstraksi tingkat tinggi pada data menggunakan sekumpulan fungsi transformasi *non-linear* yang ditata berlapis-lapis dan mendalam. Teknik dan algoritma dalam *deep learning* dapat digunakan baik untuk kebutuhan pembelajaran terarah (*supervised learning*), pembelajaran tak terarah (*unsupervised learning*) dan semi-terarah (*semi-supervised learning*) dalam berbagai aplikasi seperti pengenalan citra, pengenalan suara, klasifikasi teks, dan sebagainya.



Gambar 2.10 Perbandingan Simple dan *Deep Learning* Neural Network

G. Image Processing.

Image processing adalah memanipulasi gambar untuk menyempurnakan atau mengekstrak informasi darinya, *Image processing* terbagi menjadi 2 metode yakni *Analog Image processing* dan *Digital Image Processing*.

1. Convolution (konvolusi).

Convolution didefinisikan sebagai operasi dua fungsi. Dalam menganalisis citra, fungsi pertama merupakan nilai masukan atau nilai piksel di suatu posisi dalam citra.

$$\text{Feature map} = [W - F] + 1 \dots \dots \dots (1)$$

Keterangan:

W = Ukuran volume gambar

F = Ukuran *Filter*

2. *Stride*.

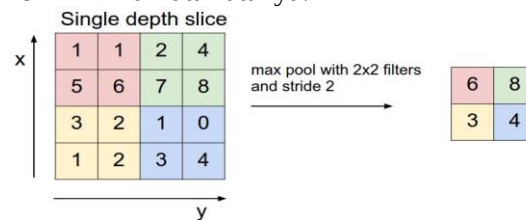
Stride adalah parameter yang menentukan berapa jumlah pergeseran *filter*. Jika nilai *stride* adalah 1, maka *filter* akan bergeser sebanyak 1 *pixels* secara *horizontal* lalu *vertical*.

3. *Padding*.

Padding atau *Zero Padding* adalah parameter yang menentukan jumlah *pixels* (berisi nilai 0) yang akan ditambahkan disetiap sisi dari *input*. Hal ini digunakan dengan tujuan untuk memanipulasi dimensi *output* dari *convolution layer* (*Feature map*).

4. *Pooling*.

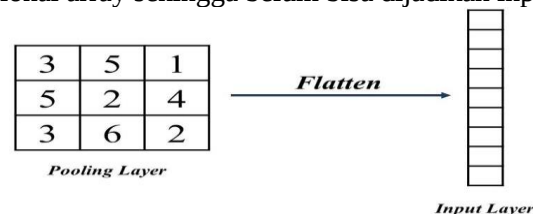
Pooling layer biasanya berada setelah *convolution layer*. Pada prinsipnya *pooling layer* terdiri dari sebuah *filter* dengan ukuran dan *stride* tertentu yang akan bergeser pada seluruh area *feature map*. *Pooling* yang biasa digunakan adalah Max *Pooling* dan Average *Pooling*. Sebagai contoh jika kita menggunakan Max *Pooling* 2x2 dengan *stride* 2, maka pada setiap pergeseran filter nilai maximum pada area 2x2 pixel tersebut yang akan dipilih, sedangkan Average *Pooling* akan memilih nilai rata-ratanya.



Gambar 2.11 *Pooling*

5. *Flatten*.

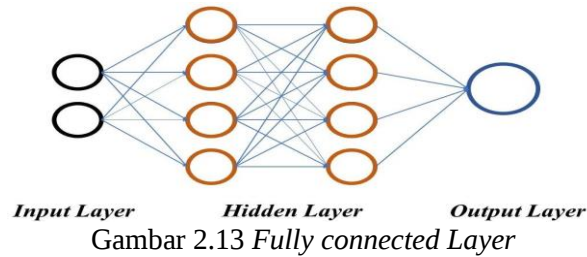
Output dari proses konvolusi yaitu setelah dilakukan operasi average pooling masih berbentuk multidimensional array sehingga belum bisa dijadikan input pada neural network.



Gambar 2.12 *Flatten*

6. *Fully connected Layer*.

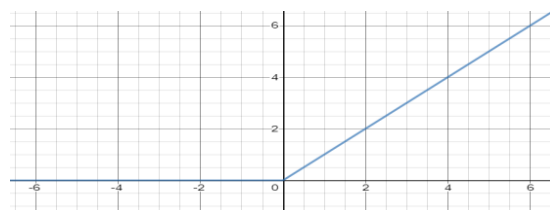
Fully connected layer berarti bahwa setiap neuron pada lapisan sebelumnya dihubungkan dengan setiap *neuron* di *fully connected layer*, seperti konvolusi, ReLU, dan pooling layer, memungkinkan terdapat satu atau lebih *fully connected layer* pada suatu arsitektur CNN.



