

PENDETEKSI VISUAL MAKANAN DAN JUMLAH KALORINYA MENGGUNAKAN ALGORITMA MASK R-CNN BERBASIS BOT TELEGRAM

Raihanaldy Ash-Shafa ^{*1)}, Pulung Nurtantio Andono ²⁾

1. Universitas Dian Nuswantoro Semarang, Indonesia
2. Universitas Dian Nuswantoro Semarang, Indonesia

Article Info

Kata Kunci: Kalori Makanan, Deteksi, Mask RCNN, Telegram

Keywords: *Food calorie, Detection, Mask RCNN, Telegram*

Article history:

Received 9 October 2024
Revised 8 November 2024
Accepted 2 December 2024
Available online 1 March 2025

DOI :

<https://doi.org/10.29100/jipi.v10i1.6972>

* Corresponding author.

Corresponding Author

E-mail address:

author@email.ac.id

ABSTRAK

Algoritma deteksi visual MASK R-CNN merupakan teknologi yang dapat membantu user menjaga pola makan sehat dengan secara otomatis mendeteksi jenis makanan yang dikonsumsi. Sistem ini melibatkan pembuatan model berdasarkan kumpulan data, mengeksplorasi data, melatih model menggunakan algoritma Mask R-CNN, menguji model menggunakan gambar, dan menghubungkan model ke bot telegram menggunakan API. Kumpulan data dikumpulkan, dilatih, dan divalidasi, serta dikelompokkan ke dalam 40 kelas dengan berbagai jenis makanan dan minuman. Model ini memiliki tingkat akurasi total 78% dari 13 jenis gambar makanan yang diuji. Metode Mask Region Convolutional Neural Network (Mask R-CNN) dirancang untuk menyediakan akses cepat dan mudah ke informasi tentang jumlah kalori. Model ini dilatih menggunakan set data dari Alcrowd Food Recognition Challenge, dan mencapai tingkat akurasi 78%. Akurasi sistem dapat ditingkatkan dengan menggunakan set data yang lebih bervariasi dan mengoptimalkan pencahayaan pada gambar.

ABSTRACT

The MASK R-CNN visual detection algorithm is a technology that can help users maintain a healthy diet by automatically detecting the type of food consumed. The system involves building a model based on a data set, exploring the data, training the model using the Mask R-CNN algorithm, testing the model using images, and connecting the model to a telegram bot using an API. The dataset is collected, trained, and validated, and grouped into 40 classes with various foods and drinks. The model has a total accuracy rate of 78% from 13 type food images. The Mask Region Convolutional Neural Network (Mask R-CNN) method is designed to provide quick and easy access to information about calorie counts. The model was trained using a dataset from the Alcrowd Food Recognition Challenge, achieving an accuracy rate of 78%. The accuracy of the system can be improved by using more varied datasets and optimizing lighting on images.

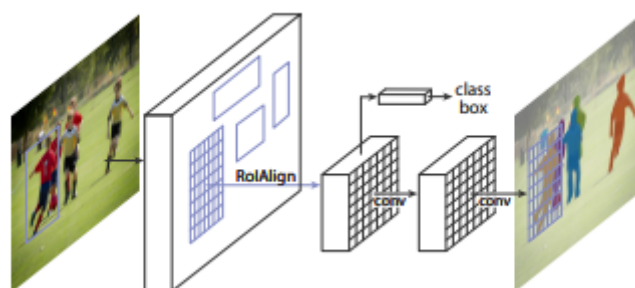
I. PENDAHULUAN

PERKEMBANGAN revolusi industri pada bidang teknologi telah memberikan dampak yang sangat signifikan bagi umat manusia. Melihat dari 10 tahun terakhir perubahan yang telah terjadi, memiliki dampak pada kondisi ekonomi, sosial dan budaya di masyarakat Indonesia [1]. Salah satunya berdampak pada pangan dan gizi. Ada berbagai macam makanan yang mungkin dengan adanya perkembangan industri membuat kandungan gizi dan kalorinya menjadi tidak seimbang salah satunya yaitu *Junk Food*. Yang mana konsumsi produk makanan dan minuman ini memiliki kalori, lemak, dan gula serta natrium yang berlebihan [2]. Sehingga tubuh menjadi kekurangan energi yang diperoleh melalui konsumsi makanan untuk beraktivitas dan melakukan metabolisme [3].

Salah satu tantangan utama menjaga pola makan yang sehat yaitu sulitnya untuk mengidentifikasi jenis makanan dan kalori yang terkandung didalamnya. Pengetahuan tentang jumlah kalori dan kandungan gizi suatu makanan seringkali terbatas, terutama di kalangan masyarakat perkotaan maupun pedesaan yang tidak memiliki akses atau pemahaman mendalam tentang komposisi gizi. Keadaan ini diperparah dengan terbatasnya alat dan metode yang dapat memberikan informasi gizi secara cepat dan juga akurat, sehingga masyarakat cenderung mengandalkan perkiraan dan informasi yang tidak akurat dalam mengambil keputusan mengenai pola makan.

Dari situlah peran teknologi hadir untuk membantu pola makan yang sehat. Dengan adanya kemajuan dalam visi komputer memungkinkan pengembangan algoritma yang dapat mengidentifikasi makanan dan menghitung kandungan kalornya. Kemajuan teknologi pada visi komputer memungkinkan deteksi dan analisis objek secara real-time, menggunakan algoritma seperti jaringan saraf konvolusional (CNN) untuk segmentasi objek yang kompleks [4]. Sedangkan pada penelitian ini algoritma yang menjanjikan untuk digunakan yaitu algoritma hasil turunan dari CNN yaitu Mask R-CNN [5], yang dapat mendeteksi dan mensegmentasi objek secara akurat, termasuk objek makanan. Beberapa bidang penelitian, termasuk manufaktur, telah menunjukkan kemampuan algoritma ini untuk mendeteksi cacat pada bagian produksi seperti komponen pengelasan [6]. Algoritma ini juga telah efektif digunakan dalam bidang transportasi seperti analisis lalu lintas, manajemen tempat parkir, dan deteksi pelanggaran lalu lintas [7]. Melalui algoritma ini, sebuah sistem dapat dikembangkan secara otomatis untuk mengenali jenis makanan dan memperkirakan kandungan kalornya berdasarkan gambar.

Berbagai tugas penelitian, termasuk segmentasi semantik, pengenalan objek, klasifikasi gambar, dan segmentasi, dilakukan dengan menggunakan visi computer [3][8]. Agar komputer dapat mengidentifikasi objek dalam bentuk foto dengan cara yang mirip dengan manusia, klasifikasi gambar bertujuan untuk meniru kemampuan manusia dalam memahami informasi gambar digital [9]. Dengan menggunakan gambar yang akan dikonversi menjadi data yang dapat diidentifikasi oleh sistem, algoritma *Mask Region Convolutional Neural Network* (MASK R-CNN) menggunakan gambar agar sistem dapat mengenali jenis makanan dengan mencocokkan gambar.



