

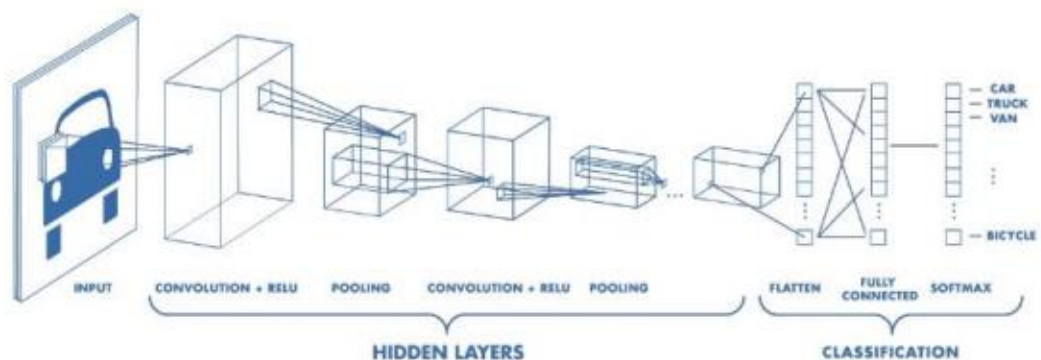
LANDASAN TEORI

2.1 Dasar Teori

Dasar teori pada penelitian ini meliputi jaringan saraf konvolusi, arsitektur VGG, CBAM, *Squeeze-and-Excitation* (SE), dan *Batch Normalization* (BN).

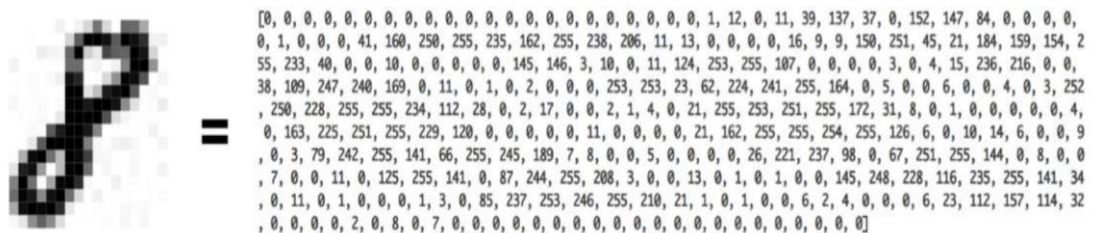
2.1.1 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan arsitektur yang memiliki 4 jenis lapisan yang diantaranya lapisan konvolusi, *pooling*, jaringan saraf, dan aktivasi. Arsitektur CNN dapat Gambar 2. 1 (Purwono et al., 2022).



Gambar 2. 1 Arsitektur CNN

Komputer melihat gambar seperti kumpulan angka yang disebut matriks. Setiap angka merepresentasikan intensitas cahaya yang disebut piksel. Berikut ilustrasi Gambar 2. 2 angka delapan diilustrasikan sebagai matriks.



Gambar 2. 2 Ilustrasi matriks gambar delapan

Tahap lapisan konvolusi sebuah filter *kernel* akan menghitung konvolusi dari gambar untuk diekstrak fitur fundamentalnya. Filter *kernel* tersebut tidak boleh melebihi ukuran gambar. *Kernel* tersebut bergeser keseluruhan gambar untuk melakukan perkalian titik matriks berdasarkan bobot filter. Perhitungan ini akan menghasilkan peta aktivasi 2D. Persamaan umum matematis konvolusi ditunjukkan pada persamaan

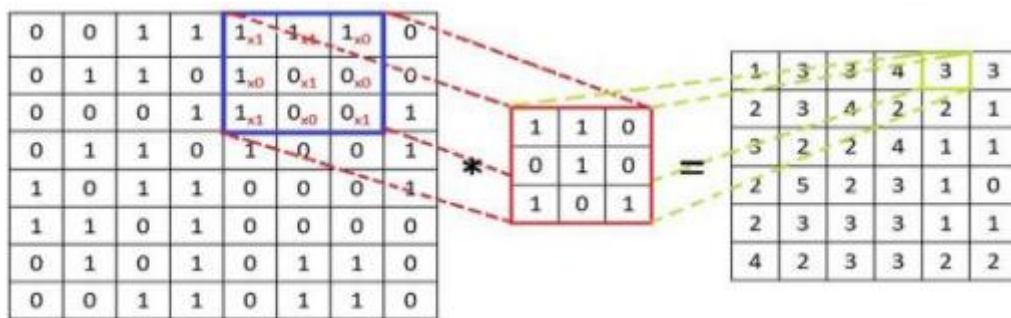
$$Activation\ map = Input * Filter$$

$$= \sum_{y=0}^{columns} (\sum_{x=0}^{rows} Input(x - p, y - q) Filter(x, y)) \quad (1)$$

dan ilustrasi konvolusi pada Gambar 2. 3 (Purwono et al., 2022).