7. Activation Layer.

Activation layer digunakan untuk menentukan apakah *neuron* dari jaringan tersebut akan diaktifkan atau dinonaktifkan. Secara umum terdapat 2 jenis fungsi aktivasi, *Linear* dan *Non- Linear Activation*. Nilai keluarannya berada dalam kisaran 0 hingga tak terbatas. persamaan 2 merupakan fungsi ReLU, dan fungsi input terhadap output dapat dilihat sebagai berikut

$$(z) = \max (0, z) \dots\dots\dots (2)$$



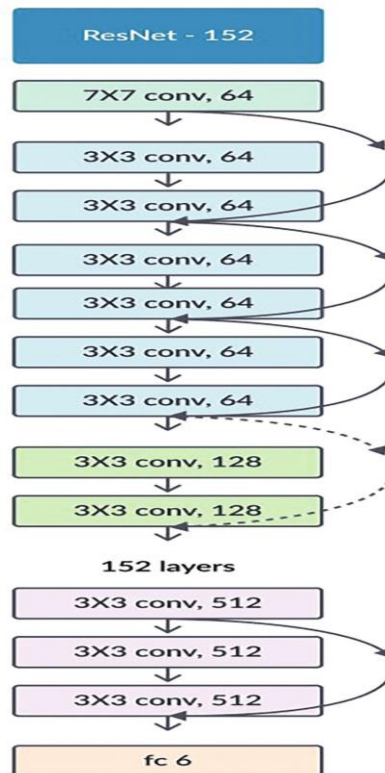
Gambar 2.14 Rectified linear unit (ReLU)

8. Softmax.

Fungsi softmax digunakan untuk metode klasifikasi dengan jumlah kelas yang banyak, seperti regresi logistik multinomial, ANN, dan lain sebagainya. Softmax merupakan sebuah fungsi yang mengubah nilai bobot keluaran dari semua proses di fully connected layer menjadi rentang nilai 0 dan 1.

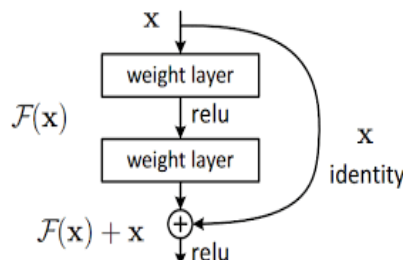
H. Residual Network.

FCN pada awalnya ditujukan dengan adaptasi dari model rekognisi VGG, tetapi FCN dapat digunakan juga untuk mengadaptasi model rekognisi lain seperti *Residual Network* (ResNet). Jaringan ini dikembangkan oleh peneliti dari *Microsoft Research*. ResNet memiliki beberapa macam jenis arsitektur yang dibedakan berdasarkan jumlah layer yang digunakan, mulai dari 18 layer, 34 layer, 50 layer, 101 layer, sampai 152 layer.



Gambar 2.15 Arsitektur ResNet 152

Model ResNet memperkenalkan blok baru yang dikenal dengan blok residual.

Gambar 2.16 Blok *Residual Network*

Blok *Residual Network* pada ResNet dapat dilakukan apabila dimensi data masukan sama dengan dimensi data keluaran. Selain itu, setiap blok ResNet terdiri dari 2 layer (untuk jaringan ResNet 18 dan ResNet 34) atau 3 layer (untuk jaringan ResNet 50, 101, 152). Dua lapisan awal dari arsitektur ResNet menyerupai GoogleNet dengan melakukan *convolution* 7×7 dan *max pooling* berukuran 3×3 dengan jumlah *stride* 2 yang dapat dilihat sebagai berikut.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix}$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix}$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix}$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix}$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Gambar 2.17 Arsitektur ResNet 152

I. Confusion Matrix.

Confusion matrix adalah salah satu metode evaluasi pada segmentasi semantik untuk mengetahui seberapa baik sistem yang dibuat dalam pengklasifikasian data. Pada *confusion matrix*, dikenal empat istilah berupa *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) untuk mempresentasikan hasil perbandingan klasifikasi sistem dengan klasifikasi sesungguhnya.

Tabel 2.1 Confusion Matrix

		Aktual	
		True	False
Prediksi	True	TP	FN
	False	FP	TN

Tabel 2.2 Confusion Matrix Multi-Class

		Aktual								
Classes		Building	Tree	Motorcycle	Car	Pedestrian	Road	Pole	Pavement	Total
Prediksi	Building	12	0	0	0	0	3	0	0	15
	Tree	0	10	2	0	0	0	4	0	16
	Motorcycle	0	0	9	2	0	0	0	0	11
	Car	0	0	2	7	0	0	0	0	9
	Pedestrian	0	3	1	0	15	0	0	0	19
	Road	2	0	0	0	0	17	0	4	24
	Pole	0	2	0	0	0	0	10	0	12
	Pavement	0	0	0	0	0	4	0	16	20
Total		14	15	14	9	15	24	14	20	96

