

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terdahulu

Beberapa penelitian terkait pengaruh kualitas *dataset* terhadap akurasi dari model deteksi objek, SSIM, dan metode yang digunakan antara lain:

##### 2.1.1 Pengaruh Bias Dataset Terhadap Performa Akurasi Deteksi Objek Pada Arsitektur Efficientdet-Lite [3]

Penelitian dengan judul “Pengaruh Bias Dataset Terhadap Performa Akurasi Deteksi Objek Pada Arsitektur Efficientdet-Lite” ini dilakukan oleh Ridho Imantiyar. Penelitian ini membahas mengenai pengaruh dari bias yang terdapat pada *dataset* yang akan digunakan untuk model deteksi objek dengan metode *Efficient-Lite* dengan menggunakan objek berupa produk retail yang tergolong dalam produk dalam kategori *Ready to Drink* (RTD). Dalam penelitiannya, penulis menggunakan 4 buah *dataset* dengan kondisi masing-masing di mana *dataset* pertama merupakan *dataset* yang diambil secara sempurna tanpa adanya pengaruh dari bias apapun. Kemudian diambil *dataset* dengan menerapkan bias arah datang cahaya sehingga terdapat pantulan cahaya pada objek yang dilakukan secara manual melalui angle foto pengambilan *dataset*. *Dataset* yang ketiga diambil dengan cara mengambil foto kaleng minuman dari angle yang miring tanpa adanya pengaruh cahaya. Dan kemudian *dataset* yang terakhir adalah kondisi *dataset* yang diberikan label. Setelah *dataset* dikumpulkan, selanjutnya masing-masing *dataset* tersebut melalui proses training model untuk menghasilkan model objek deteksi dengan akurasi yang berbeda-beda. Seluruh proses pengambilan gambar *dataset* dalam penelitian ini dilakukan secara manual. Dan proses pengujian pengaruh kualitas *dataset* terhadap model deteksi objek dilakukan tanpa menggunakan metrik pengukuran.

Kesimpulan dari penelitian ini adalah model dengan *dataset* terbaik (tanpa bias) menghasilkan AP sebesar 0,6211. *dataset* dengan kondisi pengaruh bias cahaya memiliki AP sebesar 0,3918. *dataset* dengan kondisi bias sudut pengambilan gambar memiliki AP sebesar 0,5215. dan *dataset* yang diberi label memiliki AP sebesar 0,122.

Penelitian pada jurnal ini menjadi acuan pertama penulis dalam melakukan penelitian karena terdapat kemiripan tujuan penelitian walaupun memiliki aspek penilaian yang berbeda.

### **2.1.2 SSIM Image Quality Metric for Denoised Image [7]**

Jurnal penelitian dengan judul “SSIM Image Quality Metric for Denoised Image” ini dilakukan oleh Ndajah, Hisakazu, et al. Penelitian pada jurnal ini berfokus pada penggunaan metrik SSIM dalam mengukur kualitas gambar dengan membandingkannya dengan MSE. di mana hasil dari penelitian ini menunjukkan jika nilai MSE selalu tetap walaupun gambar didistorsi hingga jauh dari kualitas awalnya. Sedangkan dengan menggunakan SSIM nilai metriknya semakin kecil saat gambar semakin didistorsi. Distorsi yang dilakukan antara lain *Contrast Stretch*, *Negative Image*, *Gaussian White Noise*, *Lossy Compression*, dan *Blurred Image*.

Menurut penelitian ini, hal tersebut bisa terjadi karena SSIM berkorelasi kuat dengan *Human Vision System* (HVS) di mana pencahayaan, kontras, dan struktur gambar menjadi parameter utama pengukuran metrik SSIM. Berbeda dengan metrik berbasis MSE yang tidak dirancang untuk menangkap distorsi struktural yang mungkin terjadi. Penelitian ini berkontribusi untuk membuat penulis memilih perhitungan kualitas gambar dengan metrik SSIM dibanding metrik yang berbasis MSE.

