

BAB II

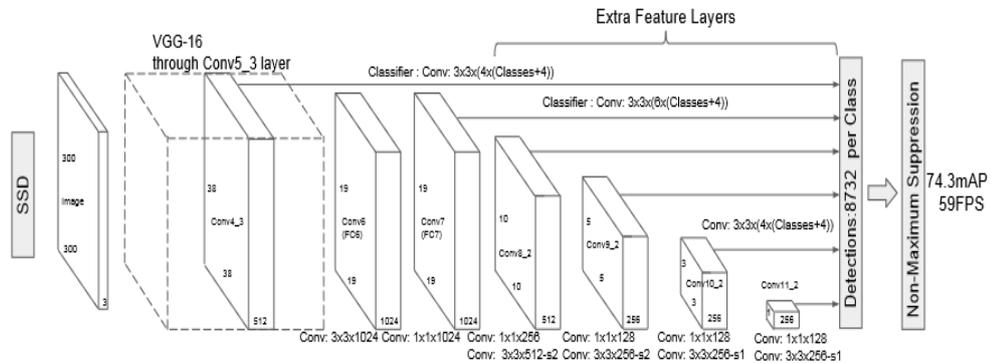
LANDASAN TEORI

Pada penelitian ini akan dibuat dengan landasan teori yang mendukung agar sistem dapat berjalan seperti yang telah direncanakan. Bab ini menjelaskan mengenai teori-teori pendukung yang berkaitan dengan deteksi area parkir menggunakan SSD-MobileNet dengan penjelasan komponen-komponen yang akan digunakan pada penelitian ini. Adapun landasan teori pada penelitian ini adalah sebagai berikut.

1.1 *Single Shot Multibox Detector* (SSD)

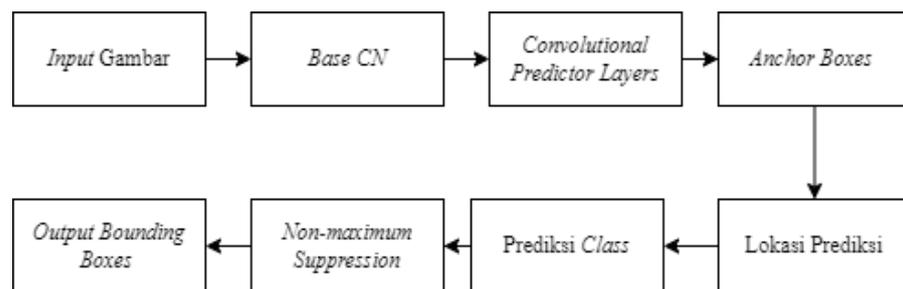
Single Shot Multibox Detector (SSD) adalah salah satu dari teknik *deep learning* yang mendiskritisasi ruang keluaran kotak pembatas ke dalam satu set kotak standar pada berbagai rasio aspek dan skala per titik peta fitur. SSD beroperasi berdasarkan pendekatan jaringan *convolutional feed-forward* [8]. SSD merupakan algoritma *single deep neural network* yang menerapkan fitur *bounding boxes* dalam memperkirakan lokasi objek yang dideteksi. **Gambar 2.1** menunjukkan bahwa SSD mempunyai dua buah arsitektur jaringan VGG-16 dan jaringan arsitektur pada *extra feature layers* yang dapat berupa *inception*, *mobilenet*, atau *Resnet*. Fitur utama dalam SSD antara lain: *Multi-Scale Feature Maps*, *Convolutional Predictor*, dan *Default Box and Aspect Ratio*. SSD mempunyai dua tipe parameter yaitu model parameter dan *hyperparameter*. Parameter merupakan konstanta model yang telah ditentukan pada set data pelatihan dan tidak dapat dirubah. Sedangkan *hyperparameter* merupakan konstanta yang dapat disesuaikan dengan kebutuhan untuk mendapatkan model dengan kinerja yang baik. berbagai

macam *hyperparameter* yaitu: *Learning Rate*, *Number of State training*, *Number of Epoch*, dan lain-lain [8].