Tabel 2.3 Akurasi *Coffusion Matrix Multi-Class*

Kelas	TP	FP	FN	TN	Akurasi
<i>Building</i>	12	3	2	83	92%
<i>Tree</i>	10	6	3	81	91%
<i>Motorcycle</i>	9	2	5	84	93%
<i>Car</i>	7	2	2	89	96%
<i>Pedestrian</i>	15	4	0	81	96%
<i>Road</i>	17	2	3	78	95%
<i>Pole</i>	10	2	4	84	94%
<i>Pavement</i>	16	4	4	76	92%

Rumus mendapatkan akurasi adalah seperti persamaan 3.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\% \dots \dots \dots (3)$$

Keterangan :

TP : Prediksi kelas label dan benar

TN : Prediksi bukan kelas label dan benar

FP : Prediksi kelas label dan aktual bukan kelas label

FN : Prediksi bukan kelas label dan aktual kelas label

J. Segmentasi Semantik

Segmentasi semantik merupakan teknik yang digunakan untuk mempartisi objek yang bermakna semantik dan diklasifikasikan kedalam kelas yang telah ditentukan (Hong, et al., 2015). Segmentasi memiliki 2 metode yakni untuk memisahkan gambar yang diinginkan (foreground) dari gambar yang tidak diinginkan (background). Metode segmentasi paling sederhana disebut *thresholding*, yang bergantung pada perubahan intensitas warna antara piksel objek dan piksel latar belakang.

K. NVIDIA Jetson AGX Xavier.

NVIDIA AGX Jetson Xavier Developer Kit adalah kit pengembangan AI yang dapat digunakan untuk menjalankan berbagai kebutuhan AI modern. Dengan *NVIDIA Jetson AGX Xavier*, pengembang dapat menjalankan *framework* dan model AI untuk aplikasi pengenalan gambar, pengenalan objek, segmentasi, pengenalan suara, dan banyak lagi.

L. Python.

Python adalah salah satu bahasa pemrograman yang dapat melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode orientasi objek (Object Oriented Programming) serta menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax.

M. PyTorch.

PyTorch merupakan pengembangan dari Torch Framework yang awalnya berbahasa pemrograman Lua. PyTorch, yang dikembangkan oleh Facebook, difokuskan sebagai framework proses komputasi Machine Learning, sehingga bisa disandingkan dengan framework sejenis seperti Tensorflow (dikembangkan oleh Google), Keras, Theano (sudah tidak dikembangkan), Caffe2, dan lainnya.

N. LabelMe.

LabelMe adalah sebuah alat yang dikembangkan oleh Massachusetts Institute Of Technology (MIT) *Computer Science and Artificial Intelligence Laboratory* (CSAIL) yang mendukung anotasi gambar yang memberikan informasi tentang ID, bentuk dan lokasi .

O. Pencitraan Termal.

Kemampuan melihat melalui intensitas yang dimiliki sebuah objek merupakan penjelasan tentang pencitraan termal. Metode yang mampu memanfaatkan suhu atau panas yang dihasilkan dari sebuah objek dengan menggunakan sensor panas pada kamera termal merupakan pengolahan citra termal.

P. FLIR (Vision Automotive Development Kit) ADK Dokumen.

FLIR ADK adalah Kamera inframerah termal adalah teknologi sensor terbaik untuk deteksi pejalan kaki dan menyediakan data penting yang diperlukan untuk meningkatkan pengambilan keputusan otomatis dalam fungsi keselamatan kendaraan termasuk pengereman darurat otomatis (AEB). Selanjutnya akan dibahas mengenai prosedur penggunaan umum dari kamera FLIR.

1. USB – Getting Started

Penggunaan aplikasi Boson dengan USB pada kamera FLIR ADK terhubung dengan port USB untuk menyalurkan daya pada kamera FLIR ADK untuk menampilkan video dan control pada kamera FLIR ADK.

a. Connect and power the USB ADK.

- 1) Hubungkan kabel USB ke komputer.
- 2) Kabel Opsional: Untuk menyalakan *window heater*, memerlukan *power suplay* sebesar 12 VDC dengan arus 0,5A dengan konektor MIZU P-25.

b. FLIR Boson GUI (Graphical User Interface).

digunakan untuk streaming gambar serta mengontrol kamera yang ada pada halaman web tersebut <https://www.flir.com/support-center/oem/camera-controller-gui-for-boson/>. temukan kamera dengan membuka Device Manager dan daftar Port yang sudah tersedia.

c. Sinkronisasi Eksternal.

pada kabel USB ADK terdapat Connector coaxial BNC yang mampu mensinkronisasikan jika diperlukan. ADK memicu ditepi jatuh dari sinyal 3-12 volt dengan lebar pulsa minimum 90nS. Hal tersebut untuk dsinkronisasikan pulsa pada 55-60Hz, bahkan jika kamera dalam output 30 Hz.