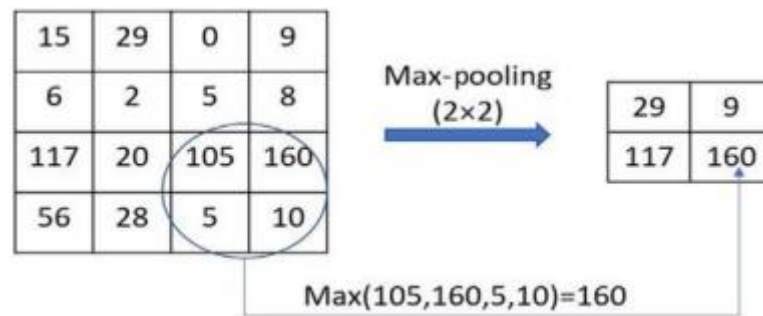
$$Activation\ map = Input * Filter$$

$$= \sum_{y=0}^{columns} (\sum_{x=0}^{rows} Input(x - p, y - q) Filter(x, y)) \quad (1)$$

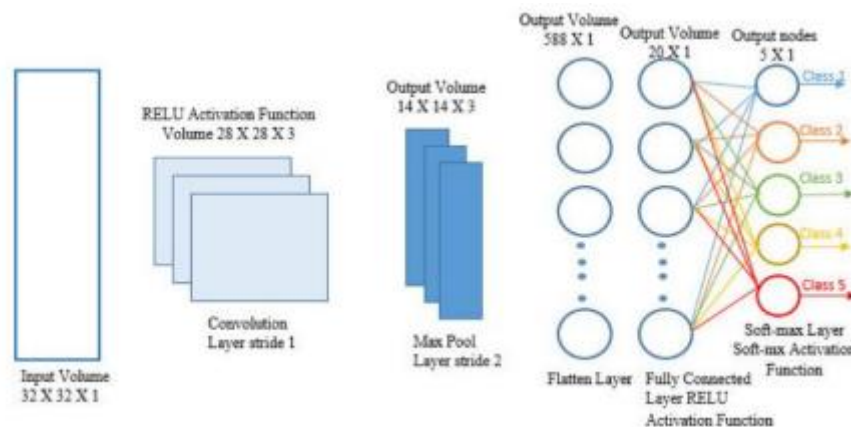


Gambar 2. 3 Ilustrasi konvolusi

Tahap lapisan *pooling* akan menggabungkan dua lapisan konvolusi secara berurutan. Metode menurunkan sampel tersebut dapat mengurangi jumlah parameter dan komputasi. Fungsi untuk melakukan hal tersebut biasanya menggunakan fungsi maksimal atau rata-rata. Berikut ilustrasi *pooling* pada Gambar 2. 4 (Purwono et al., 2022).

Gambar 2. 4 Ilustrasi lapisan *pooling*

Tahap lapisan jaringan saraf digunakan untuk mempelajari tingkat tinggi nongaris lurus kombinasi dari fitur yang direpresentasikan oleh lapisan konvolusi. Selain itu, jaringan saraf dapat digunakan sebagai lapisan aktivasi dengan menyesuaikan fungsi aktivasinya. Berikut ilustrasi jaringan saraf pada Gambar 2. 5 (Purwono et al., 2022).

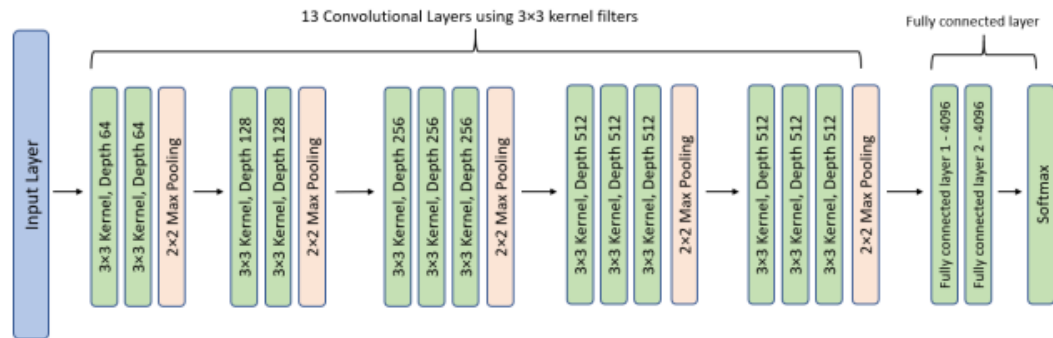


Gambar 2. 5 Ilustrasi jaringan saraf

2.1.2 Arsitektur *Visual Geometry Group (VGG)*

Arsitektur *Visual Geometry Group (VGG)* merupakan model jaringan konvolusional yang sangat dalam. Pada awalnya model tersebut digunakan untuk menyelesaikan permasalahan pada kompetisi *ImageNet Challenge* 2014 dengan menempati posisi pertama dan kedua pada lokalisasi dan trek klasifikasi. Dengan memanfaatkan filter (3x3) Model VGG memanfaatkan kedalamannya sehingga

dapat meningkatkan performa secara signifikan (Simonyan & Zisserman, 2014). Berikut merupakan penerapan arsitektur VGG-16 yang dilakukan oleh tammina dkk. pada Gambar 2. 6 (Tammina, 2019).

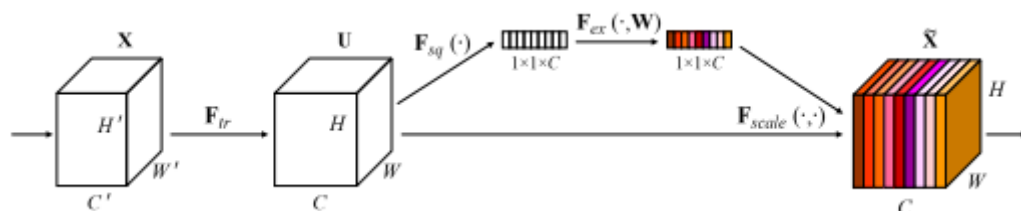


Gambar 2. 6 Penerapan VGG-16

Gambar 2. 6 merupakan penerapan VGG-16 dengan memanfaatkan 13 lapisan konvolusional dengan (3×3) *kernel* dan 3 lapis jarngan saraf. Pada lapisan *softmax* digunakan sebagai lapisan penggolong (Tammina, 2019).

