



Segmentasi Objek Berbasis Kamera Termal Menggunakan *Deep Learning (Pre-Trained Resnet 34)*

(Thermal Image-Based Object Segmentation Using Deep Learning (Pre-Trained Resnet 34))

Muhammad Fauzan Anggi F. L¹, Arief Suryadi Satyawan², Sri Desy Siswanti³

¹ Universitas Nurtanio

E-mail: fauzanlaksonoo03.fl@gmail.com

² Badan Riset dan Inovasi Nasional

E-mail: arief.suryadi@akane.waseda.jp

³ Universitas Nurtanio

desy0712unnur@gmail.com

Abstract— *Technology is developing very rapidly in a pluralistic country, among these technologies are electric vehicles without human intervention, we are more familiar with this with autonomous electric vehicles. The purpose of this autonomous electric vehicle is to suppress human negligence when driving, besides being able to make it easier for the driver to travel without the need to drive it. Before all these things can operate properly, the autonomous electric vehicle needs a detection scheme that can distinguish objects, the segmentation method uses a thermal camera sensor based on Deep Learning that is trained with the required data set. This method uses a Fully Convolutional Network (FCN) with a Residual Network 34 (ResNet 34) architectural model with an image dimension of 640x512 pixels as its feature extraction. The advantage of ResNet 34 is that it is able to do quite a lot of dataset training even though the hardware used is not the most qualified. The design of this object detection system uses semantic segmentation, Neural Network, and Image Processing methods as input signals in the form of images, and a FLIR thermal camera which is like an eye for a vehicle which receives an input signal, the process from start to finish is processed using the Jetson AGX Xavier. The capability of semantic segmentation was tested offline with 40216 image frames, there are three categories, namely, good, sufficient, and not good. Which includes the good category as many as 23389 frames, 10706 frames is a sufficient category, and for not good category there are 6120 frames. The mean Intersection over Union (IoU) obtained at the time of this study was 78.56%.*

Keywords— *Autonomous Electric Vehicles, Fully Convolutional Network, Residual Network.*

Abstrak— *Teknologi berkembang amat pesat berkembang di negara majemuk, diantara teknologi tersebut ialah kendaraan listrik tanpa intervensi manusia, kita lebih mengenal hal tersebut dengan kendaraan listrik otonom (KLO). Tujuan dari kendaraan listrik otonom ini untuk menekan kelalaian manusia pada saat berkendara, disamping itu mampu memberi kemudahan bagi pengemudinya berpergian tanpa perlu mengendarainya. Sebelum semua hal itu bisa beroperasi dengan baik, kendaraan listrik otonom tersebut membutuhkan skema pendeteksian yang bisa membedakan objek, metode segmentasi menggunakan sensor kamera termal yang didasarkan pada Deep Learning yang dilatih bersama kumpulan data yang diperlukan. Metode ini menggunakan Fully Convolutional Network (FCN) dengan model arsitektur Residual Network 34 (ResNet 34) yang berdimensi gambar sebesar 640x512 pixel sebagai ekstraksi fiturnya. Kelebihan ResNet 34 yaitu mampu melakukan training dataset cukup banyak meskipun hardware yang digunakan bukanlah yang paling mumpuni. Rancangan sistem pendeteksian objek ini mempergunakan metode segmentasi semantik, Neural Network, dan Image Processing sebagai sinyal masukan yang berupa gambar, dan kamera termal FLIR yang di ibaratkan sebagai mata untuk kendaraan yang mana menerima sinyal masukan,*

prosedur awal hingga akhir diproses menggunakan Jetson AGX Xavier. Kemampuan dari segmentasi semantik diuji secara offline dengan 40216 frame gambar, terdapat tiga kategori yaitu, baik, cukup, serta kurang baik. Yang termasuk kategori baik sebanyak 23389 frame, 10706 frame merupakan kategori cukup, dan untuk kategori kurang baik terdapat 6120 frame. Hasil mean Intersection over Union (IoU) yang didapatkan pada saat penelitian ini sebesar 78,56%.

