

Klasifikasi Kardus Barang di PT XYZ Menggunakan *Convolutional Neural Network* dengan Pendekatan *Fine Grained Image Classification*

Alief Yuwastika Firmandicky¹, Yeremia Alfa Susetyo^{2*}

^{1,2*} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

article info

Article history:

Received 6 March 2024
Received in revised form
11 May 2024
Accepted 30 July 2024
Available online October
2024.

DOI:

<https://doi.org/10.35870/jti.k.v8i4.2337>.

Keywords:

Convolutional Neural
Network; Fine Grained Image
Classification; ResNet
Architecture.

Kata Kunci:

Convolutional Neural
Network; Fine Grained Image
Classification; ResNet
Architecture.

abstract

PT XYZ requires a system to automatically validate items in storage with the system. Cardboard boxes containing items exhibit high visual similarity within a specific sub-category. Several studies have demonstrated the use of Convolutional Neural Network (CNN) as a method for image classification with a Fine Grained Image Classification (FGIC) approach for classifying data with high similarity, resulting in good accuracy, and this will be applied in this research. The ResNet architecture is used with and without ImageNet weight initialization, combined with the RMCSAM architecture, resulting in eight training configurations. Based on testing results using 172 images across 14 classes, the ResNet + RMCSAM configuration with ImageNet weight initialization and the 20 times augmentation dataset achieves the highest accuracy compared to other configurations, with an accuracy of 99.42% and a loss of 0.0004. This configuration is utilized for cardboard classification in the PT XYZ warehouse.

abstract

PT XYZ memerlukan sistem untuk dapat memvalidasi barang pada penyimpanan dengan sistem secara otomatis. Kardus barang memiliki kemiripan visual yang tinggi pada satu sub-kategori barang. Beberapa penelitian sudah membuktikan penggunaan Algoritma Convolutional Neural Network (CNN) sebagai metode dalam klasifikasi gambar dengan pendekatan Fine Grained Image Classification (FGIC) untuk pengklasifikasian data dengan tingkat kemiripan tinggi menghasilkan akurasi yang baik akan dipakai dalam penelitian ini. Penggunaan arsitektur ResNet dengan dan tidak dengan inisialisasi bobot serta dikombinasikan dengan arsitektur RMCSAM sehingga memiliki 8 konfigurasi pelatihan. Berdasarkan hasil pengujian menggunakan data 172 gambar dengan 14 kelas, konfigurasi model ResNet + RMCSAM dengan inisialisasi bobot ImageNet lalu menggunakan dataset Augmentasi 20 kali memiliki akurasi terbaik daripada konfigurasi lainnya dengan akurasi sebesar 99,42% dengan kehilangan (loss) mencapai 0,0004. Konfigurasi ini yang digunakan dalam klasifikasi kardus pada Gudang PT.XYZ.

Corresponding Author. Email: yeremiaalfasusetyo@uksw.edu ^{2}.

© E-ISSN: 2580-1643.

Copyright © 2024 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan Riset). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. 



ACM Computing Classification System (CCS)



Communication and Mass Media Complete (CMC)

1. Latar Belakang

Manajemen gudang sangat penting untuk memastikan kelancaran pasokan barang dengan baik [1]. Dengan adanya manajemen gudang yang baik, dapat memastikan persediaan dan logistik pada toko agar tidak terjadi kekurangan dengan permintaan konsumen. Alur kerja yang terjadi pada gudang secara garis besar adalah menyimpan barang sebelum dapat didistribusikan ke toko untuk dijual kepada konsumen. Tidak semua barang secara cepat dapat didistribusikan langsung kepada toko. Sebaliknya pendistribusian barang terjadi bila toko memerlukan pemasokan barang tertentu saja. Maka dari itu, diperlukannya tempat penyimpanan pada gudang untuk menyimpan barang yang datang tetapi tidak langsung dikirim kepada toko. Agar tidak terjadi kesalahan diperlukannya pengorganisasian barang dan pencatatan barang yang baik [2]. Pada gudang PT XYZ memiliki sistem informasi pencatatan barang pada gudang untuk mencatat barang yang disimpan ke tempat penyimpanan. Namun terjadi beberapa kasus barang yang ada tempat penyimpanan tidak sesuai dengan apa yang ada sistem dikarenakan oleh *human error*. Maka dari itu diperlukannya validasi pencocokan terhadap barang yang ada di tempat penyimpanan dengan yang ada di sistem. Namun pencocokan barang dengan sistem untuk sebesar tempat penyimpanan memerlukan waktu yang sangat lama jika dilakukan secara manual. PT XYZ memerlukan sistem untuk dapat memvalidasi barang pada penyimpanan dengan sistem secara otomatis.