III. METODE/MODEL YANG DIUSULKAN

A. Arsitektur Model Secara Umum.

Tabel 3.1 Tahapan Penelitian

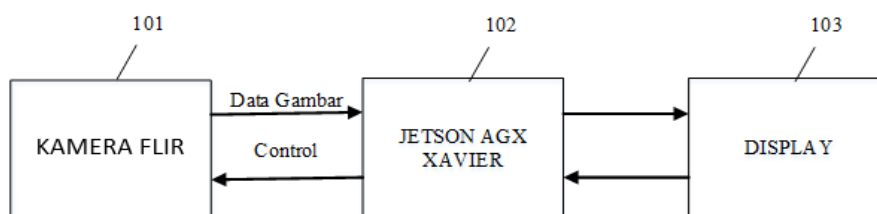
No	Tahapan	Meode	Hasil
1.	Studi literatur.	1. Survei internet. 2. Studi buku.	Pemahaman: 1. Mengenai teknologi <i>Artificial Intelligence</i> , <i>Machine Learning</i> , <i>Deep Learning</i> , <i>image processing</i> , <i>Convolutional Neural Network</i> (CNN), <i>Fully Convolutional Network</i> (FCN), <i>Residual Network</i> (ResNet), <i>Coffusion Matrix</i> , dan pemrosesannya 2. <i>Hardware</i> dan <i>Software</i> yang digunakan 3. Pemrograman dengan bahasa <i>Python</i> pada PC dan NVIDIA Jetson
2.	Desain perangkat lunak <i>system</i> pendeteksi menggunakan sensor kamera <i>thermal</i>	a. Menggunakan <i>software</i> pemrograman <i>Python</i> Anaconda b. Penggunaan model <i>software</i> ResNet 152 dan Memberi	a. Desain dan modifikasi <i>software</i> aplikasi pendeteksian objek berbasis <i>Deep Learning</i> pre-trained ResNet 152. b. Dataset pendeteksi terdiri dari 10 klasifikasi yaitu. a. Background b. Sky c. <i>Building</i> d. Pole

No	Tahapan	Meode	Hasil
		label pada gambar menggunakan Labelme	e. Road f. Pavement g. Tree h. Car i. Pedestrian j. Motorcycle
3.	Realisasi perangkat lunak pada Jetson AGX xavier	Menggunakan <i>software</i> pemrograman <i>Python</i> .	<i>software</i> aplikasi pendeteksian objek menggunakan sensor kamera <i>thermal</i> berbasis DL.
4.	Pengukuran kinerja <i>software</i> teknologi Deteksi objek	Pengukuran skala laboratorium dan lapangan.	Kinerja <i>software</i> aplikasi pendeteksian objek menggunakan sensor kamera <i>thermal</i> berbasis DL pada Jetson AGX xavier.
5	Melakukan Analisis	Analisis kuantitatif dan analisis kualitatif	Data dan kesimpulan penelitian
6	Pembuatan laporan tugas akhir	Pengetikan tugas akhir	Laporan tugas akhir

IV. HASIL/IMPLEMENTASI MODEL DAN PEMBAHASAN

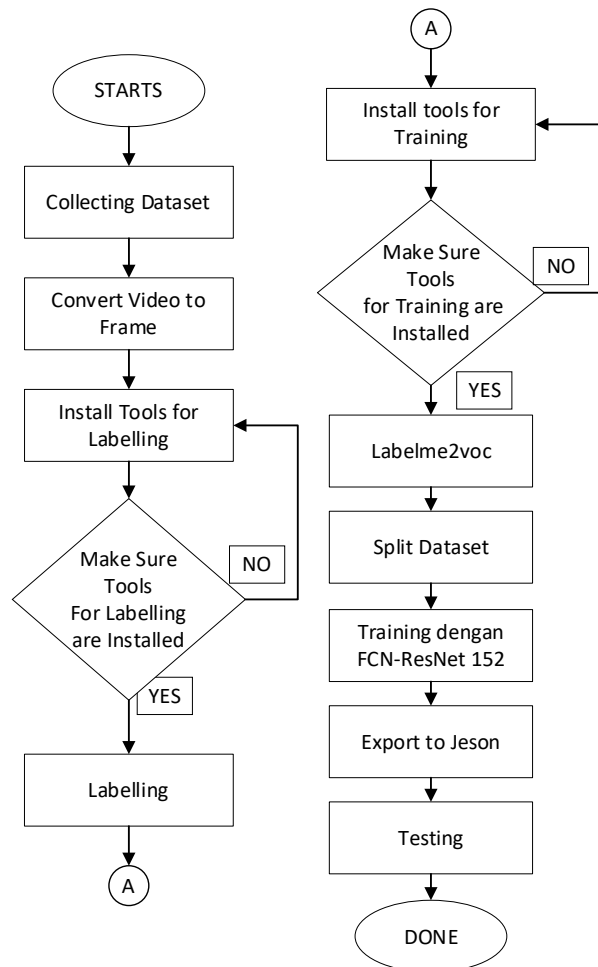
A. Desain.