2.1.3 Squeeze-and-Excitation (SE)

Squeeze-and-Excitation Networks (SE) merupakan dasar pengembangan dari metode *Convolutional Block Attention Module* (CBAM). Metode ini berfokus pada *attention* kanal. SE memiliki dua tahap yaitu meremas dan *excitation*. Berikut arsitektur dari SE pada Gambar 2. 7 (Hu et al., 2018).



Gambar 2. 7 Arsitektur SE

Tahap meremas merupakan tahap untuk menentukan koefisien dari setiap kanal. Hal tersebut dapat dilakukan dengan metode *global average pooling*. Berikut persamaan untuk *global average pooling* pada persamaan $z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$. (2) (Hu et al., 2018).

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (2)$$

Tahap *excitation* merupakan tahap untuk menentukan kanal yang lebih penting pada gambar. Untuk melakukan hal tersebut dilakukan mekanisme gerbang. Berikut persamaan mekanisme gerbang pada persamaan $s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$ (3) (Hu et al., 2018).

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

Dimana:

δ = fungsi *ReLU*

$$W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$$

$$W_2 \in \mathbb{R}^{\frac{C}{r} \times C}$$

Untuk melimit kompleksitas model dan membantu generalisasi dilakukan parameterisasi mekanisme gerbang dengan melakukan *bottleneck* pada persamaan $s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$ (3). *Bottleneck* tersebut dilakukan dengan dua lapis jaringan saraf di sekeliling nongaris lurus. Contohnya, lapisan reduksi dimensi dengan parameter W_1 dengan rasio reduksi r sebuah *ReLU* dan sebuah peningkatan dimensi dengan parameter W_2 . Untuk menghasilkan luaran akhir dapat digunakan metode *rescaling* dari hasil transformasi konvolusi U seperti yang ditunjukkan pada Gambar 2. 7 dengan

aktivasi. Berikut persamaan *rescaling* $\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c$ (4) (Hu et al., 2018).

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

Dimana:

$$\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_C]$$

$F_{scale}(u_c, s_c)$ merupakan perkalian kanal diantara peta fitur $u_c \in R^{H \times W}$ dan skalar s_c .

2.1.4 Convolutional Block Attention Module (CBAM)

Convolutional Block Attention Module (CBAM) merupakan modul *attention* yang efektif namun didesain dengan sederhana untuk konvolusional *feed-forward neural network*. Dengan adanya *intermediate feature map* modul ini memilih secara berurutan menyimpulkan *attention maps* sepanjang dua dimensi yang terpisah yaitu kanal dan spasial. Kemudian, *attention maps* dikalikan dengan *input feature map* untuk penyempurnaan fitur adaptif (Woo et al., 2018). Dari pernyataan tersebut CBAM merupakan suatu modul yang memanfaatkan dua algoritma yaitu kanal dan spasial. Adapun rangkuman persamaan algoritma pada persamaan $F'' = M_s(F') \otimes F'$, (5) dan arsitektur dari CBAM pada Gambar 2. 8 dan Gambar 2. 9 (Woo et al., 2018).

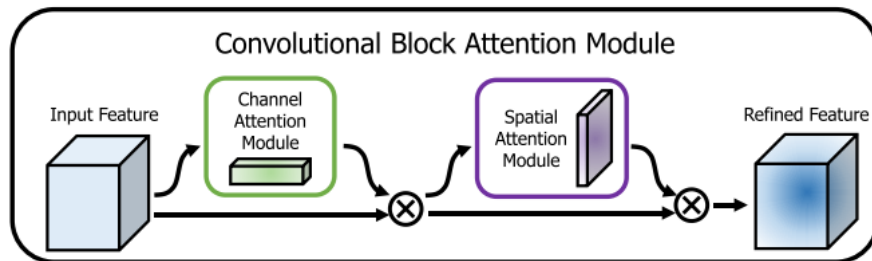
$$F' = M_c(F) \otimes F,$$

$$F'' = M_s(F') \otimes F', \quad (5)$$

$$\text{Pada persamaan } F'' = M_s(F') \otimes F', \quad (5) \otimes$$

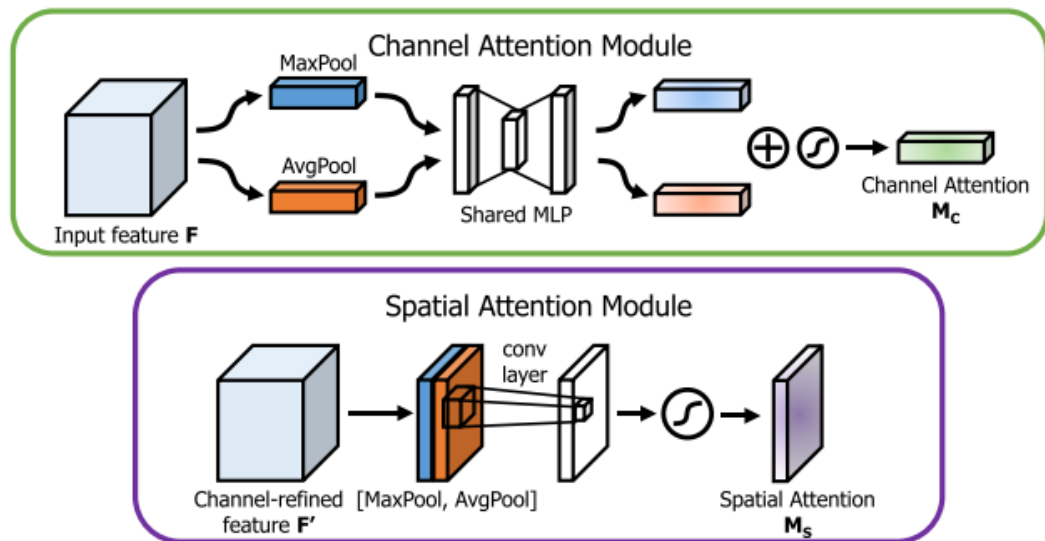
menunjukkan perkalian elemen. Selama perkalian, nilai *attention* disalin. Dengan

demikian, *attention* kanal disiarkan bersamaan dengan dimensi spasial maupun sebaliknya. F'' merupakan hasil akhir dari proses tersebut (Woo et al., 2018).