Fine Grained Image Classification merupakan pendekatan dalam *Deep Learning* yang bertujuan untuk membedakan gambar pada kategori dasar ke dalam sub-kategori berbeda misalnya perbedaan spesies burung atau perbedaan model mobil [3]. kasus yang dihadapi dalam FGIC merupakan kasus yang memiliki *dataset* yang secara visual memiliki kemiripan yang tinggi. Pendekatan yang sering digunakan dalam *Fine Grained Image Classification* melibatkan penggunaan arsitektur Jaringan Saraf yang Mendalam (*Deep Neural Networks*) dan teknik *Transfer Learning* untuk mengekstraksi fitur-fitur yang relevan dari gambar-gambar tersebut. Sehingga memungkinkan model memahami pola dan perbedaan yang halus pada kelas-kelas yang mirip secara visual.

Algoritma *Convolutional Neural Network* merupakan salah satu penerapan pada *Deep Learning* yang digunakan untuk mengolah citra gambar digital [4][5][6]. CNN sangat efektif dalam mengenali pola spasial dan memahami hierarki fitur dari data visual. CNN sangat berhasil dalam menjalankan berbagai tugas pengolahan gambar, seperti klasifikasi gambar, deteksi objek, dan segmentasi gambar. Cara kerja CNN dengan menggunakan *kernel* berisikan bobot yang dapat mengekstrak fitur dari data citra gambar dengan melalui operasi konvolusi [4]. Kardus barang memiliki kemiripan visual yang tinggi pada satu sub-kategori barang. Perbedaan-perbedaan kecil tersebut contohnya ukuran, varian, label dan warna. Untuk dapat membedakan hal tersebut perlunya model klasifikasi dengan tingkat ketelitian yang tinggi untuk dapat membangun sistem validasi barang pada tempat penyimpanan otomatis di PT XYZ.

Penelitian tentang Deteksi Klasifikasi Kendaraan pada Jalan Raya menggunakan Algoritma *Convolutional Neural Network* (CNN) membahas tentang Metode *Convolutional Neural Network* yang dapat mengenali dan mengklasifikasikan Citra Gambar kendaraan. Menggunakan 2 model Uji yaitu model sequential dan model VGG16 dengan memiliki akurasi yang tinggi[4]. Menggunakan 8 kelas dengan total data sebanyak 1406 citra data latih dan 274 citra data uji. Dengan model sequential memiliki akurasi 98,18 % dengan loss mencapai 0,103, sedangkan dengan model VGG16 memiliki akurasi sebesar 99,64 % dengan loss mencapai 0,014. Diperoleh kesimpulan bahwa model VGG16 memiliki akurasi yang lebih baik karena memiliki layer lebih banyak dan sudah dilatih sebelumnya menggunakan *dataset* ImageNet maka akan semakin banyak informasi yang dapat di peroleh dari citra gambar.

Penelitian Selanjutnya yang berjudul Deteksi Penyakit Covid-19 Pada Citra X-Ray dengan Pendekatan *Convolutional Neural Network* (CNN) yang dilakukan oleh Mawaddah Harahap dan kawan-kawan membahas tentang klasifikasi gambar hasil sinar-X pada bagian dada manusia untuk dapat mendiagnosis pasien COVID-19 positif maupun negative [7]. Dalam penelitiannya menggunakan arsitektur *Convolutional Neural Network* dengan menggunakan metode *Transfer Learning* dimana menggunakan arsitektur penelitian sebelumnya dengan kasus yang mirip untuk mengekstrak bobot dalam lapisan

tersembunyi di dalam model yang dapat dimodifikasi dan diupdate sesuai masalah dalam penelitian. Penelitian ini juga memberikan hasil pelatihan menggunakan arsitektur-arsitektur terkenal sebelumnya seperti VGG19, ResNet, ResNetV2, MobileNetV2, dan InceptionResNetV2. Dari hasil yang diberikan dengan menggunakan *dataset* gambar Sinar-X tubuh manusia untuk mendiagnosis COVID-19 diraih oleh arsitektur ResNet50V2 dengan akurasi mencapai 96 persen. Namun dikatakan bahwa dalam penelitian ini masih memerlukan data lebih banyak untuk menghindari *overfitting*.