Gambar. 1. Arsitektur Mask R-CNN

Pada penelitian ini, sistem yang dikembangkan akan mampu mendeteksi secara otomatis jenis makanan yang akan dikonsumsi oleh *user* [10] lalu didokumentasikan yang berisi data jumlah dan kadar kalori menggunakan algoritma R-CNN. Tujuan dari penelitian ini yaitu untuk mengembangkan sistem berbasis Mask R-CNN yang mampu mendeteksi dan mengidentifikasi jenis makanan secara otomatis serta memperkirakan jumlah kalori. Penelitian ini akan menggunakan algoritma *Mask Region Convolutional Neural Network* (MASK R-CNN) untuk mendeteksi objek dan untuk menghitung estimasi jumlah kalori pada makanan tersebut. Melalui platform Telegram yang mempunyai fitur bot untuk implementasi machine learning, Telegram dipilih sebagai platform untuk mengintegrasikan sistem Machine Learning karena menawarkan API yang kuat, mudah digunakan, dan terdokumentasi dengan baik, yang memudahkan komunikasi antara model Machine Learning dan user. Dengan bantuan Telegram Bot API, user dapat mengirimkan data seperti gambar atau teks yang kemudian diteruskan ke model Machine Learning untuk dianalisis dan merespons hasil inferensi dengan cepat. Kecepatan komunikasi ini didukung oleh metode seperti *webhook* atau polling melalui koneksi server Telegram, yang memungkinkan komunikasi real-time antara user dan model ML. Selain itu, Telegram juga memiliki keuntungan praktis seperti keamanan dan privasi data melalui enkripsi end-to-end, memastikan bahwa data tetap aman selama interaksi. Telegram memungkinkan pengiriman file berukuran besar hingga 50 MB, memudahkan user untuk mengunggah gambar atau dataset yang akan diproses oleh model ML. Fitur lain yang menjadi keuntungan adalah antarmuka yang intuitif dan ramah user, yang memungkinkan akses mudah di berbagai perangkat seperti Android, iOS, desktop, dan web. Dengan berbagai keunggulan ini, Telegram tidak hanya menyediakan komunikasi yang cepat dan efisien tetapi juga memastikan kenyamanan, keamanan, dan fleksibilitas dalam mengintegrasikan sistem Machine Learning dengan interaksi user

Pemilihan metode ini dipertimbangkan karena kemampuannya dalam menyediakan segmentasi secara instan [11]. Mask R-CNN memiliki keunggulan yang signifikan dibandingkan dengan metode lain seperti SSD (Single Shot Multibox Detector) dan YOLO (You Only Look Once), terutama dalam konteks tugas deteksi objek yang memerlukan segmentasi yang mendetail. Salah satu contoh keunggulan Mask R-CNN terlihat dalam penerapannya untuk deteksi kapal dalam citra pemantauan jarak jauh (remote sensing) [12]. Studi membandingkan metode SS R-CNN, SSD, dan Mask R-CNN untuk mendeteksi kapal di lingkungan laut yang kompleks menunjukkan bahwa SS R-CNN dengan fitur self-supervised learning memberikan hasil yang lebih baik dibandingkan SSD dan Mask R-CNN tradisional, terutama dalam mendeteksi kapal kecil dan sedang.

Berbeda dengan YOLO, yang merupakan model deteksi objek berkecepatan tinggi dan melakukan deteksi dalam satu tahap, Mask R-CNN memiliki keuntungan dalam hal akurasi segmentasi tingkat piksel melalui metode region proposal dan RoIAlign. Keunggulan ini memungkinkan Mask R-CNN untuk mendeteksi dan membedakan objek-objek yang berdekatan dengan detail lebih spesifik dibandingkan dengan metode YOLO yang lebih cepat tetapi terkadang kurang presisi dalam pemisahan objek dengan tumpang tindih atau objek kecil [13][14].

Dengan demikian, hasil dari sistem ini diharapkan dapat memberikan kontribusi nyata dalam mendukung pola makan yang lebih sehat dengan menyediakan informasi kalori dan gizi yang mudah diakses dan akurat kepada masyarakat umum.

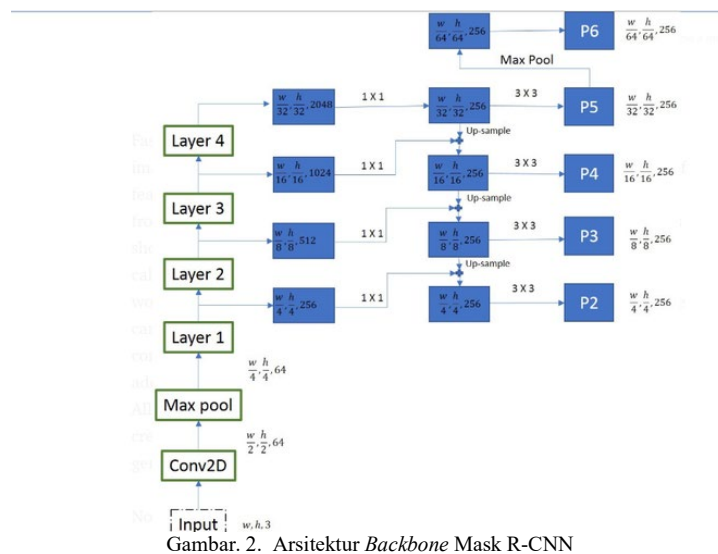
II. METODOLOGI PENELITIAN

A. Deep Learning

Deep Learning adalah cabang dari *machine learning* di mana parameternya masih diperlukan untuk menerapkan model *machine learning* untuk menghasilkan prediksi yang akurat. Dengan mensimulasikan bagaimana otak manusia menganalisis informasi, teknik ini memungkinkan mesin untuk melihat pola, mencapai kesimpulan, dan menghasilkan prediksi yang tepat. Sebuah algoritma *deep learning* menggunakan *Long Short-Term Memory*, yang memiliki banyak lapisan yang terkubur, untuk menghasilkan pola yang konsisten [15]. Beberapa lapisan jaringan saraf digunakan dalam teknik *deep learning*. Dimana data yang sebelumnya telah disederhanakan akan ditransmisikan ke lapisan berikutnya oleh jaringan saraf [16]. Untuk membuat keputusan berdasarkan penalaran manusia, *deep learning* menggunakan struktur algoritmik yang saat ini disebut sebagai *artificial neural network* (ANN). ANN sendiri adalah jenis sistem komputer yang mendapatkan inspirasi dari jaringan saraf biologis yang terlihat pada otak manusia. ANN terdiri dari banyak unit pemrosesan sederhana, atau “neuron”, yang terhubung satu sama lain. Setiap neuron menerima informasi, memprosesnya, dan kemudian mengeluarkan hasilnya. Contoh dari Model ANN, seperti *Convolutional Neural Networks* (CNN), dapat digunakan untuk mengklasifikasikan objek dalam bentuk gambar