### **2.1.3 Tinjauan Literatur Sistematis tentang Structural Similarity Index Measure untuk Deteksi Anomali Gambar [8]**

Jurnal penelitian dengan judul “Tinjauan Literatur Sistematis tentang Structural Similarity Index Measure untuk Deteksi Anomali

Gambar” ini dilakukan oleh Halim Bayuaji Sumarna , Ema Utami, dan Anggit Dwi Hartanto. Pada penelitian ini dijelaskan lebih lanjut lagi mengenai cara kerja dari algoritma SSIM yang mengukur parameter-parameter seperti *Luminance*, *Contrast*, dan *Structure*. Penelitian ini juga menjelaskan berbagai penelitian-penelitian terdahulu yang menggunakan SSIM seperti penggunaan metrik SSIM dalam menemukan ekspresi wajah yang paling mirip dengan ekspresi yang diberikan dengan memanfaatkan aspek struktural visual dari ekspresi dan beberapa penelitian lainnya berikut dengan hasilnya sehingga membuat manfaat penggunaan SSIM dalam pengukuran kualitas gambar semakin jelas. Jurnal ini berkontribusi terhadap pemahaman penulis secara *high-level* terhadap pengukuran metrik SSIM dalam penentuan kualitas gambar.

#### **2.1.4 Object Detection Techniques: A Comparison [9]**

Jurnal penelitian dengan judul “Object Detection Techniques: A Comparison” ini dilakukan oleh Priyanka Malhotra dan Ekansh Garg. Penelitian ini berfokus pada pengujian R-CNN, Fast R-CNN, dan YOLO sebagai model deteksi objek. Berdasarkan penelitian tersebut disimpulkan bahwa model YOLO memiliki kecepatan yang paling cepat dibandingkan R-CNN dan Fast R-CNN. Namun, model YOLO memiliki kelemahan tidak bisa mendeteksi objek berukuran kecil pada gambar. Berbanding terbalik dengan model YOLO, model keluarga R-CNN memiliki kemampuan untuk mendeteksi objek yang berukuran kecil pada gambar namun dengan kecepatan yang lebih lambat dibanding YOLO. Pemilihan model yang ingin digunakan tergantung dari *dataset* serta tujuan dibuatnya model tersebut. Apabila model yang dibuat diharapkan bisa memiliki kemampuan deteksi yang sangat detail, maka model keluarga R-CNN akan menjadi model yang lebih baik.

Melalui penelitian jurnal penelitian ini, penulis memutuskan untuk menggunakan model YOLO dikarenakan peneliti berharap model yang dibuat bisa mendeteksi objek secara cepat dengan beban kerja yang

ringan secara *real-time*. Selain itu, penelitian ini tidak membutuhkan model dengan akurasi yang sangat tinggi karena hanya berfokus pada dampak kualitas *dataset* terhadap akurasi model.

### 2.1.5 A Study on Object Detection [10]

Penelitian dengan judul “A Study on Object Detection” yang dilakukan oleh S.Manjula dan Dr.K.Lakshmi membahas secara *high-level* mengenai bagaimana cara deteksi objek bekerja dengan objek penangkapan gambar menggunakan *Surveillance System*. Menurut penelitian ini, deteksi objek bekerja melalui beberapa tahapan seperti *Environment Modelling* yang bertujuan untuk melakukan pemetaan terhadap gambar yang akan dideteksi, *Motion Segmentation* yaitu tahapan untuk mendeteksi area yang terdapat objek bergerak, dan *Object Classification* yang bertujuan untuk melakukan pembedaan *region* seperti *region* milik manusia dengan *region* milik benda bergerak lainnya. Penelitian ini berkontribusi untuk memberikan penulis gambaran secara *High-level* bagaimana cara kerja deteksi objek.