Konstruksi Hardware dan Software merupakan komponen penting yang digunakan untuk membantu dalam desain system pendeteksian objek di lingkungan terbatas berbasis Arsitektur, CNN Pre-trained Resnet 152 menggunakan gambar kamera thermal FLIR. Konstruksi Hardware terdiri dari kamera thermal FLIR, Jetson AGX Xavier, dan modul display. Kamera Thermal Konstruksi hardware digunakan untuk membangun system pendeteksian objek berdasarkan klasifikasinya (ditunjukkan pada nomor 101-103). Kamera thermal FLIR digunakan untuk mengambil sumber gambar dalam bentuk objek berdasarkan klasifikasinya di area terbatas (Fakultas Teknik Universitas Nurtanio Bandung). Jetson AGX Xavier digunakan untuk proses penangkapan gambar thermal FLIR, pelatihan software jaringan Deep Learning Arsitektur CNN Pre-Trained ResNet152 dan pengujian software jaringan Deep Learning Arsitektur CNN Pre-Trained ResNet 152 terlatih baik secara real-time maupun offline. Bagian diplay berfungsi untuk menampilkan gambar thermal FLIR hasil deteksi objek berdasarkan klasifikasinya.



Gambar 4.1 Konstruksi *Hardware* Pendeteksian Objek berbasis Arsitektur CNN Pre-trained Resnet-152

B. Langkah Alur Sistem Pendeteksian Objek Pada Gambar Thermal



Gambar 4.3 Flowchart Sistem Segmentasi Objek

C. Pengujian Pertama.

Pengujian pertama pada tugas akhir ini adalah pengujian sistem pendeteksian objek secara segmentasi semantik dengan menggunakan model arsitektur ResNet 152 untuk mensegmentasi objek sekitar ruas jalan di lingkungan Universitas Nurtanio Bandung. Pengujian ini berdasarkan ketepatan dan nilai keakuratan yang dapat mensegmentasi objek yang telah dibuat seperti pada Gambar 4.20 dengan baik terhadap 34030 *frame* gambar secara *offline*. Objek yang tertangkap oleh kamera 360° ternormalisasi akan terindikasi oleh warna informasi atau nilai RGB yang telah dibuat seperti pada Gambar 4.21 sesuai dengan warna yang ada pada setiap objek.

Tabel 5.3 Data Klasifikasi Objek Pada ResNet 152

No	Objek	Warna Testing Gambar	Warna Ground Truth
1	Background	Hitam	Hitam
2	Sky	Biru Muda	Merah
3	Building	Abu-abu	Hijau
4	Pole	Kuning	Kuning
5	Road	Unggu	Biru Tua
6	Pavement	Biru Tua	Unggu
7	Tree	Hijau	Biru Muda

No	Objek	Warna <i>Testing Gambar</i>	Warna <i>Ground Truth</i>
8	<i>Motorcycle</i>	Biru Tua	Merah
9	<i>Car</i>	Merah	Abu-abu
10	<i>Pedestrian</i>	Coklat	Coklat

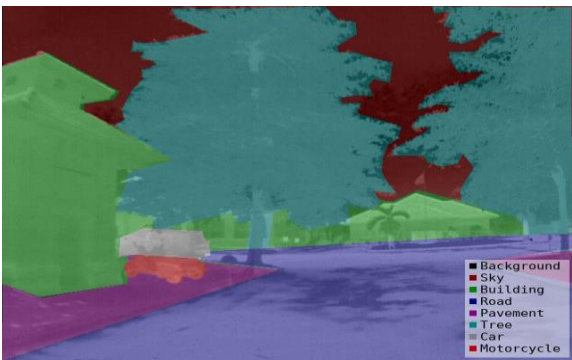
Pada Tabel 5.3 memperlihatkan objek, warna, nilai RGB yang telah dibuat serta presentase *Intersection over Union* (IoU) pada setiap objek dari hasil pelatihan model arsitektur ResNet 152. IoU merupakan nilai berdasarkan statistik kesamaan dan keragaman set sampel yang tujuannya untuk mengevaluasi area tumpang tindih antara *ground truth* dan *prediction segmentation*. *Ground truth* yang dimaksud adalah *dataset* yang telah dianotasi atau dilabeli secara manual, sedangkan *prediction segmentation* merupakan hasil segmentasi dari setiap objek yang tersegmentasi.