B. Mask Region Convolutional Neural Network

Mask Region Convolutional Neural Network (MASK R-CNN) adalah jaringan saraf dalam yang bertujuan untuk memecahkan masalah klasifikasi pola dalam pembelajaran mesin atau visi komputer. Dengan kata lain, dapat membedakan berbagai hal dalam bentuk dan video. Bedanya dengan R-CNN yaitu Mask R-CNN adalah versi yang lebih cepat dari R-CNN dengan menambahkan cabang untuk prediksi segmentasi wajah untuk setiap *Region of Interest* (RoI), secara paralel dengan cabang yang ada untuk klasifikasi dan regresi bounding box [17]. Mask R-CNN memiliki sifat yang memungkinkannya mengatasi tantangan pengenalan gambar di lanskap kompleks dengan melakukan segmentasi berdasarkan fitur spektral dan tekstur karena keanekaragaman filogenetik [18]. Mask R-CNN memiliki dua bagian. Pertama, memberikan gambaran tentang wilayah di mana benda-benda tersebut mungkin ada secara terpadu. Kedua, memprediksi kelas objek, memodifikasi kotak pembatas, dan membuat representasi objek tingkat piksel berdasarkan ide dari langkah pertama. Kedua bagian tersebut melekat pada struktur *backbone* atau jaringan neural dalam gaya *Feature Pyramid Network* (FPN). *Backbone* sendiri memiliki beberapa varian seperti ResNet, VGG, dan Inception. Arsitektur pada ResNet umumnya mirip dengan Faster R-CNN, namun ResNet-FPN menyarankan beberapa perubahan. Ini termasuk generasi ROI multi-layer dimana keakuratan desain ResNet sebelumnya ditingkatkan dengan jaringan piramida fitur multi-layer ini, yang menghasilkan ROI dengan berbagai ukuran.



Gambar. 2. Arsitektur Backbone Mask R-CNN

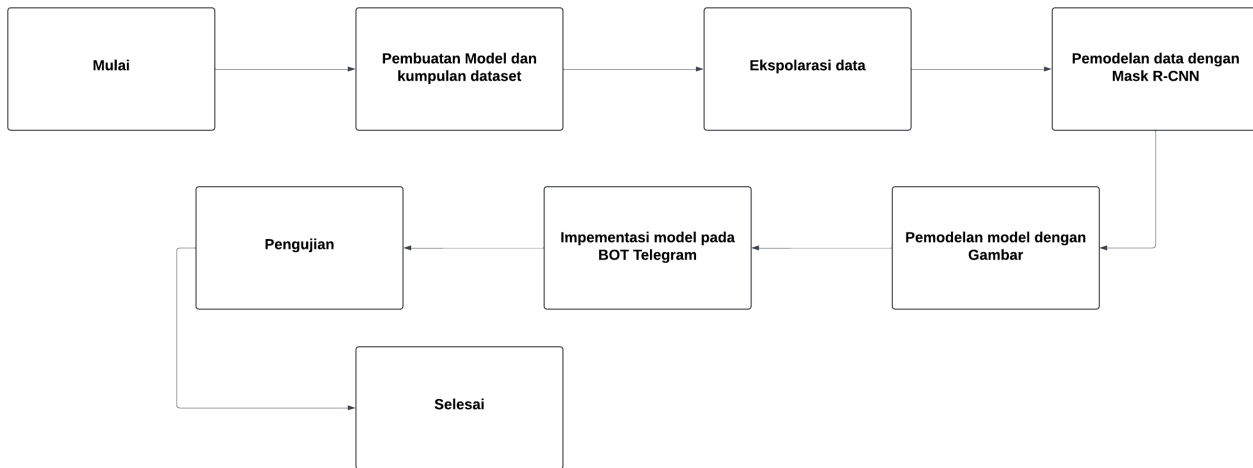
Dapat dilihat pada gambar 2, setiap bagian layer pada fitur digandakan dan ukuran layer fitur dipotong menjadi dua pada setiap tingkat. Output empat lapisan (lapisan 1, 2, 3, dan 4) digunakan. Jalur atas-bawah adalah metode yang digunakan untuk membuat fitur *maps* final. Dengan menggunakan prosedur kelas atas, dapat dimulai dengan fitur *maps* teratas ($w / 32, h / 32, 256$) dan melanjutkan ke yang lebih besar. Konvolusi 1×1 juga digunakan untuk mengurangi jumlah saluran menjadi 256 sebelum pengambilan sampel. Output up-sampling dari iterasi sebelumnya kemudian ditambah dengan ini, elemen demi elemen. Empat fitur *maps* terakhir (P2, P3, P4, dan P5) dihasilkan dengan menerapkan lapisan konvolusi tiga X tiga pada semua output. Operasi penggabungan maksimal dari P5 digunakan untuk membuat peta fitur kelima (P6).

Penerapan metode Mask R-CNN dapat disaksikan pada penelitian terkait deteksi objek sampah yang dilakukan Shuijing Li, Ming Yan [19] menggunakan metode Mask R-CNN. Pada penelitian ini membahas masalah pemilahan sampah dan menyarankan cara untuk membantu proses pemilahan sampah dan mengurangi pemborosan waktu. Proses pembuatan dataset berdasarkan kriteria klasifikasi sampah yang terdapat pada kota dan proses pelatihan data menggunakan Mask Scoring RCNN. Mask Scoring RCNN menambahkan struktur di atas Mask RCNN untuk meningkatkan struktur penilaian menjadi lebih akurat. Pada tahap pengambilan sampel, kualitas mask di Mask R-CNN biasanya tidak terkait dengan klasifikasi. Oleh karena itu, Mask Scoring R-CNN menambahkan modul MaskIoU ke basis asli dan mengintegrasikan titik prediksi dan fitur ROI yang diperoleh setelah masking dengan lapisan volume dan *full connection layer* untuk mendapatkan titik sampel. Keakuratan Mask Scoring R-CNN mencapai 65.8% karena proses pengumpulan data di bagi menjadi empat tahap. Algoritma Mask Scoring RCNN lebih baik untuk menemukan banyak objek sekaligus.

Sedangkan pada penelitian serupa dengan objek berupa makanan yang dilakukan oleh Darma Udayana, Nugraha [10] yang menggunakan metode CNN dan dataset berupa 1000 data citra makanan yang menghasilkan akurasi tertinggi sebesar 98% dan akurasi terendah sebesar 66%. Dengan Kesimpulan bahwa hasil akurasi klasifikasi yang rendah diakibatkan karena fitur citra makanan yang diuji memiliki banyak kemiripan dengan fitur citra makanan lain sehingga sistem salah dalam mengklasifikasikan makanan tersebut. Dari kedua penelitian tersebut, dapat disimpulkan bahwa usern Mask R-CNN pada useran visi komputer memiliki keunggulan pada keakuratan yang tinggi dengan keakuratan diatas 50% pada kedua penelitian diatas. Perlu ditekankan bahwa keakuratan pada kedua penelitian diatas tidak lepas pada faktor metode atau model machine learning yang digunakan, serta useran dataset yang legkap dan tepat membuat hasil menjadi lebih akurat.