Gambar 2. 8 Rangkuman arsitektur CBAM

Pada Gambar 2. 8 merupakan rangkuman arsitektur CBAM. Arsitektur CBAM terdiri dari dua submodul yaitu kanal dan spasial. Peta fitur perantara diproses secara adaptif di setiap konvolusional jaringan dalam (Woo et al., 2018).



Gambar 2. 9 Diagram submodul CBAM

Pada Gambar 2. 9 modul kanal menggunakan luaran *max-pooling* dan *average-pooling* dengan berbagi jaringan. Modul spasial menggunakan dual luaran yang digabungkan bersamaan dengan axis kanal dan melanjutkan ke lapisan konvolusional. Adapun bentuk persamaan $= \sigma \left(W_1 \left(W_0 (F_{avg}^c) \right) \right) +$

$$W_1(W_0(F_{max}^c)) \quad (6) \text{ dan } = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s]))$$

(7) dari dua modul tersebut (Woo et al., 2018).

$$\begin{aligned} M_c(F) &= \sigma \left(MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \\ &= \sigma \left(W_1 \left(W_0(F_{avg}^c) \right) + W_1 \left(W_0(F_{max}^c) \right) \right) \end{aligned} \quad (6)$$

Dimana:

σ = fungsi sigmoid

$$W_0 \in \mathbb{R}^{\frac{C}{r} \times C}$$

$$W_1 \in \mathbb{R}^{C \times C/r}$$

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \end{aligned} \quad (7)$$

Dimana:

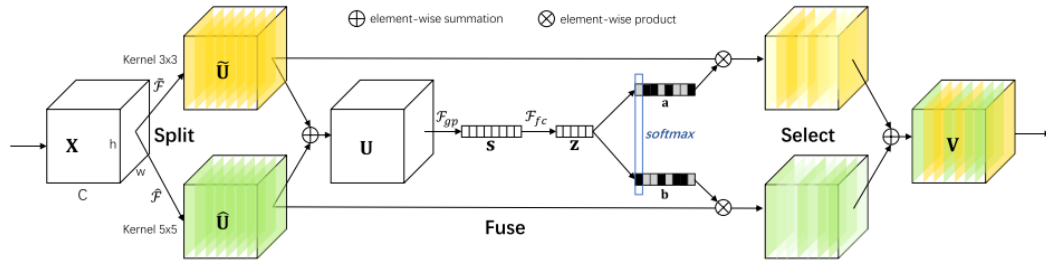
σ = fungsi sigmoid

$f^{7 \times 7}$ = konvolusi operator dengan ukuran filter (7x7)

2.1.5 Selective Kernel (SK)

Selective Kernel (SK) merupakan sebuah mekanisme CNN yang adaptif untuk dapat menyesuaikan bidang reseptifnya pada berbagai skala masukan. Dengan metode tersebut, didapatkan hasil yang melebihi *state-of-the-art* pada *dataset Imagenet* dan CIFAR dengan kompleksifitas model lebih rendah. SK memiliki tiga tahap yang diantaranya membagi, mengabung, dan memilih.

Rangkuman arsitektur dengan *kernel* (3x3) dan (5x5) terdapat pada Gambar 2. 10 (Li et al., 2019).



Gambar 2. 10 Rangkuman arsitektur SK

Tahap membagi merupakan pemrosesan peta fitur gambar. Dengan yang dicontohkan pada Gambar 2. 10 melibatkan dua operasi konvolusi dengan *kernel* (3x3) dan (5x5). Untuk meningkatkan efisiensi digunakan grup/*depthwise* konvolusi, normalisasi, dan fungsi *ReLU* (Li et al., 2019).

Tahap menggabungkan bertujuan untuk memuat bidang reseptif yang adaptif berdasarkan masukannya. Tahap ini dilakukan dengan menggunakan gerbang untuk mengontrol aliran informasi dari cabang berbeda dengan skala berbeda ke saraf di lapisan berikutnya. Untuk mencapai hal tersebut gerbang perlu mengintegrasikan informasi dari semua cabang. Berikut bentuk persamaan $U = \tilde{U} + \hat{U}$ (8) dari proses tersebut (Li et al., 2019).

$$U = \tilde{U} + \hat{U} \quad (8)$$

Kemudian kita menanamkan informasi global dengan menggunakan *global average pooling* untuk membuat statistik kanal sebagai $s \in R^C$. Proses *global average pooling* ditunjukkan pada persamaan $z_c = F_{sq}(u_c) =$

$$\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (2) \text{ (Li et al., 2019).}$$

Fitur kompak $z \in \mathbb{R}^{d \times 1}$ dibuat untuk dapat memandu sehingga memilih secara presisi dan adaptif. Hal tersebut dapat dicapai dengan lapisan jaringan saraf dengan reduksi dimensi untuk efisiensi. Berikut bentuk persamaan $z = F_{fc}(s) = \delta(B(W_s))$ (9) dari proses tersebut (Li et al., 2019).

$$z = F_{fc}(s) = \delta(B(W_s)) \quad (9)$$

Dimana:

δ = fungsi *ReLU*

B = *Batch Normalization*

$W \in \mathbb{R}^{d \times c}$

Rasio reduksi digunakan untuk mempelajari dampak dari d dalam efisiensi model. Digunakan rasio reduksi r untuk mengontrol nilai tersebut. Berikut bentuk persamaan $d = \max(\frac{c}{r}, L)$ (10) dari proses tersebut (Li et al., 2019).

$$d = \max(\frac{c}{r}, L) \quad (10)$$

Dimana:

L = nilai minimal dari d

Tahap memilih merupakan proses untuk memilih informasi skala spasial yang berbeda dari sebuah gambar. Tahap ini dilakukan dengan menggunakan mekanisme *soft attention* yang dipandu oleh fitur deksriptor yang ringkas z . Mekanismenya menggunakan operator *softmax* untuk memilih saluran yang paling penting. Berikut bentuk persamaan $a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}}$ (11) dari proses tersebut (Li et al., 2019).