Gambar 2.1 Arsitektur SSD [8]

Single Shot Multibox Detector (SSD) merupakan sebuah metode dalam bidang *computer vision* yang digunakan untuk mendeteksi objek di dalam sebuah gambar. Metode ini menggabungkan teknik pelatihan *neural network* dengan penggunaan *anchor boxes* untuk menghasilkan deteksi objek yang cepat dan akurat. Adapun diagram blok metode SSD ditunjukkan oleh **Gambar 2.2**.



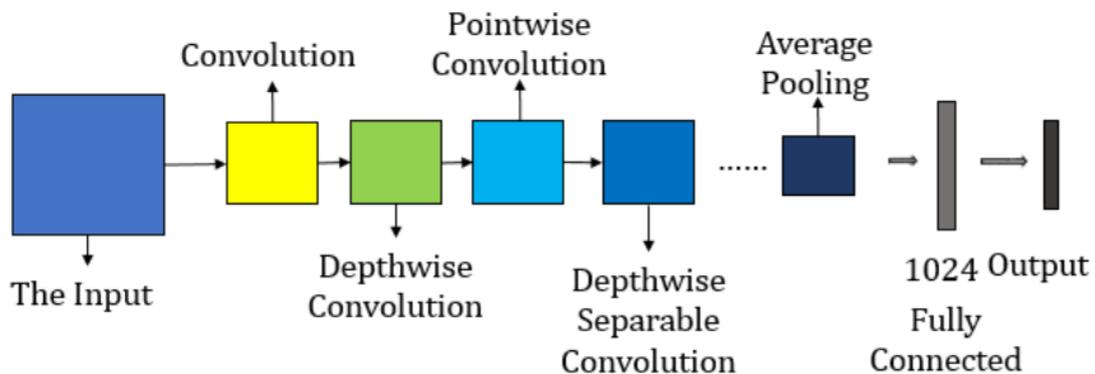
Gambar 2.2 Diagram Blok SSD

Pada **Gambar 2.2** Keseluruhan model SSD beroperasi dengan mengambil gambar masukan, mengekstrak fitur dari gambar tersebut menggunakan jaringan konvolusi, kemudian melakukan prediksi lokalisasi dan kelas untuk setiap *anchor box* pada gambar menggunakan lapisan konvolusi tambahan. Prediksi tersebut

kemudian digunakan untuk menentukan apakah setiap *anchor box* berisi objek dan di lokasi mana objek tersebut berada. Setelah itu, algoritma *non-maximum suppression* digunakan untuk menghilangkan kotak referensi yang tumpang tindih dan memilih kotak dengan skor kelas tertinggi sebagai keluaran akhir. SSD menggunakan menggunakan *anchor boxes* dalam mempercepat proses pelatihan dan meningkatkan akurasi deteksi. *Anchor boxes* merupakan kotak-kotak yang dihasilkan dengan ukuran dan aspek rasio yang bervariasi. *Bonding box* yang dihasilkan pada setiap lokasi dalam gambar kemudian dihitung perbedaannya dengan *anchor box* yang cocok dengan ukuran dan aspek rasio yang sesuai.

1.2 MobileNet

MobileNet merupakan salah satu arsitektur canggih yang dapat secara signifikan mengurangi ukuran model dan hanya memengaruhi akurasi minimal. Dasar dari arsitektur MobileNet adalah dua jenis lapisan konvolusi yang berbeda dari proses konvolusi standar. Lapisan-lapisan ini adalah lapisan konvolusi yang dapat dipisahkan secara mendalam dan lapisan konvolusi yang dapat dipisahkan. Lapisan kedalaman menggunakan filter tunggal untuk setiap saluran input, berbeda dengan konvolusi standar yang menggunakan input dengan dimensi saluran untuk satu filter. Untuk pointwise layer, kemudian lakukan operasi perkalian filter 1x1 untuk menggabungkan hasil feature map dari depthwise layer [9]. MobileNet menggunakan arsitektur *hyperparameter* yang sederhana dan efektif, yang dapat mengaktifkan lebih sedikit parameter jaringan dan mempercepat komputasi. Selain itu, jaringan ini sangat praktis untuk pengenalan ekspresi wajah [20]. Adapun **Gambar 2.3** menunjukkan struktur jaringan MobileNet.