C. Metode Penelitian

Secara umum, penelitian ini akan menghasilkan sebuah sistem yang dapat mengidentifikasi berbagai jenis makanan dan mengestimasi berapa banyak kalori yang dikandungnya [20]. Analisa masalah dimulai dari pembuatan model dan pengumpulan dataset, pengolahan model dan data menggunakan Mask R-CNN, hingga pengujian gambar dan implementasi pada *bot telegram*. Seluruh metode dilakukan dan diuji menggunakan Google Colab sebagai platform untuk melatih model dengan data, terutama dalam konteks machine learning dan deep learning.



Gambar. 3. Alur kerangka penelitian

Berdasarkan alur kerangka penelitian yang telah dibuat pada Gambar 3, Pada langkah pertama yang dilakukan yaitu pembuatan model berdasarkan pada kumpulan data yang dikumpulkan dengan menggunakan Kaggle secara langsung. Dari data yang telah dikumpulkan melalui Kaggle, dilanjutkan dengan tahap eksplorasi data untuk mengklasifikasikan data dimana pada proses ini dataset akan melakukan cleaning dan anotasi. Lalu, model akan dilatih menggunakan algoritma Mask R-CNN untuk bisa mengenali objek lewat gambar. Kemudian model akan diuji menggunakan beberapa gambar dan melihat bagaimana respon model untuk mengenali objek. Setelah pengujian model pada gambar, maka model akan langsung diimplementasikan dan disambungkan pada *bot* telegram dengan menggunakan API dan melatih *bot* tersebut. Terakhir yaitu pengujian dan dokumentasi pada *bot* telegram sera penarikan kesimpulan pada penelitian kali ini. Untuk lebih jelasnya bagaimana penelitian ini bekerja, maka akan dijabarkan sebagai berikut

1. Pembuatan model dan kumpulan dataset

Model dibuat berdasarkan dataset yang diambil pada sebuah dataset pengujian yang digunakan dari AICrowd Food Recognition Challenge [21], terdiri dari 40 jenis makanan dan minuman yang diklasifikasikan secara manual untuk kebutuhan model. Dataset diproses dan dimuat dari file JSON dan gambar yang disertakan lalu data dilakukan proses *data cleaning* dan anotasi format dataset berbasis COCO (*Common Objects in Context*) yang mencakup informasi seperti kategori makanan, bounding box, dan mask yang digunakan untuk tugas segmentasi gambar. Setiap gambar akan diubah hingga memiliki ukuran dimensi 256 x 256 piksel dengan tiga saluran warna. Kumpulan data diperiksa untuk memeriksa struktur, distribusi, dan frekuensi jenis makanan, dan divisualisasikan agar model dapat mengidentifikasi makanan berdasarkan karakteristik yang relevan. Selain itu model dibuat menggunakan versi sebagai berikut

TABEL I
 VERSI PEMODELAN

Nama Model	Versi	Fungsi
Python	3.10.12	Bahasa pemrograman utama
Numpy	1.23.5	Pustaka untuk komputasi numerik dan pengolahan data.
Tensorflow	2.12.0	Framework untuk membuat, melatih, dan menyebarkan model pengenalan citra
Keras	2.12.0	API tingkat tinggi untuk membangun dan melatih model neural network dengan antarmuka sederhana.
Scikit-image	0.19.1	Pustaka untuk pengolahan citra, menyediakan fungsi untuk analisis dan transformasi gambar.
Telebot	0.0.5	Pustaka untuk mengembangkan bot Telegram, menghubungkan aplikasi machine learning.
PyTelegramBotAPI	4.23.0	Pustaka untuk membuat dan mengelola bot Telegram.
Googletrans	4.0.0-rc1	Pustaka untuk menerjemahkan teks menggunakan Google Translate API.

2. Eksplorasi data dan segmentasi

Pada tahap ini, dataset yang telah dikumpulkan dan diproses pada segmen atau tahap sebelumnya diklasifikasikan dan dibagi meliputi proses pemahaman struktur dan sebaran data gambar pangan, identifikasi jenis pangan, dan pengorganisasian data untuk meningkatkan akurasi model klasifikasi makanan. Proses ini memverifikasi hasil anotasi dan deteksi objek pada dataset sebelum memulai tahap pelatihan model, seperti yang dilakukan dengan visualisasi dan pengecekan bounding box, mask, dan kelas objek. Lalu data akan disegmentasikan dan diisi dengan koordinat yang menggambarkan bentuk objek dalam gambar dalam bentuk format polygon. Koordinat tersebut membentuk batas objek, yang akan digunakan untuk menghasilkan mask biner. Model akan belajar untuk mengidentifikasi dan memetakan area ini sebagai bagian dari objek yang relevan sebelum dimasukkan pemodelan dengan Mask R-CNN.

3. Pemodelan data dengan Mask R-CNN

Fase ini berfokus pada penerapan jaringan Mask R-CNN untuk mendeteksi dan mengenali makanan dalam gambar yang diunggah melalui bot Telegram. Sistem akan melakukan prediksi pada model yang telah dilatih menggunakan dataset "crowdai-food-challenge". Konfigurasi ini mengoptimalkan proses inference dengan ukuran gambar yang lebih besar (320x320), satu gambar per GPU, dan pengaturan ambang batas yang lebih rendah untuk memfasilitasi deteksi objek yang lebih fleksibel meskipun dengan tingkat kepercayaan yang lebih rendah. Konfigurasi pelatihan ini memastikan bahwa model dapat dilatih dengan baik sesuai dengan ukuran gambar dan jumlah kelas dalam dataset.

4. Pemodelan dan pengujian dengan gambar

Pada Tahap ini, model yang dilatih dan dievaluasi pada tahap sebelumnya diterapkan pada gambar baru yang diunggah melalui bot Telegram Model ini mendeteksi makanan dalam gambar yang menghasilkan kotak pembatas dan masker segmentasi untuk setiap objek yang terdeteksi lalu menguji gambar dengan model Mask R-CNN untuk mendeteksi objek (makanan) dalam gambar, menampilkan hasil deteksi dalam bentuk visual dan teks (kelas dan probabilitas) untuk memastikan bahwa model berfungsi dengan baik sebelum diintegrasikan ke dalam sistem yang lebih besar, seperti bot Telegram. Hasil deteksi ini divisualisasikan dan dikirimkan kembali oleh sistem sehingga dapat melihat objek yang terdeteksi dan klasifikasinya.