Lalu Penelitian berjudul *Fine-Grained Image Analysis With Deep Learning: A Survey* menjelaskan tentang perbedaan analisis gambar pada *Computer Vision* dengan pendekatan *Fine Grained* [8]. pada pengenalan gambar umumnya yang diberikan contoh bahwa komputer dapat mengenali dan membedakan suatu objek dengan kategori dasar dengan perbedaan yang cukup besar sehingga komputer dapat mengenali dengan mudah. Lalu dijelaskan bahwa *Fine-Grained Image Analysis* ini merupakan pendekatan yang dilakukan untuk mengenali objek pada satu kategori dasar ke kategori yang lebih kecil. Tantangan yang dihadapi pada *Fine Grained* adalah komputer mampu mengenali dan membedakan objek yang memiliki kesamaan yang tinggi pada visualnya. Masalah utama yang dihadapi adalah komputer sulit sekali mengenali dengan sedikit perbedaan sehingga dengan algoritma klasifikasi biasanya tidak mampu membedakan dan memiliki ketepatan prediksi yang rendah. Maka diperlukannya metode-metode seperti menangkap bagian objek yang berbeda. Menemukan informasi kasar hingga halus pada gambar dan juga mengembangkan model yang berbasis atau khusus pada atribut yang dipakai.

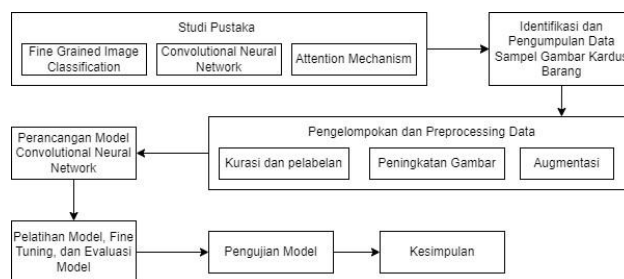
Dalam penelitian Dichao Liu, Yu Wang Kenji Mase and Jien Kato tahun 2022 tentang penerapan *Fine-Grained Image Classification* dengan menggunakan metode *Multi-Scale Attention* dan juga *Channel Spatial Attention* mengatakan bahwa pendekatan *Attention Information* atau perhatian informasi pada objek merupakan pendekatan yang efektif untuk penerapan *Fine Grained Image Classification* [3]. Cara kerjanya dengan memberikan perhatian dan melokalisasi pada bagian pada objek sehingga jika digabungkan dan diklasifikasikan mampu memberikan informasi

pelengkap dan komprehensif. Dijelaskan bahwa metode yang dengan memperoleh fitur ekstraksi dari arsitektur ResNet50 lalu mengeksploitasi pada lapisan *fully-connected*. Dari hasil fitur ekstraksi tersebut akan dilakukan 2 operasi yaitu *Multi-Scale Attention* dan *Channel Spatial Attention* yang nantinya akan digabungkan dengan fitur ekstraksi awal sehingga fitur memiliki semacam Atensi pada bagiannya. Operasi ini disebut *Attention* karena dengan menggabungkan operasi atensi dengan fitur dengan operasi penjumlahan matriks sehingga pada bagian tertentu akan menambah bobot sehingga model akan dapat fokus pada area tersebut.

Penelitian-penelitian diatas melakukan beberapa inovasi dalam mengerjakan tugas klasifikasi gambar dengan metode *Deep Learning* dengan studi kasus yang berbeda-beda. Beberapa teknik dilakukan agar dapat memberikan potensi maksimal dan hasil yang memuaskan. Sama halnya dengan penelitian ini memiliki tujuan untuk dapat mengklasifikasikan kardus barang di PT.XYZ. Melalui pendekatan *Fine Grained* dan teknik *Transfer Learning* serta melakukan beberapa konfigurasi berbeda agar dapat dibandingkan yang nantinya dapat diketahui hasil konfigurasi yang maksimal untuk dapat digunakan di PT.XYZ.

2. Metode Penelitian

Tahapan Metode Penelitian yang akan dilakukan adalah sebagai berikut.