Kata Kunci— **Autonomous Electric Vehicles, Fully Convolutional Network, Residual Network**

I. PENDAHULUAN

Seiring dengan pesatnya perkembangan teknologi di negara majemuk, diantaranya ialah kendaraan listrik tanpa intervensi manusia atau yang kita tahu dengan kendaraan listrik otonom (KLO). Pada masa ini banyaknya terjadi kecelakaan kendaraan terkhususnya di Indonesia, yang disebabkan oleh kelalaian pengendaranya seperti, kelelahan, mengantuk, tidak fokus, dan sebagainya, dengan faktor-faktor tersebut kemungkinan untuk mengurangi tingkat kecelakaan kendaraan di jalan tidaklah mudah. Suatu kendaraan dikembangkan yang bertujuan untuk mengurangi dan mengatasi kelalaian pengendara ketika berkemudi, kendaraan listrik otonom ini bertujuan menekan kelalaian manusia dalam berkendara, disamping itu mampu memberi kemudahan bagi pengemudinya berpergian tanpa perlu mengendarainya.

Sebelum semua hal itu bisa beroperasi dengan baik, kendaraan listrik otonom tersebut membutuhkan skema pendeteksian yang bisa membedakan objek, metode segmentasi yang didasarkan pada *Deep Learning* yang dilatih bersama kumpulan data yang diperlukan. Dalam tugas akhir ini, penulis akan melakukan penelitian perihal sistem pendeteksi objek dengan melaksanakan segmentasi objek menggunakan model arsitektur ResNet 34, yang berlandaskan sensor kamera termal menggunakan kamera FLIR sebagai alat pendeteksi objek, serta bisa diperluas sesuai dengan klasifikasinya (*Building, Tree, Pole, Pavement, Sky, Car, Motorcycle, Pedestrian*) sehingga gambar mampu disegmentasikan sesuai objek. Penggunaan peranti kamera termal ini dibantu oleh *mini computer Artificial Intelligence (AI)* yaitu Jetson AGX Xavier untuk mengelola data.

II. LANDASAN TEORI

Dengan semakin canggihnya teknologi dan semakin kuatnya kemampuan komputasi dari komputer, banyak aplikasi dan algoritma yang sebelumnya tidak memungkinkan untuk diterapkan di komputer jinjing atau telepon genggam sekarang bisa diimplementasikan dengan mudah. Salah satu teknik yang menjadi populer adalah kecerdasan buatan atau *Artificial Intelligence (AI)* menjadi terkenal karena banyak aplikasi jenis ini yang sekarang bisa ditemui sehari-hari seperti Siri pada iPhone dan *speech recognition* atau pengenalan suara pada telepon seluler Android. Seiring dengan munculnya istilah AI, Pembelajaran Mesin atau *Machine Learning (ML)*, Jaringan Syaraf tiruan *Artificial Neural Network (NN)*, dan *Deep Learning (DL)* masyarakat menjadi bingung dengan istilah-istilah tersebut. Pada Makalah ini, kami akan membahas mengenai perbedaan dari istilah-istilah tersebut. Pada umumnya AI adalah induk dari semua teknik tersebut, dan pada akhirnya DL adalah kasus khusus dan paling mutakhir pada AI [1].

ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*) adalah suatu kompetisi tahunan yang melombakan klasifikasi basis data gambar yang ada pada ImageNet yang memiliki 1000 kelas dan setiap tahunnya terdapat berbagai macam arsitektur jaringan *Convolutional Neural Network (CNN)* yang menjadi *state of the art*. CNN digunakan karena metode ini bekerja dengan sangat baik dengan volume data yang besar. Dari 1000 kelas, akan diambil 100

kelas yang akan dijadikan dataset penelitian untuk mengetahui strategi Pembelajaran seperti apa pada CNN yang dapat memiliki performa terbaik dengan berbagai skenario. Skenario terbaik didapatkan dengan melakukan training dengan ukuran gambar yang kecil, lalu melakukan training kembali dengan memperbesar ukuran gambar. Skenario ini mendapatkan akurasi

sebesar 75,82%, akurasi yang cukup tinggi untuk model yang dapat Mengklasifikasikan 100 kelas. Skenario ini memiliki performa yang paling baik dalam klasifikasi keseluruhan berdasarkan ukuran evaluasi confusion matrix [2].