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}}, b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}}, \quad (11)$$

Dimana:

$$A, B \in \mathbb{R}^{C \times d}$$

$$A, B = \text{soft attention vector } \tilde{U}$$

$$a, b = \text{soft attention vector } \hat{U}$$

Peta fitur akhir V didapat melalui bobot *attention* dari berbagai *kernel*. Berikut

$$\text{bentuk ilustrasi persamaan } V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c, a_c + b_c = 1, \quad ($$

12) dari proses tersebut (Li et al., 2019).

$$V_c = a_c \cdot \tilde{U}_c + b_c \cdot \hat{U}_c, a_c + b_c = 1, \quad (12)$$

Dimana:

$$V = [V_1, V_2, \dots, V_C]$$

$$V_c \in \mathbb{R}^{H \times W}$$

2.1.6 Batch Normalization (BN)

Batch Normalization (BN) merupakan sebuah teknik untuk menormalkan aktivasi ditengah lapisan jaringan saraf dalam. Teknik ini memiliki kecenderungan untuk meningkatkan akurasi dan kecepatan pada proses pelatihan sehingga banyak digemari. BN biasa digunakan pada kasus jaringan saraf konvolusi. BN memiliki empat dimensi input dan output yang dinotasikan sebagai $I_{b,c,x,y}$ dan $O_{b,c,x,y}$. Dimensi tersebut merujuk pada b untuk *batch*, c untuk kanal, dan dua dimensi spasial x, y . Pada kasus masukan gambar dengan kanal RGB. BN menerapkan normalisasi yang sama untuk semua aktivasi di setiap kanal. Adapun persamaan BN

pada persamaan $O_{b,c,x,y} \leftarrow \gamma_c \frac{I_{b,c,x,y} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \quad \forall b, c, x, y.$ (13)

(Bjorck et al., 2018).

$$O_{b,c,x,y} \leftarrow \gamma_c \frac{I_{b,c,x,y} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \quad \forall b, c, x, y. \quad (13)$$

BN mengurangi aktivasi rata-rata $\mu_c = \frac{1}{|\beta|} \sum_{b,x,y} I_{b,c,x,y}$ dari semua masukan aktivasi di kanal c , dimana β berisi semua aktivasi dalam kanal c disemua fitur b diseluruh *mini-batch* dan semua lokasi spasial x, y . Selanjutnya, BN membagi aktivasi terpusat dengan standard deviasi σ_c (ditambah ϵ untuk stabilitas numerik) yang dihitung secara analog. Selama pengujian, *running averages of mean* dan varian digunakan. Normalisasi diikuti oleh transformasi *affine* berdasarkan kanal yang diparametrikkan melalui γ_c, β_c yang dipelajari selama pelatihan (Bjorck et al., 2018).

2.2 Metrik Penelitian

Metrik Penelitian pada penelitian ini meliputi akurasi, presisi, *recall*, *f-1 score*, dan *sparse top-k categorical accuracy*.

2.2.1 Akurasi

Akurasi adalah jumlah pengklasifikasian yang benar pada seluruh set pengujian atau bisa dikatakan sebagian kecil dari prediksi pengklasifikasi yang benar di seluruh set pengujian. Adapun persamaan akurasi pada persamaan $Accuracy Rate = \frac{TP+TN}{TP+TN+FP+FN}$ (14) (Jude Chukwura Obi, 2023).

$$Accuracy Rate = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Dimana:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

2.2.2 Presisi

Presisi adalah rasio *True Positive* (TP) dari semua hasil positif yang diobservasi. Semua hasil positif ini termasuk *False Positive* (FP). Adapun persamaan presisi pada persamaan $Precision = \frac{TP}{TP+FP}$ (15) (Jude Chukwura Obi, 2023).

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

Dimana:

TP = True Positive

FP = False Positive

2.2.3 Recall

Recall atau kepekaan adalah kemampuan model untuk dapat mengidentifikasi *True Positive* (TP). Adapun persamaan kepekaan pada persamaan $Recall = \frac{TP}{TP+FN}$ (16) (Jude Chukwura Obi, 2023).

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

Dimana:

TP = True Positive

FN = False Negative

2.2.4 F-1 Score

F-1 score merupakan rata-rata tertimbang presisi dan kepekaan. Lebih lanjut, *f-1 score* merupakan pengukuran akurasi tes dan dihitung dari presisi dan kepekaan tes. Adapun persamaan *f-1 score* pada persamaan *F1 Score* =

$$\frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17) \text{ (Jude Chukwura Obi, 2023).}$$

$$F1 \text{ Score} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (17)$$

2.2.5 Sparse Top-K Categorical Accuracy

Sparse top-k categorical accuracy merupakan metrik yang digunakan dalam kasus dengan kelas banyak dimana labelnya merupakan angka (0, 1, ..., total kelas). Metrik ini membandingkan label benar dengan K prediksi model yang paling mungkin. Adapun persamaan metrik ini pada persamaan

$$SparseTopKCategorycalAccuracy = \frac{y_{pred}}{Total \text{ number of samples}} \quad (18)$$

(HAIBAL Team, 2024).

$$SparseTopKCategorycalAccuracy = \frac{y_{pred}}{Total \text{ number of samples}} \quad (18)$$

Dimana:

y_{pred} = Jumlah sampel yang label sebenarnya berada dalam K kelas prediksi yang paling mungkin

2.3 Penelitian Terkait (*State of the Art*)

Penelitian terkait mengenai teori yang digunakan disajikan dalam Tabel 2. 1.