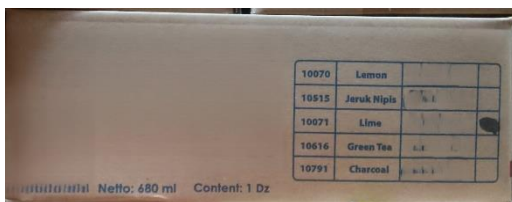


Gambar 1. Tahapan Penelitian

Persiapan dalam penelitian ini adalah studi pustaka dengan mencari referensi mengenai topik pembahasan *Fine Grained Image Classification* dalam jurnal-jurnal nasional maupun internasional, artikel beserta situs internet yang berhubungan. Pada tahap ini dilakukan pencarian *tools* untuk klasifikasi gambar

yang mungkin bisa digunakan untuk menunjang penelitian dan tahap berikutnya. Mempelajari fitur-fitur dalam *tools* dengan melakukan demo klasifikasi gambar dengan *dataset* yang sudah ada di internet untuk menentukan model yang digunakan dalam penelitian. Penekanan utama dalam penelitian ini adalah meningkatkan kemampuan sistem untuk mengenali objek pada kardus dengan tingkat detail yang tinggi, memungkinkan identifikasi yang lebih akurat dan efisien. Mengidentifikasi ruang lingkup pembahasan dengan terjun langsung di lapangan untuk menemukan masalah yang muncul untuk dibahas. Menentukan data-data apa saja yang perlu diambil lalu menentukan model arsitektur klasifikasi gambar untuk klasifikasi kardus dengan dasar *Fine Grained Image Classification*.

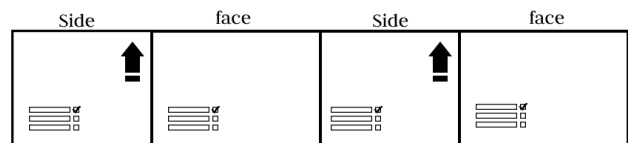
Data yang digunakan berupa citra gambar kardus barang. Populasi data merupakan kardus yang terdapat pada Gudang PT XYZ. Dengan sampel data adalah gambar kardus yang dapat berhasil diperoleh. Lalu untuk metode pengumpulan data menggunakan metode *Purposive Sampling*. Metode ini memiliki karakteristik sampel disesuaikan dengan tujuan penelitian [9]. Data yang diambil memiliki kriteria tertentu seperti memiliki kesamaan dengan tingkat tinggi untuk dapat diklasifikasikan. Ketentuan gambar yang diambil merupakan bagian depan, samping kardus yang berisikan informasi sebuah kardus seperti nama, varian, rasa, jenis barang, produksi dan lain-lain.



Gambar 2. Sampel data kardus bagian depan

Tahap Pengelompokan dan Praproses data (*Preprocessing Data*) dilakukan pengelompokan data yang berhasil diambil. Data yang diperoleh akan dipilih sesuai ketentuan awal. Dari data tersebut dilakukan pengelompokan sesuai dengan informasi yang tercantum pada gambar dengan urutan nama produk lalu dilanjutkan dengan varian dan juga ukuran. Kardus memiliki beberapa sisi layaknya balok yang memiliki 6 sisi, data yang diperoleh merupakan data 2 sisi panjang dan 2 sisi lebar. Maka dari satu kardus di bagi menjadi 2 kelas yaitu sisi depan atau

face dan sisi samping atau *side*. Dari pengelompokan sesuai informasi dikelompokkan lagi menjadi *face* dan *side* dengan ketentuan bila pada sisi tersebut memiliki informasi pembeda dengan sub-kategori lain yang hampir sama. Bila tidak memiliki pembeda seperti logo, dan nama produk tidak dimasukan kedalam *dataset*. Tujuan ini bermaksud untuk membedakan antara kardus sisi depan dan samping walaupun merupakan 1 kardus yang sama agar komputer lebih mudah membedakan.



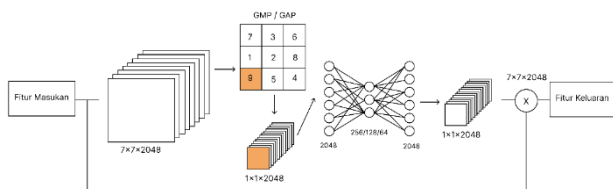
Gambar 3. Kerangka kardus

Data yang berhasil dikelompokkan dan juga diberi label data tersebut akan dilakukan praproses data dengan melakukan beberapa proses seperti peningkatan kualitas gambar, pemotongan gambar, mengubah ukuran, mengatur warna dan lain-lain. Lalu dilakukan augmentasi gambar untuk memperbanyak *dataset* gambar dengan membuat salinan gambar serta memanipulasi transformasi gambar [10]. Diproses dengan beberapa kombinasi seperti rotasi, *mirroring*, dan mengubah kecerahan pada gambar sehingga menjadi data baru yang berbeda dengan data aslinya. Setelah data berhasil dikumpulkan dan menjadi *dataset* yang dapat untuk diuji. Data akan dibagi menjadi 2 bagian yaitu data latih yang digunakan untuk melatih model dan data validasi untuk memvalidasi hasil dari model tersebut.