5. Implementasi bot telegram

Selama fase implementasi bot Telegram, model Mask R-CNN yang terlatih akan diintegrasikan ke dalam bot, memungkinkan mengunggah gambar makanan dan mendapatkan hasil pengenalan secara instan. Tahapan bermula pada *bot* yang dikembangkan menggunakan API Python Telegram dan terhubung ke server yang menjalankan model Mask R-CNN. Bot lalu dikonfigurasi untuk mengenali dan merespons berbagai jenis pesan, termasuk gambar, teks, atau perintah. Saat mengirim sebuah gambar, bot mendeteksi pesan jenis "Foto" dan mengunduh gambar dari server Telegram. File gambar ini disimpan sementara dengan nama khusus berdasarkan ID untuk menghindari duplikasi. Kemudian bot memproses gambar melalui mask R-CNN dan menganalisis gambar untuk mendeteksi objek makanan. Model ini membuat masker objek, kotak pembatas, dan label untuk jenis makanan pada gambar. Lalu yang terakhir yaitu fungsi penerjemah dimana nama jenis makanan yang dikenali model dari bahasa Inggris di terjemahkan ke bahasa Indonesia. Hasil pengenalan dan terjemahannya dikemas menjadi teks yang responsif. Lalu gambar hasil deteksi dengan masker dan label makanan dikirim melalui bot bersama dengan pesan teks yang merinci item dan jumlah kalori yang terdeteksi.

6. Tingkat akurasi uji coba

Pengujian merupakan langkah penting dalam penelitian untuk melihat apakah suatu algoritma berhasil. Pada tahap ini dilakukan pengujian menggunakan algoritma masked R-CNN untuk jumlah kalori berbasis bot telegram. Berikut rumus yang digunakan menghitung tingkat akurasi pada uji coba gambar:

$$Akurasi = \left(\frac{\sum Data\ benar}{\sum Data\ uji} \right) \times 100\%$$

Keterangan:

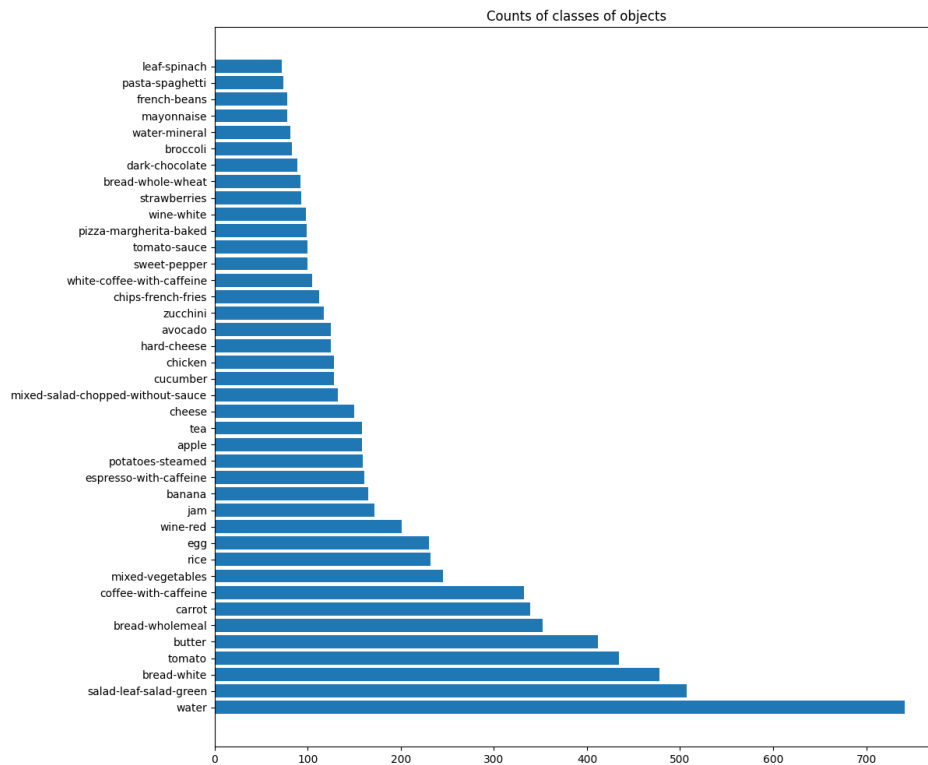
\sum Data Benar = Banyak Data Benar

\sum Data Uji = Banyak Data Uji

Data uji sendiri terdiri dari 13 jenis makanan dan minuman yang dipilih secara acak berdasarkan dari klasifikasi dataset. Uji keakuratan data uji dengan menggunakan rumus yang tercantum di atas. Dengan kata lain, menghitung nilai akurasi dengan membagi jumlah data pengujian yang benar dengan jumlah data pengujian dan dikalikan dengan 100%.

III. HASIL DAN PEMBAHASAN

Setelah dataset dikumpulkan pada tahap awal, maka pada tahap eksplorasi dataset model di training sebanyak 5545 gambar dan divalidasi sebanyak 291 gambar. Lalu setelah itu, dataset akan *dimasking* untuk dapat memberikan ID dan menamai setiap makanan atau minuman yang terdeteksi.



Gambar. 4. Klasifikasi dataset penelitian



Gambar. 5. Anotasi sampel data

Sampel diambil dari kumpulan data pelatihan untuk memeriksa apakah kumpulan data tersebut dikenali dengan benar seperti pada gambar 5 dimana hasil sampel pada gambar dataset dipilih secara acak untuk memverifikasi bahwa anotasi sudah akurat menunjukkan gambar dataset makanan. Setelah berhasil, langkah selanjutnya mensegmentasikan objek dari model deteksi agar gambar dapat dilihat dan divisualisasikan lalu menghasilkan *anchors* yang berguna untuk mencocokkan objek yang sebenarnya dengan prediksi yang dihasilkan oleh model. total jumlah anchors yang dihasilkan yaitu 16368 dengan skala (32, 64, 128, 256, 512) dan rasio [0.5, 1, 2] menghasilkan 5 level yang berbeda. Pemilihan skala anchor seperti 32, 64, 128, 256, dan 512 berguna dalam menangani berbagai ukuran objek. Misalnya, skala yang lebih kecil seperti 32 cocok untuk mendeteksi objek berukuran kecil dalam gambar dengan resolusi tinggi, sedangkan skala yang lebih besar seperti 128 berguna untuk mendeteksi objek berukuran besar atau dalam lingkungan dengan resolusi rendah. Penyesuaian ini memungkinkan model untuk beradaptasi dengan variasi ukuran dan bentuk objek dalam dataset yang berbeda. Pada level 0 menghasilkan jumlah anchors tertinggi, yaitu 12,288. Pada level 1 menghasilkan 3,072 anchors, level 2 menghasilkan 768, level 3 menghasilkan 192 dan level 4 yang hanya menghasilkan 48 yang biasanya terjadi pada level dengan resolusi paling rendah. Distribusi anchors di level 0 yang dominan sering kali menunjukkan adanya bias terhadap ukuran objek tertentu dalam