Tabel 2. 1 Penelitian terkait

No	Konten	Deskripsi
1	Judul paper dan Penulis	Attention mechanisms in computer vision: A survey (M. H. Guo et al., 2022)
	Permasalahan	Menemukan bagian yang penting pada gambar yang kompleks secara efektif
	Kontribusi	Menyediakan tinjauan dari banyak mekanisme <i>attention</i> dalam bidang visi komputer dan mengkategorikannya berdasarkan pendekatannya
	Metode/ Solusi	Survei
	Batasan	Mekanisme <i>attention</i> dibidang visi komputer
2	Judul paper dan Penulis	AVNC: Attention-Based VGG-Style Network for COVID-19 Diagnosis by CBAM (S.-H. Wang et al., 2022)
	Permasalahan	Mendeteksi pasien COVID-19 lebih akurat dan presisi
	Kontribusi	12 lapisan dengan basis <i>attention</i> dan arsitektur VGG
	Metode/ Solusi	12 lapisan dengan basis <i>attention</i> dan arsitektur VGG
	Batasan	Penelitian ini berfokus pada pengembangan arsitektur VGG dengan <i>attention</i> mekanisme CBAM dengan kasus COVID-19
3	Judul paper dan Penulis	A quality grade classification method for fresh tea leaves based on an improved YOLOv8x-SPPCSPC-CBAM model (Zhao et al., 2024)
	Permasalahan	Mencegah masalah umum yang terjadi pada penilaian daun teh yang dikarakterkan memiliki penilaian kerusakan yang tinggi dan akurasi yang buruk. selain itu, penilaian secara presisi dengan integrasi dengan pembelajaran mesin dan visi mesin yang terbatas.
	Kontribusi	Menampilkan sebuah pendekatan yang inovatif untuk mengklasifikasikan mutu daun teh
	Metode/ Solusi	Mengintegrasikan algoritma rekognisi gambar dan jaringan saraf dalam untuk mengklasifikasi daun teh secara akurat dengan membedakan antara kombinasi kuncup dan daun
	Batasan	Penelitian ini berfokus pada metode <i>YOLOv8x-SPPCSPC</i> dan mekanisme <i>attention</i> CBAM untuk mengklasifikasikan kualitas daun teh
4	Judul paper dan Penulis	CNTCB-YOLOv7: An Effective Forest Fire Detection Model Based on ConvNeXtV2 and CBAM (Xu et al., 2024)

	Permasalahan	Mengidentifikasi fitur dan aspek yang substil dari api yang dapat memberikan tantangan yang signifikan untuk model pembelajaran dalam.
	Kontribusi	Meningkatkan model untuk dapat merepresentasikan fitur dan presisi dalam deteksi. penelitian ini membandingkan metode <i>ConvNeXtV2</i> dan <i>Conv2Former</i> pada algoritma <i>You Only Look Once version 7</i> (YOLOv7) lalu membandingkan hasilnya dengan YOLOv7 yang orisinal
	Metode/ Solusi	Perbandingan komprehensif metode yang digunakan dengan YOLOv7 orisinal pada kasus deteksi kebakaran hutan
	Batasan	Penelitian ini berfokus pada metode <i>ConvNeXtV2</i> dan <i>Conv2Former</i> pada algoritma dasar YOLOv7 dengan kasus deteksi kebakaran hutan
5	Judul paper dan Penulis	Large Selective Kernel Network for Remote Sensing Object Detection (Y. Li et al., 2023)
	Permasalahan	Mengabaikan pengetahuan sebelumnya yang unik dalam skenario penginderaan remot
	Kontribusi	Model ringan LSKNet menjadi <i>state-of-the-art</i> pada <i>dataset</i> tolak ukur
	Metode/ Solusi	Pada penelitian ini mempertimbangkan pengetahuan sebelumnya dan mengajukan model ringan <i>Large Selective Kernel Network</i> (LSKNet)
6	Batasan	Penelitian ini berfokus pengembangan dengan metode jaringan <i>selective kernel</i> pada kasus dekteksi objek penginderaan remot
	Judul paper dan Penulis	SK-MobileNet: A Lightweight Adaptive Network Based on Complex Deep Transfer Learning for Plant Disease Recognition (Liu et al., 2023)
	Permasalahan	Gambar latar belakang terkadang kompleks dan model dapat mendeteksi lesi tanaman. Namun, sulit digunakan dan sulit mendeteksi spesifik lokasi hama. Kompleksitas model yang tinggi sehingga tidak kondusif pada penyebaran dan pengembangan perangkat lunak seluler
	Kontribusi	Menginvestigasi jaringan konvolusi dalam dengan pembelajaran transfer untuk rekognisi penyakit tumbuhan
7	Metode/ Solusi	Mengusulkan model <i>Selective Kernel MobileNet</i> (SK-MobileNet) yang cukup ringan sehingga dapat mengurangi komputasi pada penyebaran di server
	Batasan	Penelitian ini berfokus pada pendekatan <i>selective kernel</i> dan MobileNet pada kasus rekognisi penyakit tumbuhan
	Judul paper dan Penulis	LungSeek: 3D Selective Kernel residual network for pulmonary nodule diagnosis (Zhang & Zhang, 2023)
	Permasalahan	Meningkatkan kemungkinan bertahan kanker paru-paru dengan deteksi awal dan diagnosis nodul paru
	Kontribusi	Mengusulkan kerangka LungSeek berdasarkan 3D SK-Resnet untuk deteksi nodul dan klasifikasi nodul dari CT nodul dada