Dalam rancangan model *Convolutional Neural Network*. Model CNN ini mengimplementasikan beberapa kovolusi yang menghasilkan fitur ekstraksi ketika data citra gambar diproses sebagai masukan. Dari fitur ekstraksi tersebut dapat tersambung dengan lapisan disebut dengan *fully-connected* yang bertindak sebagai pengklasifikasi sesuai dengan kelas yang telah tertera. Pembuatan model arsitektur CNN menggunakan metode *Transfer Learning* dimana metode ini menggunakan pembelajaran sebelumnya, menggunakan kerangka model dari pembelajaran dengan masalah yang hampir mirip lalu memodifikasi dan mengupdate paramater sesuai dengan permasalahan dalam penelitian [11]. Jenis *Transfer Learning* yang digunakan ialah ResNet dan untuk penerapan FGIC menggunakan kerangka RMCSAM.

Residual Network (ResNet) merupakan jenis jaringan saraf tiruan (JST) menggunakan operasi konvolusi yang memungkinkan pelatihan dengan lapisan lebih dalam yang termasuk ke dalam *Deep Neural Network*. ResNet memiliki mekanisme residu yaitu menghubungkan masukan langsung dengan lapisan berikutnya tanpa melalui pemrosesan pada lapisan tersebut. Koneksi ini bertujuan untuk mengurangi terjadinya gradien yang hilang dan peledakan gradien yang dapat ditemui dalam jaringan saraf tiruan yang memiliki lapisan yang sangat dalam. ResNet merupakan salah satu arsitektur terkenal yang hingga saat ini masih digunakan karena memiliki akurasi yang terus bertambah seiring bertambahnya penelitian dengan memodifikasinya [12].

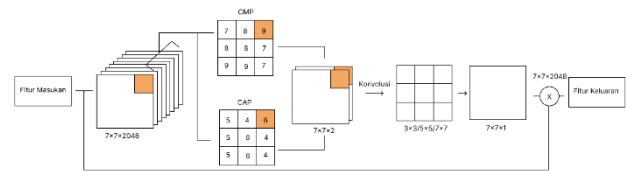
Recursive Multi-Scale Channel-Spatial Attention (RMCSAM) memanfaatkan fitur ekstraksi model CNN dengan menambahkan beberapa operasi pemrosesan nantinya pada lapisan sebelum *fully-connected* pada model ResNet yang dipakai. Fitur ekstraksi setelah melalui pemrosesan pada Arsitektur Resnet berupa citra gambar berukuran $7 \times 7 \times 2048$ dibagi menjadi 2 yaitu proses *Channel Wise Attention* dan *Spatial Wise Attention*. *Channel Wise Attention* memiliki cara kerja dengan memproses fitur ekstraksi dengan lapisan *Global Pooling* berdasarkan rata-rata (GAP) atau nilai maksimal (GMP) pada satu *kernel*, maka fitur ekstraksi memiliki ukuran menjadi $1 \times 1 \times 2048$ diubah menjadi lapisan *linier* menggunakan *flatten* berubah menjadi 2048 *neuron*. Lapisan dilakukan operasi *multilayer perceptron* (MLP) dengan satu *hidden layer* dengan neuron 256/126/64 dan fungsi aktivasi berupa *Sigmoid*.



Gambar 4. Alur *Channel Wise Attention*

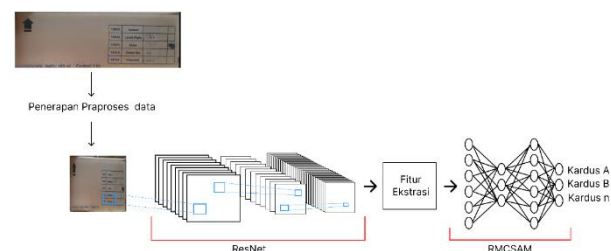
Sub-modul *Spatial Wise Attention* merubah fitur ekstraksi dengan memberikan operasi *Channel Pooling* berdasarkan channel pada fitur ekstraksi dengan hasil ukuran menjadi $7 \times 7 \times 1$. Operasi *pooling* sama dengan proses *Channel Wise Attention* dengan memperoleh rata-rata dan nilai maksimal dari setiap indeks *pooling*. Fitur ekstraksi rata-rata (CAP) dan nilai maksimal (CMP) digabungkan (*concatenate*) berdasarkan sumbu

channel menjadi ukuran $7 \times 7 \times 2$. Lalu dilakukan operasi konvolusi dengan ukuran kernel $3 \times 3 / 5 \times 5 / 7 \times 7$ dan melalui fungsi aktivasi Sigmoid. Hasil fitur ekstraksi *Channel Wise Attention* dan *Spatial Wise Attention* masing-masing dilakukan operasi perkalian matriks dengan fitur ekstraksi awal lalu ditambahkan. Maka akan diberikan keluaran berukuran $7 \times 7 \times 2048$.