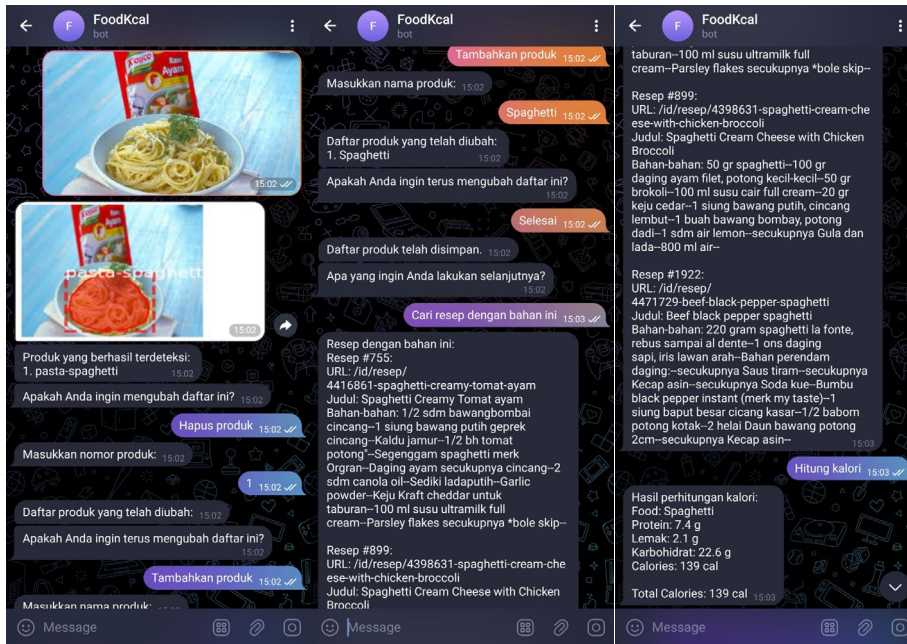
model deteksi objek. Bias ini muncul karena anchor boxes pada level ini biasanya memiliki ukuran dan skala yang tetap atau predefinisi yang tidak cukup fleksibel untuk menyesuaikan dengan berbagai variasi ukuran dan bentuk objek dalam citra. Tetapi dengan metode seperti *dynamic anchor assignment* dan *probabilistic anchor strategies*, model dapat lebih fleksibel dan mengurangi bias tersebut, yang pada akhirnya meningkatkan akurasi deteksi dan segmentasi objek pada berbagai ukuran dan kondisi gambar. Di level 0, jumlah anchor yang dominan adalah 12,288, sedangkan pada level berikutnya jumlahnya berkurang menjadi 3,072, 768, 192, dan hanya 48 pada level dengan resolusi paling rendah. Dengan cara ini, model dapat belajar dari distribusi data dan mengurangi kesulitan dalam mendeteksi berbagai ukuran dan bentuk objek. Anchors yang dihasilkan oleh model ditampilkan pada gambar yang dipilih secara acak. Hal ini dapat membantu dalam memahami bagaimana model mengenali berbagai hal dan memberikan pandangan yang jelas mengenai bagaimana anchors didistribusikan ke seluruh gambar.

Selanjutnya batch data dari dataset divisualisasikan menggunakan ROIs serta anchor yang terdeteksi. Dengan menggunakan generator data, model dapat mengambil gambar dan informasi terkait lainnya untuk analisis lebih lanjut. ROIs yang terdeteksi akan ditampilkan dalam bentuk positif dan negatif sebanyak 66 dan 134 sedangkan untuk positif rasionya mencapai 0.33, juga dari 200 ROIs yang dihasilkan semuanya bersifat unik Dimana seluruh data yang dibuat tidak memiliki duplikasi. Langkah selanjutnya yaitu membuat model menggunakan algoritma Mask R-CNN dengan cara menggabungkan semua proses sebelumnya menjadi satu serta melatih model untuk dapat melakukan prediksi pada gambar baru lalu membandingkan hasil *ground truth* dengan prediksi model pada dataset validasi. Hal ini bertujuan agar model hasil prediksi pada *ground truth* dengan model pada dataset validasi memiliki dimensi yang sama. Tahap berikutnya yaitu pengujian terhadap gambar yang telah dipilih secara acak dengan berbagai tahapan seperti melakukan serangkaian langkah untuk memuat gambar, menyesuaikan ukurannya, dan menjalankan inferensi menggunakan model Mask R-CNN seperti tahap sebelumnya.



Gambar 6. Pengujian sampel gambar

Pada gambar 6, dijelaskan bahwa terdapat beberapa informasi seperti jenis atau nama makanan serta rasio probabilitas hasil dari prediksi tersebut. Tertulis bahwa mesin masih sedikit mengalami kesulitan dalam membedakan telur dan juga tomat namun mesin sudah dapat membedakan sayur dengan bahan makanan yang lainnya. Hal ini terjadi dikarenakan telur dan tomat memiliki sedikit kesamaan pada bentuk serta warna sehingga dataset yang lebih banyak serta pencahayaan pada foto memiliki unsur penting untuk melatih mesin dapat membedakan makanan lebih akurat. Sedangkan Langkah terakhir yaitu menyiapkan bot telegram untuk pengujian serta implementasi makanan. Pada mulanya bot telegram dibuat dengan menggunakan salah satu layanan utama pada telegram yang bernama "BotFather". Layanan ini memungkinkan untuk membuat serta mengelola bot secara pribadi tanpa harus membuatnya secara manual. Setelah bot terbuat maka Langkah selanjutnya yaitu menyambungkan API bot pada model mesin yang telah dibuat lalu membuat fungsi perintah pada bot sehingga bot akan menjadi lebih interaktif ketika dijalankan. Salah satu contohnya seperti menambahkan fungsi untuk menambah produk dan menghilangkan produk dimana fitur ini dibuat untuk mengkoreksi mesin secara manual sehingga jika terdapat beberapa kesalahan seperti pengujian sebelumnya maka mesin dapat membuat hasil estimasi kalori menjadi jauh lebih akurat.



(a) (b) (c)

Gambar. 7. (a) pengenalan gambar makanan (b) fungsi interaktif sistem bot (c) Estimasi kalori makanan

Ketika bot sudah dapat mengenali makanan dan mengestimasi jumlah kalorinya, maka tahap terakhir yaitu pengujian akurasi dengan menggunakan gambar berupa 13 jenis makanan dan minuman secara acak berdasarkan pada klasifikasi seperti gambar 4. Dari ke 13 jenis yang sudah dilakukan pengujian maka dibuatlah tabel uji data mengenai makanan dan minuman dengan parameter total sampling yaitu seberapa banyak total jumlah makanan pada satu gambar lalu ada parameter benar, salah, dan tidak terdeteksi sebagai bagian untuk menguji seberapa bagus mesin bekerja dalam menangkap gambar, dilanjut dengan tingkat akurasi dimana rumus pengujian memiliki rumus seperti yang sudah dijelaskan pada bab sebelumnya. kalori yang ada.