A. Artificial Intelligent

Kecerdasan buatan atau *Artificial Intelligence* ialah satu dari beberapa cabang ilmu komputer yang mempelajari bagaimana melatih suatu mesin agar bisa melakukan sebuah pekerjaan yang apik dan rapi selayaknya manusia, atau bahkan dapat melakukan lebih baik dari manusia[3]. Kecerdasan buatan adalah kecerdasan yang ditambahkan oleh manusia ke dalam suatu sistem teknologi, diatur dan dikembangkan dalam konteks ilmiah, bentukan dari kecerdasan entitas ilmiah yang ada. Pada dasarnya AI adalah suatu pengetahuan yang membuat komputer dapat meniru kecerdasan manusia sehingga komputer dapat melakukan hal-hal yang dikerjakan manusia dimana membutuhkan suatu kecerdasan; misalkan melakukan analisa penalaran untuk mengambil suatu kesimpulan atau keputusan atau penerjemahan dari satu bahasa ke bahasa lain.

B. Deep Learning

Deep Learning adalah salah satu teknik pada *machine learning* yang memanfaatkan banyak *layer* pengolahan informasi *non-linier* untuk melakukan ekstraksi fitur, pengenalan pola, dan klasifikasi. *Deep learning* adalah salah satu cabang dari *machine learning* untuk penyelesaian masalah dengan dataset yang cukup besar menggunakan jaringan syaraf tiruan. Metode ini memiliki arsitektur yang amat kuat untuk *15 supervised learning*. *Supervised Learning* yaitu tipe pembelajaran yang berisi variabel masukan dan variabel keluaran dengan menggunakan satu algoritma atau lebih untuk mempelajari fungsi penggambaran dari masukan ke keluaran. Dengan lapisan yang ditambahkan lebih banyak maka model pembelajaran bisa mewakili data citra berlabel dengan lebih baik. Selain itu, program *machine learning* pada umumnya hanya menggunakan CPU dan RAM dalam proses komputasi, yang menentukan kecepatan dalam mengkomputasi data adalah spesifikasi dari CPU dan RAM. Selain membutuhkan CPU dan RAM pada proses komputasi, kemampuan GPU juga diperlukan pada proses ini agar dalam pemrosesan data yang besar dapat berlangsung dengan cepat [4].

Convolutional Neural Network (CNN) merupakan pengembangan dari *multilayer perceptron* (MPL) yang didesain untuk mengolah data dua dimensi dalam bentuk citra. CNN ini termasuk kedalam jenis *Deep Neural Network* karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra. CNN adalah sebuah arsitektur yang dapat dilatih dan terdiri dari beberapa tahap, Masukan dan keluaran dari setiap tahap itu terdiri dari beberapa *array* yang biasa disebut *feature map*. Setiap tahap terdiri dari tiga layer yaitu konvolusi, fungsi aktivasi *layer* dan *pooling layer*.

III. METODE/MODEL YANG DIUSULKAN

A. Metodologi Penelitian

Metodologi penelitian ini melakukan penelitian secara sistematis yang dimulai dari tahapan studi literatur sampai pembuatan laporan tugas akhir ini. Tahapan, metode, dan hasil penelitian akan diperlihatkan pada Tabel 1 dibawah ini.