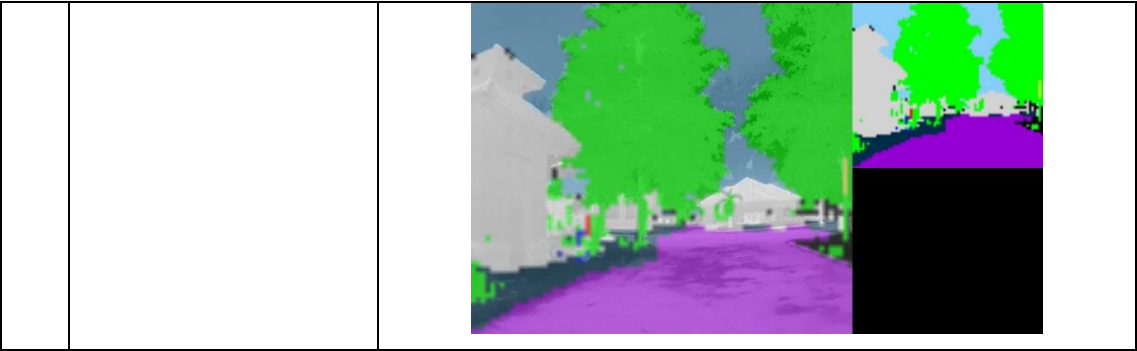
Pengujian pertama menggunakan model arsitektur ResNet 152 dikategorikan menjadi 3 kategori yaitu baik, Pendeteksian Kategori Cukup dan Pendeteksian Kurang Baik. Berikut ini masing masing kategori akan dijelaskan

1. Pendeteksian Baik

Pada Tabel 5.4 memperlihatkan perbedaan antara hasil deteksi objek secara segmentasi semantik dari *ground truth* dengan hasil *prediction segmentation*.

Tabel 5.4 Hasil Pendeteksian Kategori Baik Pada ResNet 152

No	Model	Hasil
1	Gambar <i>Thermal</i>	
2	<i>Ground Truth</i>	
3	<i>Prediction Segmentation</i>	



Pada Tabel 5.4 memperlihatkan perbedaan antara hasil deteksi objek secara segmentasi semantik dari *ground truth* dengan hasil *prediction segmentation*. Diperoleh hasil bahwa segmentasi objek pada *prediction segmentation* mendekati hasil *dataset* yang telah dianotasi (*ground truth*). Oleh sebab itu, pendeteksian dikategorikan dengan kategori sempurna, hasilnya diindikasikan oleh warna yang ada pada setiap objek yang tersegmentasi dengan baik. Hal ini disertai dengan presentase IoU dari setiap objek yang tersegmentasi.

2. Pendeteksian Kategori Cukup.

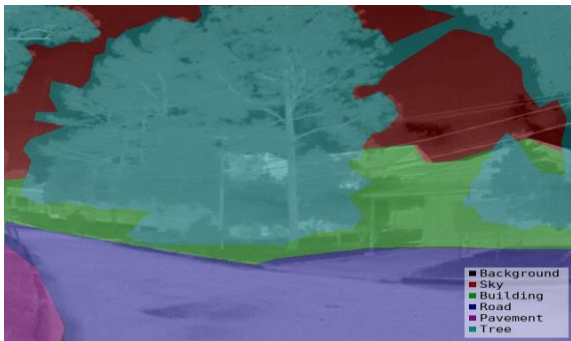
Tabel 5.5 Hasil Pendeteksian kategori Cukup Pada ResNet 152

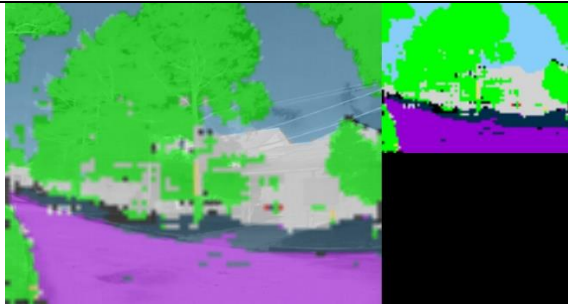
No	Model	Hasil
1	Gambar <i>Thermal</i>	
2	<i>Ground Truth</i>	

3	<i>Prediction Segmentation</i>	
---	--------------------------------	--

Pada Tabel 5.5 memperlihatkan perbedaan antara hasil deteksi objek secara segmentasi semantik dari *ground truth* dengan hasil *prediction segmentation*. Diperoleh hasil bahwa segmentasi objek pada *prediction segmentation* terdapat penyatuan indikasi warna pada satu objek, seperti pada objek jalan yang mengindikasikan warna objek bangunan dan tidak terdeteksinya sebagian objek. Pendeteksian Kategori Kurang Baik.

Tabel 5.6 Hasil Pendeteksian Kategori Kurang Baik Pada ResNet 152

No	Model	Hasil
1	Gambar <i>Thermal</i>	
2	<i>Ground Truth</i>	

3	Prediction Segmentation	
---	-------------------------	--

Pada Tabel 5.6 diperoleh hasil bahwa segmentasi objek pada *ground truth* dan *prediction segmentation* terdapat perbedaan yang terlihat jelas. Pada hasil *prediction segmentation* deteksi objek secara segmentasi semantik tidak tersegmentasi dengan baik pada setiap objek seperti bangunan, tumbuhan, jalan, motor dan mobil, sehingga pendeteksian ini dikategorikan dengan kategori pendeteksian gagal. Hasilnya diindikasikan oleh warna yang ada pada setiap objek-objek. Hal ini disertai dengan persentase IoU dari setiap objek yang tersegmentasi.