Gambar 5. Alur *Spatial Wise Attention*

Proses penggabungan dari fitur ekstraksi dengan sub-modul *Channel Wise Attention* dan *Spatial Wise Attention* ini merupakan teknik modul atensi yang merupakan salah satu cara dalam pendekatan FGIC. Setelah fitur keluaran diperoleh dengan ukuran $7 \times 7 \times 2048$, fitur dioperasikan dengan lapisan *Average Pooling* dengan menghitung rata-rata setiap *channel* sehingga ukuran fitur menjadi $1 \times 1 \times 2048$. Lalu masuk ke dalam *fully-connected* menggunakan *flatten* dan dilakukan pengklasifikasian.



Gambar 6. Alur Model

Tahap pelatihan merupakan tahap dimana arsitektur model yang sudah dibuat dilakukan pelatihan dimana model CNN melakukan serangkaian prosedur dimana gambar masuk ke dalam model dan keluaran berupa prediksi. Dari prediksi tersebut disandingkan dengan label pada data. Nilai selisih label dengan prediksi tersebut merupakan nilai *loss* atau kehilangan yang nantinya diproses dengan *optimizer* untuk mengupdate bobot pada model dalam proses *backpropagation*. Evaluasi model setelah melalui tahap pelatihan merupakan tahap dimana model yang sudah dibuat diberikan evaluasi terhadap akurasi prediksi model dengan label data yang benar. Perhitungan akurasi dapat dinyatakan sebagai berikut.

$$\text{Akurasi} = \frac{\text{jumlah prediksi benar}}{\text{jumlah keseluruhan prediksi}} \times 100\%$$

Penekanan Akurasi yang tinggi merupakan fokus utama dari penelitian ini, kecepatan prediksi model dan optimasi ukuran model bukan merupakan fokus penelitian ini. Lingkungan penelitian ialah rak penyimpanan pada gudang PT. XYZ.

3. Hasil dan Pembahasan

Tahap persiapan dilakukan dengan mengumpulkan alat yang perlu digunakan dalam penelitian. Proses pengambilan data menggunakan kamera ponsel dan alat untuk mencatat data berupa Microsoft Excel. Setelah data gambar diperoleh maka informasi pada gambar dicatat di dalam Microsoft Excel dan diberi kode kelas kardus tersebut. *Tools* yang digunakan untuk augmentasi data menggunakan Torchvision. Untuk melatih model menggunakan pustaka pytorch. Dan menggunakan Google Collab sebagai sarana pembuatan aplikasi dengan basis bahasa python.

Pada proses pengambilan data, kriteria kardus yang diambil ialah kardus yang dinilai memiliki kemiripan dengan varian kardus yang lain pada satu kelas besar dan melalui ketentuan-ketentuan kurasi data sebelumnya maka diperoleh data kelas dengan sub-kelas sebagai berikut.

Tabel 1. Data kelas pada kategori

Kategori	Kelas (Sub-kategori)
Kategori A	7, 8
Kategori B	22, 23, 24, 25
Kategori C	36, 38
Kategori D	62, 63, 67

Pada tabel 1 merupakan perolehan kelas dan juga pengelompokan kelas terhadap kategorinya. Kategori tersebut merupakan nama/jenis barang secara keseluruhan dan sub-kategori merupakan varian atau ukuran dari nama barang tersebut. Contohnya kategori sabun cuci piring dengan varian jeruk nipis 800 ml. Lalu setelah pengelompokan berdasarkan kategori dan sub-kategori, data dipisah lagi menjadi satu gambar per kardus dengan perbedaan kardus bagian muka (*face*) dan bagian samping (*side*). maka diperoleh data dengan 14 kelas dengan total data sebanyak 172 gambar.