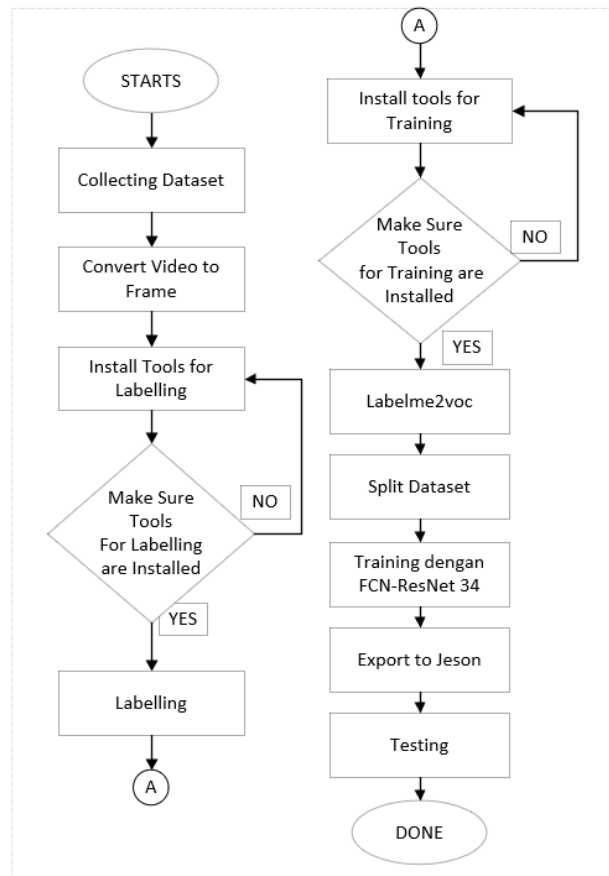
Tabel 1. Tahapan Penelitian

No	Tahapan	Metode	Hasil
1.	Studi kepustakaan	1. Peninjauan internet 2. Studi buku serta jurnal terkait	Interpretasi: 1. Pembahasan teknologi <i>Artificial Intelligence</i> , <i>Machine Learning</i> , <i>Deep Learning</i> , <i>Image processing</i> , <i>convolutional Neural Network</i> , <i>Fully Convolutional Network</i> , <i>Residual</i>

			<i>Network</i>, <i>Coffusion Matrix</i>, dan pemrosesannya. 2. Perangkat keras dan perangkat lunak yang digunakan. 3. Pemrograman menggunakan bahasa Python di PC serta NVIDIA Jetson.
2.	Desain system Software pendeteksi objek memakai sensor kamera termal	1. Memakai software perancangan python Anaconda prompt 2. Penerapan model software ResNet 34 serta melabeli gambar mengenakan Labelme	1. Memodifikasi serta mendesain <i>software</i> pendeteksi objek mendasar pada <i>Deep Learning</i> pre-trained ResNet 34. 2. Pengklasifikasian pada dataset hasil pendeteksian ada 10, yaitu : a. <i>Background</i> b. <i>Sky</i> c. <i>Building</i> d. <i>Pole</i> e. <i>Road</i> f. <i>Pavement</i> g. <i>Tree</i> h. <i>Car</i> i. <i>Pedestrian</i> j. <i>motorcycle</i>
3.	Dokumentasi <i>dataset</i>	1. <i>Dataset</i> diambil menggunakan kamera flir 2. Pengklasifikasian data memakai label	Menggunakan gambar sebanyak 660 sebagai <i>dataset</i> segmentasi semantik
4.	Pengaktualan <i>software</i> dalam Jetson AGX Xavier	1. Memakai <i>python</i> sebagai <i>software</i> pada pemrogramannya 2. Hasil dari model arsitektur FCN ResNet 34 diuji	Pengaktualan sensor kamera termal didukung <i>software</i> aplikasi pendeteksian objek yang berbasis <i>deep learning</i> .
5.	Pengukuran kemampuan <i>software</i> segmentasi semantik memakai FCN ResNet 34	1. Digunakannya metode <i>Confusion Matrix</i> menjadi konsep pencari nilai akurat pada model. 2. Memakai metode komparasi <i>ground truth</i> dengan model hasil bertujuan mengkatagorikan hasil uji terbaik, cukup serta gagal.	Percobaan <i>software</i> di Jetson AGX Xavier.

B. Flowchart

Gambar 1 di bawah ini merupakan langkah dan alur dari *system* segmentasi objek pada penelitian ini.

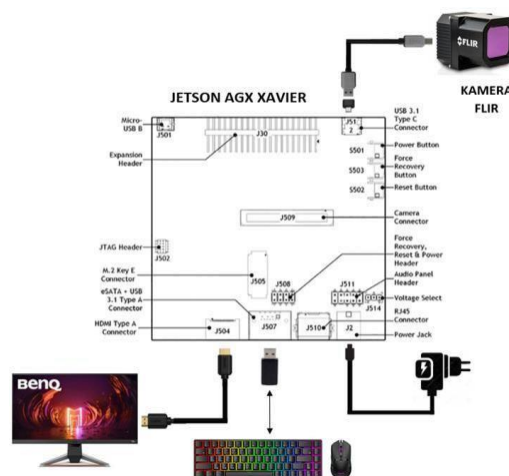


Gambar 1. Flowchart Sistem Segmentasi Objek

IV. HASIL/IMPLEMENTASI MODEL DAN PEMBAHASAN

Bagian keempat ini menjelaskan perihal desain hingga pengaplikasian segmentasi objek secara rinci, sesuai seperti klasifikasinya di area terbatas yang berdasar *Deep Learning* menggunakan arsitektur CNN Pre-trained ResNet-34 menggunakan gambar dari kamera termal FLIR.

A. Desain Model



Gambar 2. Ilustrasi Hardware

Gambar diatas menunjukan desain Gambar Hardware yang digunakan pada pembuatan rancang bangun segmentasi objek berbasis gambar thermal menggunakan deep learning (*pre-trained resnet 34*).