TABEL II
 HASIL

Dataset	Total sampling	Benar	Salah	Tidak Terdeteksi	Tingkat Akurasi	Kalori (kkal)
Apel	5	5			100%	58.0
Brokoli	6	5	1		83%	0
Buncis	5	3	2		60%	35.0
Keju	5	3	2		60%	326.0
Kentang	6	4	2		67%	83.0
Kopi	5	4	1		80%	352.0
Nasi	6	4	1	1	67%	178.0
Pisang Mas	6	6			100%	127.0
Roti	5	5			100%	248.0
Spaghetti	7	4	2	1	57%	139.0
Teh	5	3	1	1	60%	132.0
Telur Ayam	5	4	1		80%	162.0
Wortel	5	5			100%	42.0
Total Akurasi					78%	

Pada tabel 2, sistem yang dikembangkan menggunakan algoritma Mask R-CNN menunjukkan tingkat akurasi keseluruhan sebesar 78% pada deteksi makanan dari 13 jenis makanan yang diuji (seperti apel, brokoli, kopi, dan roti). Hasil tersebut menunjukkan bahwa model ini memiliki tingkat akurasi yang baik dalam mendeteksi sebagian besar objek makanan yang diuji, mencapai akurasi 100% untuk beberapa kelas makanan seperti apel, pisang mas, roti, dan wortel meskipun hasil ini relevan dengan tujuan penelitian untuk menyediakan akses cepat terhadap informasi kalori, masih ada ruang untuk perbaikan dalam akurasi secara keseluruhan. Meskipun akurasi keseluruhan mencapai 78%, penting untuk mengevaluasi kontribusi setiap parameter mask R-CNN secara detail. Jumlah data pelatihan memiliki pengaruh signifikan terhadap performa, di mana dataset yang lebih besar dapat meningkatkan generalisasi model. Konfigurasi parameter, seperti threshold untuk ROI, berperan dalam menentukan keakuratan

area yang dianalisis. Di sisi lain, metode augmentasi kumpulan data seperti rotasi, pencahayaan, dan penskalaan dapat membantu meningkatkan keragaman data dan mengurangi overfitting, yang pada akhirnya berdampak positif pada akurasi model.

Penelitian ini menunjukkan konsistensi dengan penelitian lain yang menggunakan Mask R-CNN untuk segmentasi objek. Mask R-CNN memiliki keunggulan dibandingkan model lain seperti YOLO dan Faster R-CNN dalam mendeteksi makanan dengan fitur visual yang kompleks. Keunggulannya adalah memungkinkan segmentasi objek yang lebih detail pada tingkat piksel, sehingga model dapat membedakan bagian makanan yang memiliki variasi bentuk dan tekstur yang tumpang tindih atau kompleks. YOLO dan Faster R-CNN memprioritaskan deteksi kotak pembatas yang lebih cepat, sedangkan Mask R-CNN memberikan akurasi yang lebih tinggi dengan memahami fitur visual spesifik setiap objek dalam gambar makanan. Contoh penelitian Mask R-CNN yang sudah disinggung sebelumnya tentang deteksi objek sampah menggunakan mask scoring RCNN yang dilakukan oleh Shuijing Li dan Ming Yan [19] menunjukkan bahwa Mask Scoring RCNN diterapkan pada klasifikasi objek sampah, mencapai akurasi 65,8%. Meskipun aplikasi berbeda, pendekatan serupa menggunakan Mask R-CNN menunjukkan keunggulan dalam segmentasi objek yang kompleks. Hal ini mengindikasikan Mask R-CNN dapat diadaptasi ke berbagai domain, tetapi akurasi dipengaruhi oleh kualitas dataset dan karakteristik objek. Sedangkan penelitian Dharma Udayana [10] yang menggunakan CNN pada dataset gambar makanan, menghasilkan akurasi antara 66%-98%. Akurasi yang lebih rendah terjadi karena kemiripan fitur visual antar gambar makanan. Hasil ini sejalan dengan temuan penelitian ini, di mana makanan dengan fitur visual serupa seperti telur dan tomat seringkali salah diklasifikasikan. Oleh karena itu, hasil penelitian ini mendukung temuan sebelumnya dan menunjukkan bahwa Mask R-CNN efektif untuk segmentasi objek, tetapi akurasinya lebih rendah dibandingkan kumpulan data, terutama untuk objek yang sangat mirip secara visual. Oleh sebab itu, Model dapat diimprovisasi menggunakan semi-supervised learning atau transfer learning yang memanfaatkan kumpulan data pangan global yang besar. Pendekatan ini memungkinkan model untuk belajar dari data yang lebih beragam, mengurangi risiko overfitting sekaligus meningkatkan kemampuannya untuk mengenali berbagai jenis makanan dengan akurasi dan generalisasi yang lebih tinggi.

Selama pengujian juga terdapat anomali atau kelainan pada beberapa makanan. Kelainan tersebut terjadi pada makanan yang bentuknya mirip, seperti telur dan tomat, namun sering salah klasifikasi karena kesamaan bentuk dan warna. Kesulitan serupa juga terjadi ketika membedakan minuman berwarna gelap seperti kopi. Lalu terdapat kelainan data lain pada jenis makanan brokoli yang tidak dapat menampilkan jumlah kalori, Analisis mengungkapkan bahwa hal ini mungkin disebabkan oleh kurangnya data yang representatif, sehingga menyulitkan model untuk memahami sifat visual unik brokoli. Selain itu, anomali dapat terjadi karena terdapat perbedaan yang signifikan pada properti gambar kedua set data, hal ini mungkin juga disebabkan oleh ketidakcocokan antara set data pelatihan dan validasi. Faktor-faktor ini menunjukkan bahwa pencahayaan dan variasi kumpulan data serta kesesuaian dataset mempengaruhi akurasi deteksi.

Keterbatasan penelitian ini juga tidak lepas dari terbatasnya cakupan kumpulan data, yang mungkin berdampak pada kemampuan model untuk memahami variasi makanan yang lebih kompleks. Selain itu, mengingat keterbatasan Mask R-CNN dalam membedakan objek dengan sifat visual serupa, bias mungkin terjadi dalam memilih sifat visual dari sampel dan makanan yang sangat mirip. Untuk mengatasi keterbatasan ini, penelitian di masa depan dapat mempertimbangkan penggunaan dataset yang lebih beragam, penerapan metode augmentasi yang lebih kompleks guna meningkatkan kemampuan model dalam menggeneralisasi pola dari berbagai karakteristik makanan.

IV. PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dikembangkan dengan menggunakan metode yang digunakan, beberapa dapat ditarik hasil dan kesimpulan bahwa metode *Mask Region Convolutional Neural Network* (Mask R-CNN) dirancang untuk membantu user mengatur kebiasaan makan mereka dengan menyediakan akses cepat dan mudah terhadap informasi tentang jumlah kalori. Sistem ini telah berhasil mendeteksi makanan dan jumlah kalorinya dengan menggunakan beberapa tahapan seperti pembuatan model, training data, implementasi terhadap gambar menggunakan algoritma, hingga pembuatan bot pada telegram. Setelah rancangan penelitian selesai maka dilakukanlah pengujian dengan menggunakan 10 jenis sampel gambar makanan dan minuman secara acak untuk melihat hasil dari keakuratan mesin. Menggunakan dataset dari AICrowd Food Recognition Challenge, model ini

dilatih hingga mampu mendeteksi berbagai jenis makanan dengan tingkat akurasi sebesar 78%. Meskipun begitu, pengujian juga menunjukkan bahwa terkadang model mungkin tidak mengenali kelas produk yang sudah dipelajari dalam foto. Model ini sering membingungkan objek yang serupa, seperti minuman gelap dengan kopi, hingga sayuran merah terang dengan tomat. Begitu pun dengan ketidaksesuaian antara dataset untuk melatih model dengan dataset yang dipakai untuk mengestimasi kalori sehingga jumlah kalori pada makanan tersebut tidak keluar.