Tahap berikutnya ialah dengan melakukan augmentasi data menggunakan fungsi *transformation* dalam pustaka torchvision dengan kombinasi dari beberapa konfigurasi seperti pengaturan kecerahan dengan maksimal kecerahan 50 persen, rotasi gambar secara acak dengan maksimal 25 derajat putar, membalik gambar secara horizontal maupun vertikal dengan masing masing probabilitas 0.5 sebanyak 10 kali dan 20 kali penambahan gambar pada setiap kelas.

Dilakukan praproses data dengan kustomisasi *dataset* menggunakan pustaka torchvision untuk mengubah gambar yang sudah diaugmentasi menjadi ukuran 224x224x3. Lalu data juga dibagi menjadi data latih dan juga data validasi. Dalam pembagian data menggunakan metode *stratified sampling* dimana membagi rata secara komposisi pada setiap kelas yang disebut strata[13] ke dalam data latih dan validasi secara acak dengan perbandingan 3:1 maka dengan konfigurasi diatas dapat diperoleh *dataset* siap uji sebagai berikut.

Tabel 2. Rincian data augmentasi

Data	Augmentasi x 10	Augmentasi x 20
Latih	1290	2580
Validasi	430	860
Jumlah	1720	3440

Pembuatan model CNN menggunakan ResNet yang dimodifikasi dengan RMCSAM dibuat menggunakan pustaka pytorch berbasis python. Model dirancang berdasarkan arsitektur sesuai dalam artikel RMCSAM dengan sedikit penyesuaian seperti terdapat 14 kelas pada lapisan klasifikasi. Model yang sudah terbuat dapat dilatih dengan beberapa konfigurasi pelatihan. Dalam penelitian terdapat beberapa kali pelatihan dilakukan dengan konfigurasi berbeda-beda yang bertujuan untuk membandingkan akurasi dan model terbaik untuk kasus dengan *dataset* kardus. Perbedaan tersebut dengan mengombinasikan antara penggunaan arsitektur ResNet dan ResNet+RMCSAM, *dataset* Augmentasi x10 dan Augmentasi x20, lalu dengan menginisialisasi bobot pada arsitektur dengan bobot yang pernah dilatih (*pretrain*) dengan *dataset* pada ImageNet dan inisialisasi secara random. Dari kombinasi tersebut dapat diberikan konfigurasi pelatihan sebagai berikut.

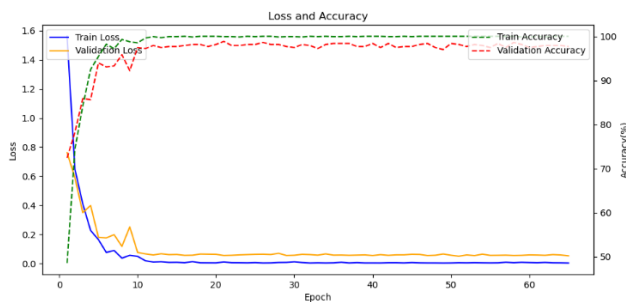
Tabel 3. Konfigurasi model

Arsitektur	Dataset	Bobot
ResNet	Augmentasi x10	Random
		ImageNet
	Augmentasi x20	Random
		ImageNet
RMCSAM	Augmentasi x10	Random
		ImageNet
	Augmentasi x20	Random
		ImageNet

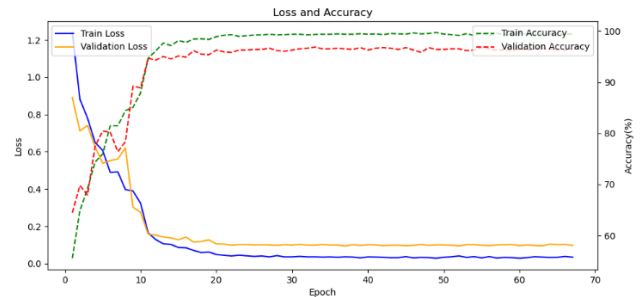
Tabel 3 merupakan rincian konfigurasi setiap model yang akan dilatih. Nama konfigurasi disesuaikan dengan kombinasi pada kolom tabel diatas, contohnya ResNetx10_pretrain adalah konfigurasi dengan arsitektur ResNet lalu dataset augmentasi x10 serta dengan menggunakan inisialisasi bobot ImageNet. Semua konfigurasi menggunakan pengoptimalan(optimizer) Adam dengan *Learning Rate* sebesar 0,0005. Lalu *loss function* menggunakan *CrossEntropyLoss*. Dilatih sampai dengan *Epoch* ke-60, *batch size* 8, *Dropout Rate* sebesar 0,2 dan untuk *Flatten* pada konfigurasi ResNet Menggunakan *Global Average Pooling*. Dengan konfigurasi diatas maka hasil pelatihan dapat didapat seperti berikut.