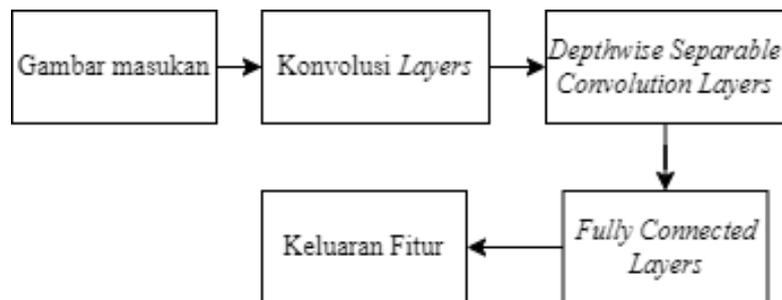


Gambar 2.3 Struktur dari MobileNet [20].

Pada **Gambar 2.3** lapisan inti MobileNet adalah filter yang dapat dipisahkan dalam. Konvolusi yang dapat dipisahkan dalam adalah bentuk dekonvolusi. Operasi konvolusi standar dibagi menjadi dua langkah: (1) mengekstraksi peta fitur, (2) melapiskan peta fitur yang diekstraksi. Konvolusi yang dapat dipisahkan kedalaman memisahkan jalur dari dua lapisan. Pertama, satu lapisan adalah konvolusi kedalaman, yang digunakan untuk mengekstraksi fitur untuk setiap saluran. Terakhir, lapisan lainnya adalah konvolusi titik demi titik, yang menggunakan konvolusi 1×1 untuk mengintegrasikan output dari langkah pertama [20].

MobileNet merupakan arsitektur yang dirancang untuk mengurangi ukuran dan kompleksitas model tanpa mengurangi kinerja pengenalan objek. MobileNet menggunakan teknik *Depthwise Separable Convolution*, yang terdiri dari dua tahap operasi konvolusi yaitu; *Depthwise Convolution* dan *Pointwise Convolution*, *depthwise convolution* memproses setiap channel pada masukan secara terpisah, dan *pointwise convolution* menggabungkan hasil dari *depthwise convolution* ke dalam satu *channel*. Dengan teknik ini, MobileNet dapat melakukan operasi

konvolusi yang lebih efisien dan menghasilkan model yang lebih ringan dibandingkan dengan arsitektur konvensional. Adapun diagram blok ekstraksi gambar menggunakan MobileNet ditunjukkan pada **Gambar 2.4**.



Gambar 2.4 Ekstraksi Gambar menggunakan MobileNet

Pada **Gambar 2.4** proses pertama yang dilakukan pada ekstraksi gambar adalah masukan gambar yang akan dimuat ke dalam jaringan MobileNet. Selanjutnya *convolution layers* akan melakukan ekstraksi fitur dengan melakukan operasi konvolusi pada gambar masukan. *Depthwise Separable Convolution* merupakan layer yang terdiri dari dua tahap yaitu *depthwise convolution* dan *pointwise convolution*, proses ini dilakukan untuk mengurangi jumlah parameter dan meningkatkan efisiensi komputasi. *Fully Connected Layers* yang akan melakukan proses klasifikasi dengan menghubungkan setiap masukan pada setiap *neuron* pada *layer* berikutnya. Kemudian, fitur keluaran merupakan fitur-fitur yang dihasilkan setelah proses ekstraksi pada gambar masukan, dan selanjutnya akan digunakan pada kasus deteksi lahan parkir.

1.3 Segmentasi Citra

Segmentasi citra adalah proses membagi citra digital menjadi beberapa segmen atau bagian, dengan tujuan untuk mengidentifikasi dan memisahkan berbagai objek dalam citra. Teknik ini biasa digunakan dalam pemrosesan gambar

dan visi komputer untuk memahami struktur dan isi gambar secara lebih detail. Tujuan utama dari segmentasi citra adalah untuk mengidentifikasi batas-batas antar objek yang berbeda dalam citra.