## 2.2 Tinjauan Teori

Beberapa penjelasan dasar teori terkait penelitian ini diantaranya:

### 2.2.1 Object Detection

*Computer Vision* adalah salah satu bidang kecerdasan buatan (AI) yang membuat mesin/komputer memiliki kemampuan untuk mengambil informasi yang berguna dari gambar digital [11]. *Computer Vision* memiliki cara kerja yang sama seperti penglihatan manusia. Manusia dapat melihat apa yang ada di depannya kemudian bisa melakukan pengolahan informasi terhadap apa yang dilihatnya. *Computer Vision* bertujuan untuk membuat mesin/komputer bisa melakukan hal serupa. Dan salah satu tugas yang bisa dilakukan oleh *Computer Vision* adalah deteksi objek. Deteksi objek bertujuan untuk melakukan pemetaan terhadap objek di dalam gambar (*object localization*) kemudian menentukan kategori setiap objek yang dideteksi (*object classification*) [12].

## 2.2.2 Structural Similarity Index Measure (SSIM)

Dalam membuat model *object detection*, *dataset* adalah salah satu faktor penentu yang membuat tingkat akurasi dari model yang dibuat [3]. Terdapat beberapa cara yang dapat diterapkan untuk melakukan pengukuran terhadap kualitas dari suatu gambar, salah satunya adalah dengan melakukan pengukuran terhadap metrik structural similarity index measure (SSIM) gambar. metrik ini umum digunakan untuk menilai kualitas gambar, dan cara kerjanya adalah dengan membandingkan fitur struktural dari gambar, dan kualitas gambar dijelaskan oleh kesamaan struktural [8].

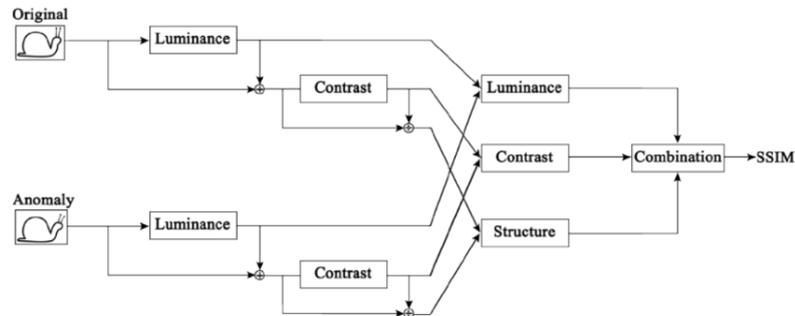
Gambar referensi dibutuhkan dalam melakukan pengukuran dengan menggunakan metrik SSIM. Kemudian, gambar uji dan gambar referensi akan dibandingkan guna mendapat informasi berupa kesamaan struktur kedua gambar. Semakin tinggi kesamaan antara gambar uji dengan referensi, maka kualitas gambar akan semakin tinggi, sedangkan jika kesamaan semakin rendah, maka kualitas gambar uji tersebut juga semakin rendah. Selain digunakan untuk menilai kualitas suatu gambar, SSIM dapat digunakan sebagai index pengukuran dalam menganalisis perbedaan gambar, sehingga kita dapat mengetahui perbedaan dari perbandingan dua gambar tersebut berdasarkan data struktural dari sebuah gambar [8].

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

**Gambar 2.1** Rumus Perhitungan SSIM

Rumus pada gambar 2.1 merupakan rumus perhitungan yang digunakan untuk menghitung nilai SSIM pada suatu gambar melalui perbandingan dengan gambar referensi. Secara sederhananya  $\mu$  (Mu) merupakan nilai *Luminance* suatu gambar dan  $\sigma$  (sigma) merupakan nilai

*contrast* suatu gambar.  $C1$  dan  $C2$  adalah sebuah konstanta yang menjamin kestabilan apabila pembagi bernilai 0.