D. Analisa Hasil.

Pengukuran hasil tingkat akurasi dan persentase *Intersection over Union* seperti pada Tabel 5.2 menjadi tolak ukur utama dalam pengujian menggunakan 34030 *frame* gambar, hal ini dimaksudkan untuk mengetahui kinerja sistem pada model arsitektur ResNet 152 saat menguji cobakan pada gambar secara *offline*. Pada Tabel 5.11 diperlihatkan data pengukuran menggunakan model arsitektur ResNet 152 yang telah dilatih dan diujicobakan pada 34030 *frame* gambar.

Tabel 5.11 Data Pengukuran ResNet 152

No	Nomor Frame	Baik	Cukup	Kurang Baik
1	0-10000	5021	3005	1974
2	10001-20000	5325	3152	1523
3	20001-34030	8230	3176	2594

Pada Tabel 5.12 diperlihatkan nilai persentase hasil uji fungsi ResNet 152 pada 34030 *frame* gambar sesuai dengan kategori yang ada yaitu sempurna, tidak sempurna dan gagal. Pengkategorian dilakukan dengan cara pengamatan visual dengan membandingkan *dataset* yang teranotasi (*ground truth*) dengan hasil 34030 *frame* gambar yang tersegmentasi menggunakan model arsitektur ResNet 152 yang telah dilatih (*prediction segmentation*). Hasil persentase uji fungsi ResNet 152 didapat dengan rumus sebagai berikut :

$$\text{Presentasi} = \frac{\text{Jumlah Gambar}}{\sum \text{Jumlah Gambar}} \times 100\%$$

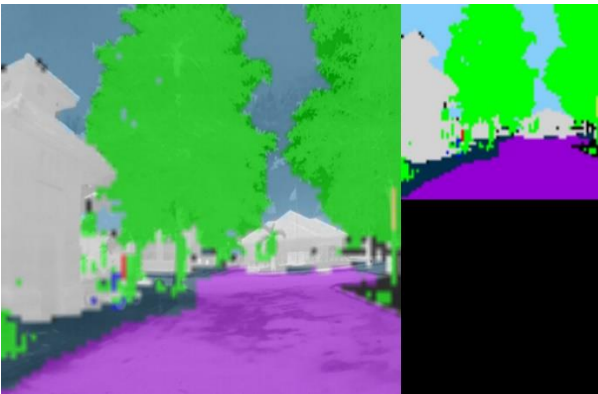
Tabel 5.12 Nilai Persentase Uji Fungsi ResNet 152 Pada 34030 *Frame* Gambar Sesuai Kategori

No	Kategori	Jumlah Gambar	Persentase
1	Baik	18576	$\frac{18576}{34030} \times 100\% = 54,58\%$
2	Cukup	9333	$\frac{9333}{34030} \times 100\% = 34,43\%$

No	Kategori	Jumlah Gambar	Persentase
3	Kurang Baik	6093	$\frac{6093}{34030} \times 100\% = 17,90\%$

Pada Tabel 5.15 diperlihatkan hasil dari kamera 360⁰, hasil pengujian dengan menggunakan model arsitektur ResNet 152 .

Tabel 5.8 Hasil Pengujian

No	Model	Keterangan
1	Kamera <i>Thermal</i>	
2	ResNet 152	

Pada Tabel 5.15 diperlihatkan hasil pengujian yang dilakukan dengan menggunakan model arsitektur ResNet 152 yang telah dilatih, yang mana gambar diambil menggunakan kamera 360⁰ yang sudah ternormalisasi. Terlihat hasil dari pengujian untuk mendeteksi objek yang telah dibuat sesuai dengan kelas label seperti bangunan, tumbuhan, motor, mobil, manusia dan jalan menggunakan arsitektur ResNet 152. Hal itu disebabkan karena pendeteksian objek secara segmentasi menggunakan ResNet 152 memiliki nilai presentasi *Mean IoU* 152 80,48%.

V. KESIMPULAN

A. Kesimpulan

Pada bab VI menunjukan hasil dari tugas akhir yang telah dilakukan pendeteksian objek secara *segmentasi semantic* dengan menggunakan metode FCN-Resnet 152, yang dapat dimanfaatkan untuk aplikasi kendaraan otonom. Adapun hasil dari penelitian dapat disimpulkan sebagai berikut:

1. Pendeteksian objek *segmentasi semantic* telah berhasil dilakukan pada objek sesuai dengan klasifikasi yang sudah dibuat dalam format txt yang selanjutnya dilakukan *labelling* pada *labelme*.
2. Kinerja sistem segmentasi semantik yang diujikan pada 34030 frame secara *offline* menghasilkan 18576 *frame* dengan kategori baik, 9333 *frame* kategori cukup dan 6121 *frame* kategori kurang baik dengan menggunakan model arsitektur ResNet 152 yang telah terlatih. Pelatihan dengan menggunakan model arsitektur Resnet 152 menghasilkan 80,48% *mean Intersection over Union* (IoU) dan 92.23% *Accuracy*.