Segmentasi citra mengacu pada penguraian suatu gambar menjadi sejumlah area bermakna yang tidak tumpang tindih dengan atribut yang sama. Teknologi segmentasi gambar dianggap penting dalam pemrosesan gambar digital, dan akurasi segmentasi secara langsung dipengaruhi oleh efektivitas tugas-tugas tingkat lanjut. Karena tingkat kerumitan dan kesulitannya, sejumlah algoritma segmentasi telah mencapai tingkat keberhasilan yang berbeda-beda, namun penelitian pada aspek ini masih menghadapi berbagai tantangan. Sebagai salah satu teknologi utama dalam pemrosesan citra digital, segmentasi citra banyak digunakan di bidang visi mesin, pengenalan wajah, pengenalan sidik jari, sistem pemantauan lalu lintas, penentuan posisi objek dalam citra satelit (seperti jalan, hutan, dll), deteksi pejalan kaki, pencitraan medis, dan berbagai bidang lainnya [21].

Algoritma segmentasi diusulkan untuk gambar skala abu-abu, terutama dipertimbangkan dalam kaitannya dengan kesamaan tingkat keabuan dalam wilayah yang sama dan diskontinuitas tingkat keabuan antar wilayah yang berbeda. Secara umum pembagian wilayah didasarkan pada kesamaan tingkat keabuan, dan deteksi tepi didasarkan pada diskontinuitas tingkat keabuan. Segmentasi gambar berwarna melibatkan pemanfaatan kesamaan antar piksel untuk mengelompokkan gambar ke dalam wilayah atau superpiksel berbeda, yang kemudian digabungkan [22].

1.4 *Thresholding*

Thresholding merupakan salah satu metode yang sering digunakan dalam segmentasi citra, dimana berguna untuk memisahkan latar depan dan latar belakang. Dengan memilih nilai ambang batas atau *threshold* yang tepat, maka citra tingkat keabuan dapat diubah menjadi citra biner. Citra biner harus memuat semua informasi penting mengenai posisi dan bentuk objek yang menjadi fokus (*foreground*). Keuntungan mendapatkan citra biner terlebih dahulu adalah pengurangan kompleksitas data dan penyederhanaan proses pengenalan dan klasifikasi. Pendekatan umum yang digunakan untuk mengubah citra tingkat keabuan menjadi citra biner adalah dengan memilih nilai ambang batas (T). Dengan cara ini, semua nilai tingkat keabuan yang kurang dari T akan diklasifikasikan sebagai hitam (0), sedangkan nilai di atas T akan menjadi putih (1). Oleh karena itu, pemilihan nilai ambang batas T menjadi salah satu persoalan penting dalam proses segmentasi [23]. Secara umum, proses *thresholding* pada citra skala abu-abu atau *grayscale* dilakukan untuk menghasilkan citra biner [24]. Secara matematis, hal ini dapat dinyatakan pada persamaan 1 berikut.

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) > T \\ 0 & \text{if } f(x, y) \leq T \end{cases} \dots\dots\dots (1)$$

$g(x, y)$ digunakan untuk merepresentasikan citra biner yang berasal dari citra *greyscale* $f(x, y)$, sedangkan nilai ambang batas T ditentukan dengan menggunakan pendekatan ambang batas global.

1.5 *Laplacian*

Laplacian merupakan salah satu teknik yang digunakan dalam operasi pengolahan citra dalam mengekstraksi fitur-fitur dalam cira berdasarkan perubahan intensitas cahaya pada objek. Dalam mengukur intensitas cahaya dalam citra,

operator *laplacian* menggunakan diferensial kedua. Dalam domain spasial, fungsi laplacian menjadi dua dimensi $f(x, y)$ dan dapat dinyatakan sebagai berikut.

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \dots\dots\dots(1)$$

∇^2 = Operator *Laplacian*

$\frac{\partial^2}{\partial x^2}$ = Operator diferensial terhadap koordinat x

$\frac{\partial^2}{\partial y^2}$ = Operator diferensial terhadap koordinat y

Dalam konteks yang lebih umum, persamaan yang dijelaskan sebelumnya menggabungkan operasi diferensial dua detik (pada koordinat x dan y) untuk menghitung operator Laplacian dalam gambar dua dimensi atau fungsi dua dimensi. Hasil perhitungan ini berupa sejumlah nilai terukur perubahan intensitas pada setiap titik pada gambar.

Operator Laplacian digunakan dalam berbagai aplikasi dalam pemrosesan gambar, seperti deteksi tepi, analisis tekstur, dan tugas lainnya. Hal ini memungkinkan wilayah dalam gambar untuk diidentifikasi dimana intensitasnya berubah secara signifikan, yang sering kali merupakan fitur penting yang ingin dianalisis atau diekstraksi dalam konteks pemrosesan gambar.