B. Pembahasan

Tahapan ini menjelaskan tahapan analisis pada sistem pendeteksian objek yang terintegrasi dengan segmentasi semantik, yang diperuntukan kendaraan listrik otonom menggunakan metode FCN bermodel arsitektur ResNet 34 yang dilakukan pengujian secara *offline*.

Pada penelitian ini pengujian terhadap model arsitektur ResNet 34, segmentasinya adalah objek ruas jalan dan sekitarnya. Hasil pengujian yang diambil dalam penelitian ini ialah ketepatan serta nilai keakuratan dalam mengsegmentasi objek yang sudah dibuat dengan baik terhadap 40216 *frame* gambar secara *offline*. Objek yang tertangkap oleh kamera termal akan teridentifikasi, lalu informasi tersebut akan ditandai dengan warna sesuai dengan warna *testing* yang dapat dilihat pada Tabel 2.

Tabel 2. Data Klasifikasi Objek Pada ResNet 34

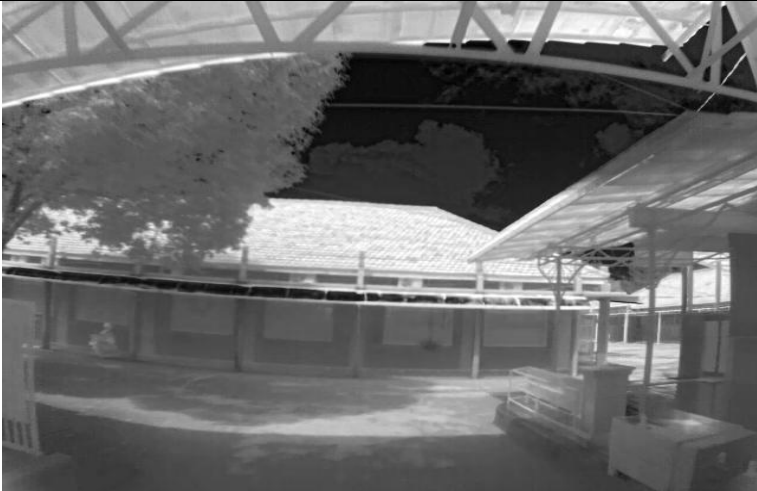
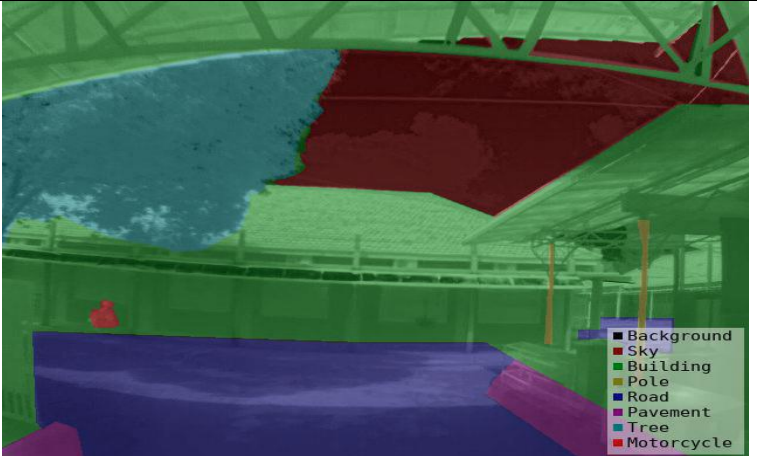
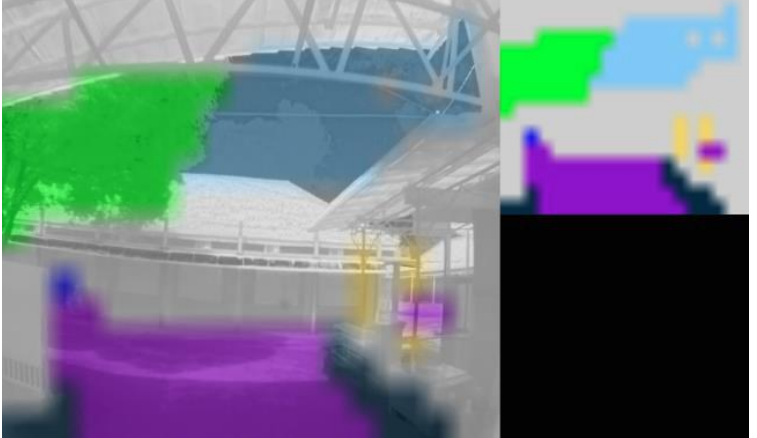
No	Objek	Warna <i>Testing Gambar</i>	Warna <i>Ground Truth</i>
1.	<i>Background</i>	Hitam	Hitam
2	<i>Sky</i>	Biru Muda	Merah
3.	<i>Building</i>	Abu-abu	Hijau
4.	<i>Pole</i>	Kuning	Kuning
5.	<i>Road</i>	Unggu	Biru Tua
6.	<i>Pavement</i>	Biru Tua	Unggu
7.	<i>Tree</i>	Hijau	Biru Muda
8.	<i>Motorcycle</i>	Biru Tua	Merah
9.	<i>Car</i>	Merah	Abu-abu
10.	<i>Pedestrian</i>	Coklat	Coklat