	Metode/ Solusi	Mengusulkan sebuah sistem otomisasi diagnoisi kanker paru LungSeek
	Batasan	Penelitian ini berfokus pada pendekatan jaringan sisa dan jaringan 3D <i>selective kernel</i> pada kasus deteksi nodul dan klasifikasi nodul dari CT nodul dada
8	Judul paper dan Penulis	An Integrated Bearing Fault Diagnosis Method Based on Multibranch SKNet and Enhanced Inception-ResNet-v2 (B. Hu et al., 2024)
	Permasalahan	diagnosis kegagalan <i>roller bearing</i>
	Kontribusi	mengintegrasikan jaringan saraf dalam yang menggabungkan jaringan <i>selective kernel</i> yang ditingkatkan dengan <i>Inception-ResNet-v2</i> yang ditingkatkan
	Metode/ Solusi	mengintegrasikan jaringan saraf dalam yang menggabungkan jaringan <i>selective kernel</i> yang ditingkatkan dengan <i>Inception-ResNet-v2</i> yang ditingkatkan
	Batasan	Penelitian ini berfokus pada pengintegrasian <i>selective kernel</i> dengan <i>Inception-ResNet-v2</i> pada kasus diagnosis kegagalan <i>roller bearing</i>
9	Judul paper dan Penulis	SSGNet: Selective Multi-Scale Receptive Field and Kernel Self-Attention Based on Group-Wise Modality for Brain Tumor Segmentation (B. Guo et al., 2024)
	Permasalahan	algoritma pemrosesan gambar medis tidak efektif dalam memanfaatkan karakteristik multi-modalitas untuk mengekstrak fitur lebih lanjut
	Kontribusi	dengan metode yang diajukan mendapatkan performa <i>state-of-the-art</i> jika dibandingkan dengan lebih dari 12 tolok ukur
	Metode/ Solusi	Menggabungkan beberapa pendekatan yang diantaranya <i>selective multi-scale receptive field</i> (SMRF), <i>selective kernel self-attention</i> (SKSA) <i>skip connection attention module</i> (SCAM) dengan basis arsitektur <i>Unet</i>
	Batasan	Penelitian ini berfokus pada pendekatan beberapa pendekatan yang diantaranya <i>selective multi-scale receptive field</i> (SMRF), <i>selective kernel self-attention</i> (SKSA) <i>skip connection attention module</i> (SCAM) dengan basis arsitektur <i>Unet</i> pada kasus segmentasi tumor otak
10	Judul paper dan Penulis	Pulmonary Nodule Classification Using a Multiview Residual Selective Kernel Network (Chang et al., 2024)
	Permasalahan	Diperlukan peningkatan pada diagnosis berbasis komputer (CAD) pada kasus kanker paru
	Kontribusi	menginvestigasi diagnosis berbasis komputer (CAD) untuk memprediksi kemungkinan tumor ganas pada komputasi tomografi (CT) nodul paru dengan pendekatan pembelajaran dalam
	Metode/ Solusi	efisien <i>residual selective kernel</i> (RSK) blok untuk mengatasi keberagaman nodul paru dengan berbagai bentuk dan struktur yang tidak jelas

	Batasan	Penelitian ini berfokus pada pengembangan dengan pendekatan <i>residual</i> dan <i>selective kernel</i> pada prediksi tumor paru
11	Judul paper dan Penulis	EFFICIENT LEARNED IMAGE COMPRESSION WITH SELECTIVE KERNEL RESIDUAL MODULE AND CHANNEL-WISE CAUSAL CONTEXT MODEL (Fu et al., n.d.)
	Permasalahan	kompleksitas yang cukup tinggi pada pendekatan dengan basis pembelajaran kompresi gambar
	Kontribusi	pada hasil eksperimen membuahkan hasil yang lebih baik pada pendekatan kompresi gambar terbaru
	Metode/ Solusi	menggunakan dua efisien modul yang pertama modul <i>selective kernel residual</i> yang kedua kanal causal kontek yang ditingkatkan
	Batasan	penelitian ini berfokus pada pengembangan metode dengan pendekatan <i>selective kernel</i> dan kanal causal kontek pada kasus pembelajaran kompresi gambar
12	Judul paper dan Penulis	Data-Augmentation Based CBAM-ResNet-GCN Method for Unbalance Fault Diagnosis of Rotating Machinery (H. Wang et al., 2024)
	Permasalahan	sulit dalam mendapatkan sample data untuk melatih model dalam kasus kegagalan mesin berputar
	Kontribusi	dengan metode yang diajukan mendapatkan hasil yang baik pada <i>dataset</i> tolok ukur
	Metode/ Solusi	penelitian ini menggunakan <i>Convolutional Attention Residual Network</i> (CBAM-ResNet) dengan jaringan konvolusi grafik
	Batasan	Penelitian ini berfokus pada pengembangan dengan pendekatan <i>Convolutional Attention Residual Network</i> (CBAM-ResNet) dengan jaringan konvolusi grafik pada kasus kegagalan mesin berputar
13	Judul paper dan Penulis	PH-CBAM: A Parallel Hybrid CBAM Network with Multi-Feature Extraction for Facial Expression Recognition (Liao et al., 2024)
	Permasalahan	Terdapat tantangan dalam memfokuskan dan mengekstrak fitur wajah secara efektif
	Kontribusi	penelitian ini mengembangkan metode dengan memfokuskan pada fitur lokal pada wajah medapat hasil lebih baik pada <i>dataset</i> tolok ukur jika dibandingkan dengan model yang sering digunakan
	Metode/ Solusi	penelitian ini menggunakan metode dengan pendekatan <i>Convolutional Block Attention Module</i> (CBAM) dan <i>Parallel Hybrid Attention Model</i>
	Batasan	Penelitian ini berfokus pada pendekatan <i>Convolutional Block Attention Module</i> (CBAM) dan <i>Parallel Hybrid Attention Model</i> pada kasus rekognisi ekspresi wajah