1.6 Deteksi Tepi

Deteksi tepi dianggap sebagai salah satu masalah utama dan mendasar dalam bidang visi komputer dan pemrosesan gambar. Kontur tepi yang diekstraksi dari gambar sering kali digunakan sebagai petunjuk penting untuk berbagai tugas pemahaman gambar seperti segmentasi gambar, deteksi objek, ekstraksi fitur, dan deteksi sudut [25]. Deteksi tepi dilakukan dengan tujuan untuk mengidentifikasi batas antara dua wilayah citra homogen yang mempunyai tingkat kecerahan

berbeda. Proses deteksi tepi pada citra digital melibatkan pencarian perbedaan intensitas yang menandai batas objek (sub-gambar) pada seluruh citra digital. Tujuannya untuk menyempurnakan tampilan batas area atau objek pada suatu citra, yang dilakukan melalui deteksi tepi [25]. Sebagai proses gambar yang penting dalam pemrosesan gambar dan pengenalan pola, deteksi tepi adalah dasar ekstraksi, pengenalan, dan deteksi fitur target. Sebagai bagian penting dari pencitraan, deteksi tepi menyediakan sarana teknis yang efektif untuk mengirimkan citra, pengenalan pola, dan pencocokan gambar [26].

1.7 Webcam

Webcam adalah perangkat keras yang digunakan untuk merekam video dan foto dari komputer atau perangkat elektronik lainnya, seperti laptop atau tablet. Webcam dapat digunakan untuk beberapa hal, termasuk konferensi video, panggilan video, dan pengambilan gambar untuk diunggah ke berbagai kebutuhan. Webcam sering terhubung ke komputer menggunakan kabel USB. Lensa kamera dan sensor gambar yang disertakan di sebagian besar webcam memungkinkan mereka merekam foto dan video dengan berbagai resolusi dan kualitas. Webcam juga sering kali dijadikan sebagai perangkat untuk kebutuhan deteksi seperti deteksi objek [27] [28] [29].



Gambar 2.5 Webcam

Penggunaan webcam untuk deteksi objek dengan menggunakan metode SSD-MobileNet digunakan pada penelitian yang dilakukan oleh Mathurabai dkk. [29] menjelaskan tentang deteksi objek menggunakan SSD-MobileNet. Deteksi dalam penelitian ini menggunakan webcam yang dilakukan untuk menunjukkan deteksi objek dengan sumber daya yang terbatas. Webcam digunakan sebagai masukan gambar secara real-time yang kemudian akan dilakukan proses oleh metode SSD-MobileNet dalam melakukan pendeteksian. Hasil pada penelitian ini menyebutkan deteksi menggunakan metode SSD-MobileNet menunjukkan tingkat akurasi yang baik pada tiap objek yang dideteksi oleh sistem dengan menggunakan metode SSD-MobileNet. Adapun hasil deteksi pada penelitian ini ditunjukkan pada **Gambar 2.7**.



Gambar 2.6 Hasil Deteksi Menggunakan Webcam [29]

Gambar 2.7 menunjukkan pada penelitian ini pendeteksian objek mencapai hasil yang baik dengan webcam berkualitas rendah dan dapat mendeteksi objek secara *real-time* dengan akurasi yang baik. Dalam percobaan, Mathurabai dkk. memberikan gambar yang berbeda sebagai masukan dan model telah diidentifikasi kemudian dengan akurasi yang baik. Kemudian penggunaan webcam untuk mendeteksi objek secara *real-time* yang juga memberikan hasil yang diinginkan. Namun, penurunan penggunaan memori pada sistem dan peningkatan kinerja tidak disebutkan dalam makalah penelitian ini [29]. Adapun spesifikasi webcam yang akan digunakan pada penelitian ini ditunjukkan pada **Tabel 2.1**.

Tabel 2.1 Spesifikasi Webcam

| | |
|-----------------------|-------|
| Resolusi Lensa Kamera | 2MP |
| Resolusi Video | 1080p |
| <i>Field Angle</i> | 150° |
| Tegangan Masukan | 5V |
| <i>Focal Lens</i> | 2.8mm |