Tabel 2. menunjukan hasil dari pelatihan model arsitektur ResNet 34 yang berupa objek serta warna pada tiap-tiap program *Testing* dan *Ground Truth*. IoU adalah nilai dari statistik kesamaan beserta keragaman set contoh yang bertujuan mengevaluasi daerah berpotongan antara *ground truth* dan *prediction segmentation*. *Ground truth* adalah *dataset* yang telah diberi label secara manual dan *prediction segmentation* ialah hasil segmentasi yang didapatkan dari tiap objek yang sudah tersegmentasi. Pengetesan ini memakai model arsitektur ResNet 34 yang dikategorikan mejadi 3 yaitu, baik, cukup, serta kurang baik. Kategori tersebut akan dijelaskan secara teratur dibawah ini.

1. Pendeteksian Berkategori Baik

Hasil dari pendeteksian objek secara gambar termal, segmentasi semantik yang berasal dari *ground truth* dengan hasil *prediction segmentation*, yang ditunjukkan Tabel III

Tabel 3. Hasil Pendeteksian Berkategori Baik

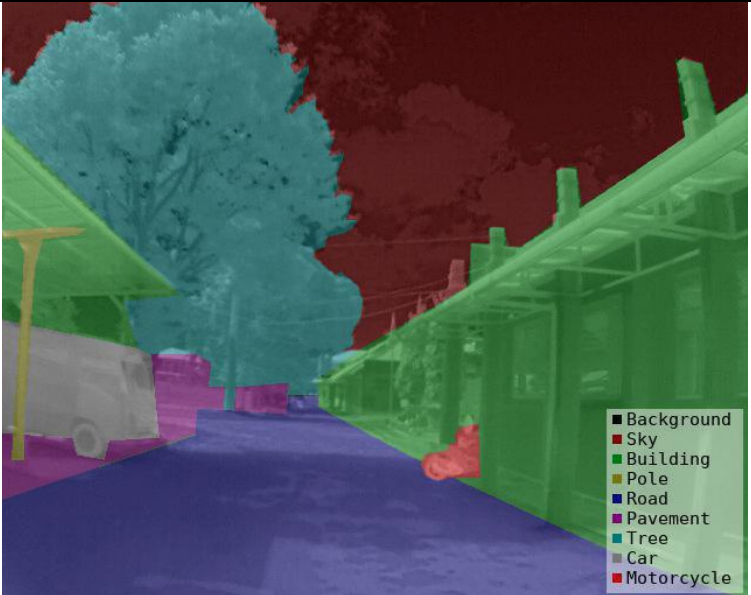
No	Model	Hasil Offline
1.	Gambar Termal	
2.	Ground Truth	
3.	Prediction Segmentation	