**Gambar 2.2** Cara Kerja SSIM Secara Sederhana

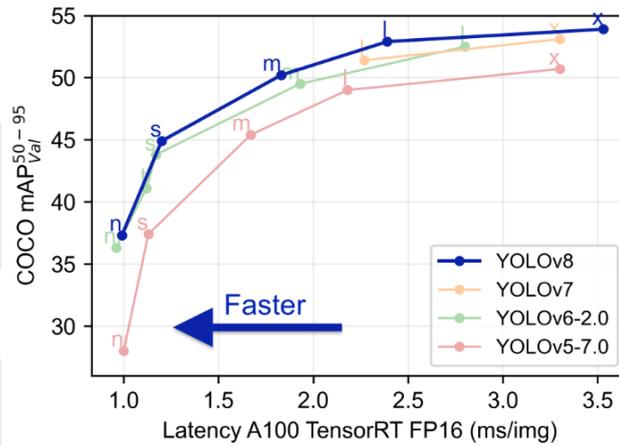
Berdasarkan gambar 2.2, dapat disimpulkan jika SSIM membagi luminasi (seberapa banyak *white & black* dalam gambar), kontras (intensitas bayangan dalam gambar) dan struktur dari dua buah gambar kemudian membandingkannya untuk mendapat nilai kesamaan dari kombinasi kedua gambar tersebut. Faktor yang membedakan pengukuran kualitas gambar dengan menggunakan metrik SSIM dengan metrik berbasis MSE adalah SSIM merupakan metrik yang berkorelasi kuat dengan *Human Vision Structure (HVS)* sedangkan MSE tidak dirancang untuk menangkap distorsi struktural yang mungkin terjadi. Sehingga beberapa penelitian berpendapat jika pengukuran kualitas gambar dengan metrik SSIM lebih baik dibanding dengan menggunakan metrik berbasis MSE [6][7].

### 2.2.3 YOLO

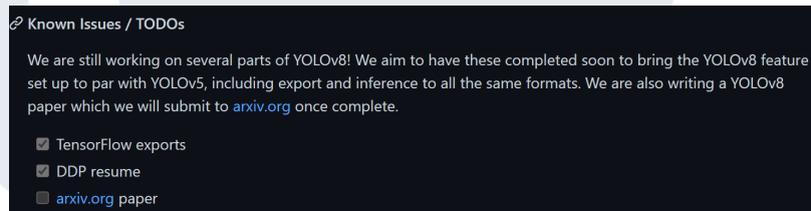
*Deep learning* merupakan metode dalam kecerdasan buatan (AI) yang memiliki tujuan untuk melakukan pengajaran terhadap komputer untuk melakukan pemrosesan data dengan cara yang terinspirasi oleh otak manusia [13]. Model *deep learning* dapat melakukan tugas seperti pengenalan terhadap pola kompleks dalam suatu gambar, teks, suara, dan data lainnya dengan tujuan untuk menghasilkan wawasan berupa data dan prediksi yang akurat terhadap objek yang dideteksi.

*Deep learning neural networks*, merupakan suatu upaya untuk melakukan peniruan terhadap kinerja otak manusia melalui kombinasi dari serangkaian informasi. Informasi-informasi tersebut akan diolah untuk melakukan pengenalan, klasifikasi serta pendeskripsian yang terdapat dalam suatu data [14]. *Convolutional neural networks (CNN)*, merupakan salah satu tipe neural network yang utamanya digunakan dalam *Computer Vision* dan *Image Classification* seperti yang dilakukan dalam model deteksi objek.

Salah satu model yang menggunakan CNN adalah *You Only Look Once* atau yang lebih akrab disebut YOLO. YOLO merupakan model *Single-shot object detection* yaitu jenis model objek deteksi yang hanya menggunakan *input image* untuk melakukan deteksi sekaligus klasifikasi objek dalam gambar tanpa memerlukan *Region Proposal*. Dalam penelitian ini YOLOv8 dipilih untuk digunakan sebagai model yang akan menerima training *dataset* dikarenakan YOLOv8 merupakan versi terbaru dari model YOLO yang dirilis pada 10 Januari 2023 dan memiliki akurasi tertinggi pada keluarga YOLO. Selain itu, dipilihnya YOLOv8 dibanding model lain seperti Faster RCNN adalah karena model YOLO memiliki kecenderungan kecepatan yang tinggi namun akurasi yang lebih rendah dibanding model lainnya. Sehingga dibutuhkan lebih banyak *dataset* dalam proses pengembangannya. Melalui penelitian ini penulis memilih model YOLO dengan harapan dapat membantu pengembangan objek deteksi model YOLO dengan memberikan gambaran atau ide untuk mengurangi mahalnya proses *labeling* (waktu dan biaya) dari penggunaan *dataset* yang banyak. Namun, hingga saat penelitian ini ditulis, masih belum terdapat paper resmi mengenai performa dari YOLOv8 dibandingkan dengan versi pendahulunya. Penulis hanya mendapatkan informasi menurut halaman github resmi dari YOLOv8. Performa dari YOLOv8 dibanding dengan versi terdahulunya dapat dilihat pada gambar di bawah [15]:



Gambar 2.3 Perbandingan YOLOv8 dengan versi sebelumnya



Gambar 2.4 Belum dirilisnya paper YOLOv8

## 2.2.4 Dataset

*Dataset* merupakan sekumpulan data yang akan digunakan dalam pelatihan model Machine Learning. Dataset dapat berupa gambar, suara, signal, ataupun tulisan tergantung dari tujuan pembuatan model. *Dataset* bisa didapat melalui proses pengumpulan secara mandiri maupun melalui pengambilan melalui sumber-sumber yang menyediakan *dataset* untuk keperluan pembelajaran ataupun penelitian.

Dalam penelitian ini, penulis akan menggunakan *dataset* berupa objek-objek yang bisa diklasifikasikan sebagai sampah. Akan terdapat enam kelas dalam penelitian yang dilakukan dan sumber dari dataset akan diambil melalui website Kaggle.

### 2.2.5 Dataset Modification

Pada penelitian ini, modifikasi *dataset* dilakukan melalui cara yang sama seperti proses augmentasi data. di mana augmentasi data adalah sebuah teknik untuk memperbanyak jumlah *dataset* dengan membuat duplikat data dengan kondisi yang sudah dimodifikasi [16]. Tujuan utama dari data augmentation adalah untuk membuat model memiliki lebih banyak referensi *dataset* sehingga berpotensi untuk menaikkan hasil akurasi dari model tersebut. Beberapa hal yang bisa dilakukan data augmentation pada *dataset* berupa gambar adalah: [16]

- 1) ***Geometric transformations***, proses augmentasi untuk memutar, memotong, menarik, atau memperbesar gambar secara acak.
- 2) ***Color space transformations***, mengubah nilai RGB suatu gambar secara acak.
- 3) ***Kernel Filters***, mengubah ketajaman serta kejernihan suatu gambar secara acak.
- 4) ***Random erasing***, menghapus bagian gambar secara acak.
- 5) ***Mixing images***, mengkombinasikan dua buah gambar.

Namun dalam penelitian ini, istilah ‘modifikasi data’ digunakan karena tujuan utama dilakukannya augmentasi data adalah untuk menurunkan kualitas *dataset* tanpa memperbanyak jumlah *dataset* sehingga bisa didapatkan hasil akurasi yang berbeda dengan model yang dilatih menggunakan *dataset* utama. Proses modifikasi data yang akan digunakan adalah dengan menambahkan *noise* serta mengubahnya kedalam *grayscale* berdasarkan gambar pada *dataset* utama. Pemilihan perubahan nilai-nilai tersebut didasari oleh karakteristik dari metrik SSIM yang sangat bergantung terhadap kondisi struktural suatu gambar dikarenakan SSIM memiliki kriteria yang sama seperti sistem visual manusia dalam menilai suatu gambar [7]. Dan kondisi *noise* pada suatu gambar merupakan nilai-nilai yang dapat dinilai berdasarkan persepsi manusia.

### 2.2.6 Roboflow

Roboflow merupakan sebuah platform yang bisa diakses melalui *website* secara daring untuk melakukan pengolahan *dataset* yang akan digunakan dalam proses pelatihan model pembelajaran mesin. Dengan menggunakan Roboflow kita bisa melakukan anotasi, *pre-processing*, Augmentasi gambar dan beberapa hal lainnya yang berkaitan dengan pengolahan *dataset* [17]. Dalam penelitian ini, penulis hanya akan memanfaatkan fitur anotasi dan *preprocessing resize* yang disediakan oleh Roboflow.