14	Judul paper dan Penulis	Real-Time Underwater Fish Detection and Recognition Based on CBAM-YOLO Network with Lightweight Design (Z. Yan et al., 2024)
	Permasalahan	diperlukan suatu alat yang dibutuhkan robot bawah air untuk menemukan dan mengakui kehidupan laut secara autonomus
	Kontribusi	Penelitian ini mengajukan sebuah metode jaringan saraf konvolusi untuk mendeteksi 6 jenis ikan secara cerdas
	Metode/ Solusi	Penelitian ini menggunakan metode dengan pendekatan YOLO dan <i>Convolutional Block Attention Module</i> (CBAM)
	Batasan	Penelitian ini berfokus pada pengembangan dengan pendekatan YOLO dan <i>Convolutional Block Attention Module</i> (CBAM) pada kasus deteksi dan rekognisis jenis ikan
15	Judul paper dan Penulis	Enhanced object detection in pediatric bronchoscopy images using YOLO-based algorithms with CBAM attention mechanism (J. Yan et al., 2024)
	Permasalahan	Terdapat tantangan dalam memvisualkan saluran bronkial dan paru dikarenakan berbagai macam objek seperti mukus, darah, dan benda asing.
	Kontribusi	Dengan metode yang digunakan terdapat peningkatan akurasi dan keandalan secara signifikan pada gambar bronkoskopi
	Metode/ Solusi	Penelitian ini menggunakan metode dengan pendekatan YOLO dan <i>Convolutional Block Attention Module</i> (CBAM)
	Batasan	Penelitian ini berfokus pada pengembangan dengan pendekatan YOLO dan <i>Convolutional Block Attention Module</i> (CBAM) pada kasus deteksi objek gambar bronkoskopi
16	Judul paper dan Penulis	Lightweight YOLOv7 Algorithm for Multi-Object Recognition on Contrabands in Terahertz Images (Ge et al., 2024)
	Permasalahan	Gambar terahertz memiliki kualitas yang buruk dan kurang detail tekstur. Selain itu, metode tradisional <i>target detection</i> memiliki deteksi yang lambat, misdeteksi, dan kehilangan barang selundupan.
	Kontribusi	Penelitian ini menggunakan berbagai metode praproses gambar. Selain itu, penelitian ini mengimplementasikan metode dengan pendekatan YOLO yang ringan dan <i>selective kernel</i> yang besar
	Metode/ Solusi	Penelitian ini menggunakan berbagai metode praproses gambar. Selain itu, penelitian ini mengimplementasikan metode dengan pendekatan YOLO yang ringan dan <i>selective kernel</i> yang besar
	Batasan	Penelitian ini berfokus pada pengembangan dengan pendekatan YOLO yang ringan dan <i>selective kernel</i> yang besar. Selain itu, berbagai <i>pre-prcessing</i> gambar diterapkan

Teknisi jaringan merupakan penelitian yang paling penting pada kasus visi komputer karena jaringan yang didesain dengan baik memastikan peningkatan performa dibanyak aplikasi. Banyak implementasi arsitektur CNN skala besar seperti *Inception*, *ResNet*, dan yang lainnya. Semua metode tersebut berfokus pada kedalaman, lebar, dan kardinalitas. Penelitian ini berfokus pada pengembangan *attention* salah satu aspek yang membuat penasaran sistem penglihatan manusia (Woo et al., 2018).

Mekanisme *attention* mempunyai fungsi yang penting pada penglihatan manusia. Dengan mekanisme tersebut penglihatan manusia tidak melihat secara keseluruhan. Namun, serangkaian pandangan sekilas parsial dan secara selektif fokus pada bagian menonjol untuk melihat (Woo et al., 2018). Terdapat banyak mekanisme *attention* yang cukup populer diantaranya CBAM dan SK.

CBAM merupakan peningkatan dari SE dimana CBAM tidak hanya meningkatkan secara kanal namun secara spasial. Selain itu, CBAM mudah diintegrasikan dengan model CNN. Pada penelitian yang dilakukan Woo dkk. mendapatkan hasil yang baik dengan eror 21.07% pada kasus klasifikasi gambar *ImageNet-1K* (Woo et al., 2018).

SK merupakan peningkatan dari *Dynamic Convolution* dimana SK dapat secara adaptif memilih bidang persepsi. Pada penelitian yang dilakukan oleh X.Li dkk. mendapat hasil yang baik dengan eror 20.19% pada kasus klasifikasi gambar *ImageNet* (X. Li et al., 2019).

Penelitian ini berkontribusi dalam mengintegrasikan SK dan CBAM. Dengan mengintegrasikan metode tersebut dapat menyelesaikan permasalahan

penelitian yaitu memperkaya ekstraksi fitur dan mengadaptifkan *kernel* model. Selain itu, BN akan diterapkan model untuk mempercepat konvergensi dan meningkatkan generalisasi model. Dengan demikian, metode yang digunakan pada penelitian ini sejalan dengan tujuan penelitian.