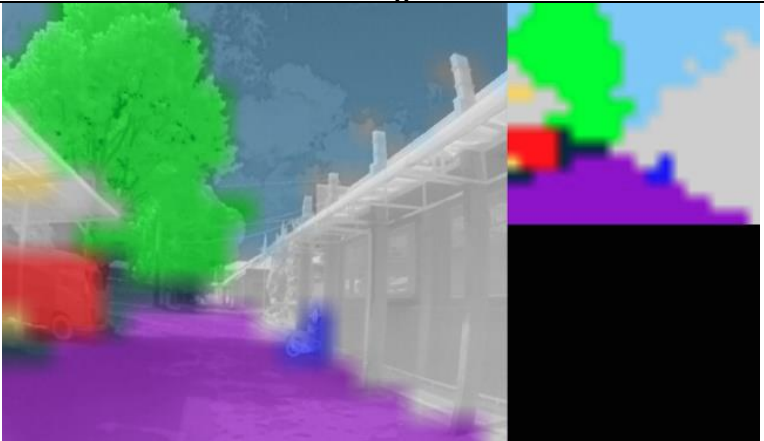
Tabel diatas menunjukkan perbedaan antara hasil pendeteksian objek menggunakan gambar termal, segmentasi semantik yang berasal dari *ground truth* serta hasil *predetection segmentation*. *Prediction segmentation* menunjukan bahwa hasul dari segmentasi objek mendekati hasil *dataset* yang telah diberi anotasi (*ground trurh*). Maka dari itu hasil dari pendeteksian masuk ke dalam kategori baik, hasilnya dapat dilihat pada warna yang ada ditiap objek yang tersegmentasi dengan baik

2. Pendeteksian Berkategori Cukup

Tabel dibawah ini menampilkan hasil pendeteksian objek secara segmentasi semantik dari *ground truth* dengan hasil *prediction segmentation*. Yang mana hasil tersebut didapatkan dari proses *training* yang menggunakan data gambar sebanyak 40.216 *frame* secara *offline*.

Tabel 4. Hasil Pendeteksian Berkategori Cukup

No	Model	Hasil Offline
1.	Gambar Termal	
2.	Ground Truth	

No	Model	Hasil Offline
3.	Prediction Segmentation	

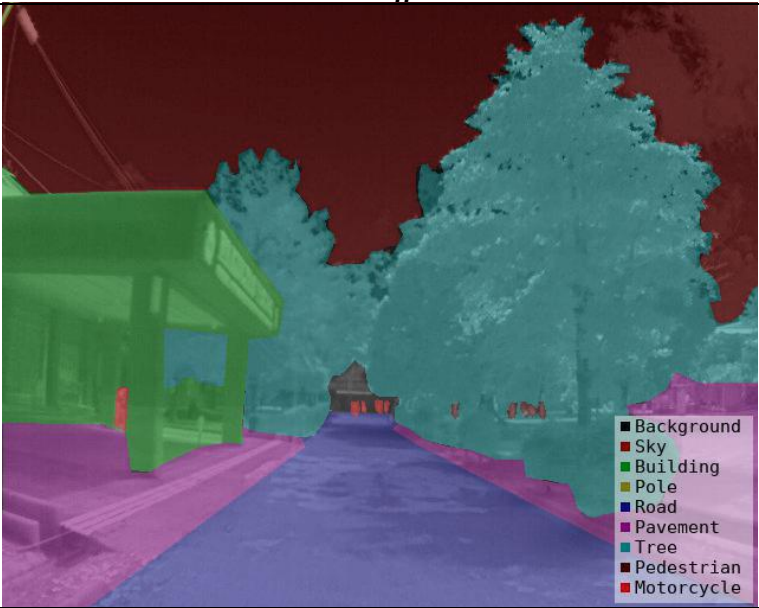
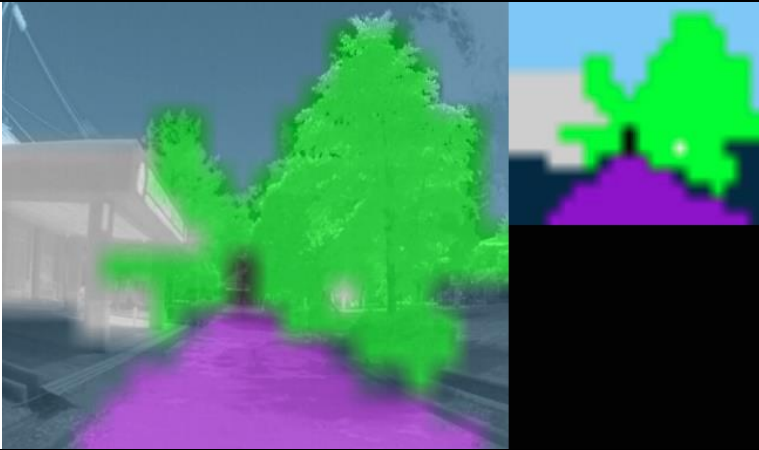
Tabel 4 menunjukkan perbedaan hasil antara deteksi objek segmentasi semantik dari *ground truth* dengan hasil *prededction segmentation*. Hasil yang didapatkan pada gambar *prediction segmentation* ialah penyatuan warna pada sebuah objek, contohnya pada gambar tersebut ialah objek yang seharusnya dideteksi *Pole* namun akhirnya menjadi objek *Car* dikarenakan posisi dua objek tersebut saling tumpang tindih. Maka dari itu hasil pendeteksian diatas masuk ke dalam kategori cukup, dengan alasan terdapat perbedaan hasil segmentasi objek di *ground truth* serta *prededction segmentation*.