B. Saran.

1. Sistem pendeteksian objek secara segmentasi semantik menggunakan metode FCN-Resnet 152 ini dapat dikembangkan dengan menggunakan lebih banyak *dataset* dan *epoch* yang lebih tinggi agar dapat meningkatkan *Intersection over Union* pada hasil *Training*.
2. Melatih *dataset* pada lingkungan yang berbeda agar model *software* yang akan dilatih dapat memiliki banyak data untuk pembelajaran.
3. Dapat menggunakan model arsitektur lainnya seperti ResNext 50 dan ResNext 152.

UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada pihak-pihak yang telah berkontribusi dalam penelitian ini hingga makalah ini dapat di publikasikan pada Seminar Nasional Sains Teknologi dan Inovasi Indonesia 2021, diantaranya kepada:

1. LPDP (Lembaga Pengelola Dana Pendidikan) sebagai sponsor penelitian.
2. BRIN (Badan Riset dan Inovasi Nasional) sebagai penyedia prasarana penelitian.
3. Universitas Nurtanio sebagai penyedia prasarana penelitian.
4. AAU (Akademi Angkatan Udara) sebagai sebagai konferensi dan prosiding.
5. Letda Lek Kevin Bayu Putra Sugandi, S.T. sebagai support system.
6. Serta rekan-rekan – rekan peneliti lainnya.

REFERENSI

- [1] Amer Sallam. (2021). *Early detection of Glaucoma using Transfer Learning from Pre-Trained CNN Models*. Diakses pada 08 Agustus 2022 dari, [Badan Pusat Statistik \(bps.go.id\)](https://bps.go.id), 2 - 4.
- [2] Radhika V M. (2020). *Movie Genre Prediction and Recommendation Using Deep Visual Features from Movie Trailers*. 3-5.
- [3] Agil Bintang Pratama. (2022) Deteksi Ruang Kosong pada Jalan Menggunakan *Semantic Segmentation* pada Mobil Otonom dari Institut Teknologi Sepuluh Nopember (ITS). 2-3.
- [4] Andreanov Ridhovan, dkk (2022) Penerapan Metode *Residual Network* (RESNET) dalam Klasifikasi Penyakit pada daun Gandum dari Universitas Singaperbangsa Karawang. 7-8.
- [5] Widi Harsono, Dkk (2021) Klasifikasi Covid-19 *Chest X-Ray* dengan Tiga Arsitektur CNN (*Resnet-152, Inception Resnet-V2, MobileNet-V2*). 7-9.
- [6] Avneet Pannu. (2015) *Artificial Intelligence and Application in Different Areas* dari DAV Institute of Engineering and Technology, Jalandhar India. 2-3.
- [7] John McCarthy. (1956) *The Dartmouth Summer Research Project On Artificial Intelligence* dari *Massachusetts Institute of Technology*. IEE 2-3
- [8] Marlina, N. N. A., Ariffin, D. M., Satyawan, A. S., Asysyakuur, M. I., Utamajaya, M. F., Satria, R. A., ... & Ema, E. (2021, December). Sistem Pendeteksi Pejalan Kaki di Lingkungan Terbatas Berbasis SSD Mobilenetv1 Menggunakan Gambar 360 Ternormalisasi. In *Prosiding Seminar Nasional Sains Teknologi dan Inovasi Indonesia (SENASTINDO)* (Vol. 3, pp. 111-122).
- [9] (Pooja Mahajan (2020) Fully Connected vs Convolutional Neural Network)
- [10] Caroline dkk. (2019). *Identifikasi Jalan Kampus Universitas Sriwijaya Berbasis Fully Convolutional Network*. *SURYA ENERGY*, 2(1), 353-358
- [11] (Baishideng. (2021). *Infographic With Icons And Timeline For Artificial Intelligence, Machine Learning And Deep Learning*.
- [12] (Lina, Q. (2019) *Apa Itu Convolutional Neural Network*)

-
- [13] He dkk. (2016). *Deep Residual Learning For Image Recognition*. *Proceeding of the IEEE Conference Computer Vision and Pattern Recognition*, 770-778.
 - [14] Aditama, Nadia. (2020). Segmentasi Citra Berbasis *Deep Learning* <https://informatika.stei.itb.ac.id/~rinaldi.munir/Citra/2021-2022/27-Kuliah-tamu-1.pdf>
 - [15] NVIDIA Developer. (2018). *Jetson AGX Xavier Developer Kit*. Diakses pada 16 Januri 2022 dari, <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>
 - [16] NVIDIA Developer. (2018). *Jetson AGX Xavier Developer Kit*. Diakses pada 16 Januri 2022 dari, <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>
 - [17] *LabelMe*. (2021). *LabelMe*. Diakses pada 16 Januari 2022 dari, <http://LabelMe2.csail.mit.edu/>
 - [18] FLIR ADK Thermal Vision Automotive Development Kit Teledyne FLIR
-