Ini menunjukkan bahwa useran dataset yang lebih bervariasi dan pencahayaan optimal pada gambar dapat meningkatkan akurasi sistem. Penelitian ini membuka peluang untuk pengembangan lebih lanjut. ini termasuk peningkatan akurasi deteksi makanan dan kalori dengan menambah variasi dataset serta mengoptimalkan parameter model. Dengan begitu, sistem yang dibuat ini diharapkan bisa membantu orang untuk melacak pola makan dengan lebih mudah dan akurat. Sistem ini juga bisa dikembangkan lagi untuk berbagai keperluan di bidang kesehatan dan juga nutrisi.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih sebanyak-banyaknya kepada Tim *Jurnal ilmiah penelitian dan pembelajaran informatika (JIFI)* yang telah membuat dan Menyusun template yang saya gunakan sebagai acuan untuk Menyusun jurnal. Saya juga ingin berterima kasih kepada Bapak Pulung Nurtantio Andono selaku pembimbing saya, atas bimbingan serta dukungannya yang sangat berperan selama pengembangan dan penyusunan jurnal ini. Saran dan motivasinya sangat membantu saya dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] S. Bachri and M. Mulyati, "Pola Hidup Sehat Masyarakat Di Era Revolusi Industri 4.0," *J. Pengabd. Teratai*, vol. 2, no. 2, pp. 79–84, 2021, doi: 10.55122/teratai.v2i2.243.
- [2] P. Sofia Rincón-Gallardo *et al.*, "Effects of menu labeling policies on transnational restaurant chains to promote a healthy diet: A scoping review to inform policy and research," *Nutrients*, vol. 12, no. 6, pp. 1–27, 2020, doi: 10.3390/nu12061544.
- [3] M. E. Putri, M. Khairi, M. Furqan, and B. Yusman, "Jurnal Kecerdasan Buatan , Komputasi dan Teknologi Informasi Deteksi Objek Untuk Menghitung Perkiraan Kalori Makanan Menggunakan Metode R-CNN Mask Berbasis Web," vol. 5, no. 1, pp. 84–92, 2024.
- [4] M. Gao, G. Zou, Y. Li, and X. Guo, "Recent Advances in Computer Vision: Technologies and Applications," *Electronics*, vol. 13, no. 14, 2024, doi: 10.3390/electronics13142734.
- [5] I. Prihatini and R. Kumala dewi, "Kandungan Enzim Papain pada Pepaya (Carica papaya L) Terhadap Metabolisme Tubuh," *J. Tadris IPA Indones.*, vol. 1, pp. 449–458, Nov. 2021, doi: 10.21154/jtii.v1i3.312.
- [6] M. Ummah, A. Sasmito, R. T. Wahyuningrum, and M. I. Mustajib, "Implementasi Mask R-CNN untuk Identifikasi Cacat pada Pengelasan Shielded Metal Arc Welding (SMAW)," 2024.
- [7] M. M. Rana, T. Akter Tithy, and M. M. Hasan, "Vehicle Detection and Count in the Captured Stream Video Using Opencv in Machine Learning," *Comput. Sci. Eng. An Int. J.*, vol. 12, no. 3, pp. 1–9, 2022, doi: 10.5121/cseij.2022.12301.
- [8] N. S. Ketkar and J. T. Moolayil, "Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch," *Deep Learn. with Python*, 2021, [Online]. Available: <https://api.semanticscholar.org/CorpusID:233187730>.
- [9] A. Kholik, "Klasifikasi Menggunakan Convolutional Neural Network (Cnn) Pada Tangkapan Layar Halaman Instagram," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, p. 10, 2021, doi: 10.33365/jdmsi.v2i2.1345.
- [10] I. P. A. E. Darma Udayana and P. G. S. C. Nugraha, "Prediksi Citra Makanan Menggunakan Convolutional Neural Network Untuk Menentukan Besaran Kalori Makanan," *J. Teknol. Inf. dan Komput.*, vol. 6, no. 1, pp. 30–38, 2020, doi: 10.36002/jutik.v6i1.1001.
- [11] S. Muhammad Rizqi Zamzami, Dahnia and H. Fitriyah, "Sistem Identifikasi Jenis Makanan dan Perhitungan Kalori berdasarkan Warna HSV dan Sensor Loadcell menggunakan Metode K-NN berbasis," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 936–942, 2021.
- [12] Q. Wu *et al.*, "Improved Mask R-CNN for Aircraft Detection in Remote Sensing Images," *Sensors*, vol. 21, no. 8, 2021, doi: 10.3390/s21082618.
- [13] R. Sapkota, D. Ahmed, and M. Karkee, "Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments," *Dec. 2023*, doi: 10.1016/j.aiaa.2024.07.001.
- [14] L. Jian, Z. Pu, L. Zhu, T. Yao, and X. Liang, "SS R-CNN: Self-Supervised Learning Improving Mask R-CNN for Ship Detection in Remote Sensing Images," *Remote Sens.*, vol. 14, no. 17, 2022, doi: 10.3390/rs14174383.
- [15] R. A. Tilasefana and R. E. Putra, "Penerapan Metode Deep Learning Menggunakan Algoritma CNN Dengan Arsitektur VGG NET Untuk Pengenalan Cuaca," *J. Informatics Comput. Sci.*, vol. 05, no. 1, pp. 48–57, 2023.
- [16] J. Nurhakiki *et al.*, "Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya," *J. Pendidik. Berkarakter*, no. 1, pp. 270–281, 2024, [Online]. Available: <https://doi.org/10.51903/pendekar.v2i1.598>.
- [17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, pp. 386–397, 2020, [Online]. Available: <https://api.semanticscholar.org/CorpusID:264031695>.
- [18] J. G. C. Ball *et al.*, "Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using Mask R-CNN," *bioRxiv*, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:250533961>.
- [19] S. Li, M. Yan, and J. Xu, "Garbage object recognition and classification based on Mask Scoring RCNN," *2020 Int. Conf. Cult. Sci. & Technol.*, pp. 54–58, 2020, [Online]. Available: <https://api.semanticscholar.org/CorpusID:227221748>.
- [20] K. Jozwik, T. Kietzmann, R. Cichy, N. Kriegeskorte, and M. Mur, "Deep Neural Networks and Visuo-Semantic Models Explain Complementary Components of Human Ventral-Stream Representational Dynamics," *J. Neurosci.*, vol. 43, Feb. 2023, doi: 10.1523/JNEUROSCI.1424-22.2022.
- [21] AICrowd, "Food Recognition Challenge," 2023. <https://www.aicrowd.com/challenges/food-recognition-challenge>.