3. Pendeteksian Berkategori Kurang baik

Tabel dibawah ini menampilkan hasil pendeteksian objek secara segmentasi semantik dari *ground truth* dengan hasil *prediction segmentation*. Yang mana hasil tersebut didapatkan dari proses *training* yang menggunakan data gambar sebanyak 40.216 *frame* secara *offline*

Tabel 5. Hasil Pendeteksian Berkategori Kurang

No	Model	Hasil Offline
1.	Gambar Termal	

No	Model	Hasil Offline
2.	Ground Truth	 The image shows a street scene with various objects segmented into different colors. A legend on the right identifies the classes: Background (black), Sky (dark blue), Building (green), Pole (yellow), Road (blue), Pavement (purple), Tree (light blue), Pedestrian (brown), and Motorcycle (red). The scene includes a building on the left, trees on the right, and a road leading into the distance.
3.	Prediction Segmentation	 The image shows the same street scene as the ground truth, but with the model's prediction. The segmentation is less accurate, with some objects incorrectly classified. For example, the road is mostly purple (Pavement) instead of blue (Road). The legend on the right is the same as in the ground truth image.

Pada Tabel V didapatkan hasil bila segmentasi objek pada *ground truth* serta *prediction segmentation* adanya sebuah perbedaan yang sangat jelas, setiap objek pada gambar hasil *prediction segmentation* tidak tsegmentasi dengan benar, pada hasil ini yang tidak terdeteksi dengan baik adalah *Motorcyle*, dan juga yang seharusnya *Road* berada pada jalanan namun pada hasil ini melebihi *Pavement*, sehingga hasil pendeteksian ini masuk ke dalam kategori yang kurang baik.

V. KESIMPULAN

Bagian kelima ini adalah kesimpulan dari pendeteksian objek secara segmentasi semantik pada penelitian ini menggunakan metode arsitektur CNN ResNet-34, yang mana akan dimanfaatkan terhadap kendaraan listrik otonom. Adapun kesimpulan yang didapatkan dari hasil penelitian ini adalah sebagai berikut:

1. Pendeteksian objek pada beberapa objek berhasil dilakukan, dengan membuat *classes* untuk label yang berformat txt. Selanjutnya melabeli gambar objek dengan aplikasi *LabelMe*.
2. Hasil daripada *training* model arsitektur ResNet-34 mendapatkan persentase sebesar 78,56% *mean Intersection over Union* (IoU) serta akurasi sebesar 95,73%.

3. Pada penelitian segmentasi semantik ini yang memakai model arsitektur ResNet-34 diuji secara *offline* menggunakan 40.216 *frame* yang menghasilkan 23.389 *frame* berkategori baik, 10.706 *frame* berkategori cukup serta 6120 *frame* yang berkategori kurang baik

UCAPAN TERIMA KASIH

Ucapan terima kasih kami sampaikan kepada pihak-pihak yang telah berkontribusi dalam penelitian ini hingga makalah ini dapat dipublikasikan pada Seminar Nasional Sains Teknologi dan Inovasi Indonesia 2022, 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 penyelenggara konferensi dan prosiding.

Serta rekan-rekan peneliti lainnya.

REFERENSI

- [1] Abu Ahmad. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning.
- [2] Kamal Hasan Mahmud, Adiwijaya, Said Al Faraby,. (2019). Klasifikasi Citra Multi-Kelas Menggunakan *Convolutional Neural Network*.
- [3] Athanasia Octaviani Puspita Dewi (2020), Kecerdasan Buatan sebagai Konsep Baru pada Perpustakaan
- [4] Triano Nurhikmat (2018). Implementasi *Deep Learning* untuk *Image Classification* Menggunakan Algoritma *Convolutional Neural Network* (CNN) pada Citra Wayang Golek.