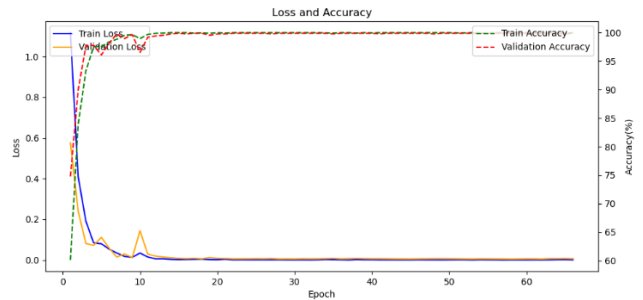
Grafik 7. Pelatihan ResNetx10



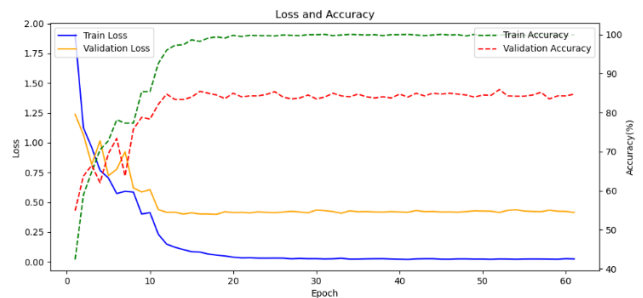
Gambar 8. Pelatihan Resnetx10_pretrain



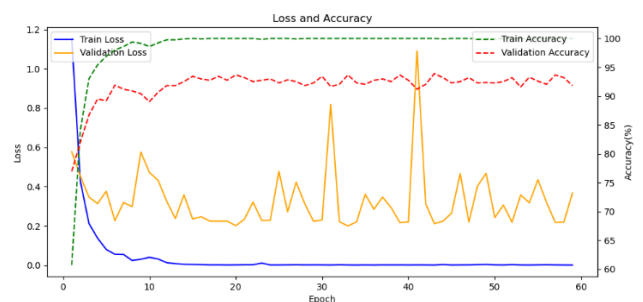
Gambar 9. Pelatihan ResNetx20



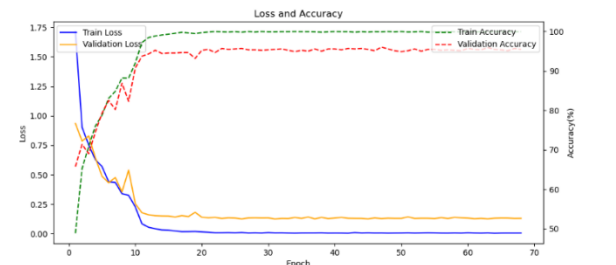
Gambar 10. Pelatihan ResNetx20_pretrain



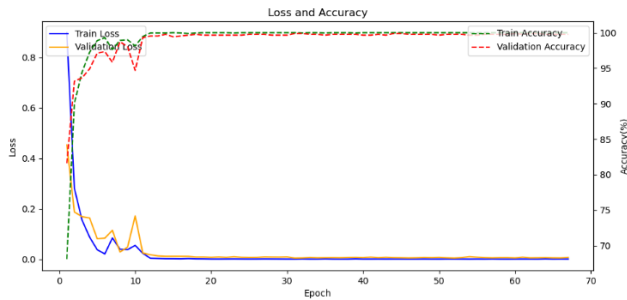
Gambar 11. Pelatihan RMCSAMx10



Gambar 12. Pelatihan RMCSAMx10_pretrain



Gambar 13. Pelatihan RMCSAMx



Gambar 14. Pelatihan RMCSAMx20_pretrain

Kumpulan gambar diatas merupakan semua grafik kehilangan (*loss*) dan akurasi pada semua model dengan konfigurasi sebelumnya. Perbandingan dapat dilihat ketika inisialisasi bobot secara random mengakibatkan *overfitting* dimana model kesusahan

atau menurun performanya ketika dibuat prediksi menggunakan data baru [14]. Lalu terlihat juga bahwa model yang menggunakan *dataset* Augmentasi x20 terlihat memiliki akurasi yang tinggi dibandingkan dengan model menggunakan Augmentasi x10. Untuk performa arsitektur ResNet lebih optimal daripada memodifikasi ResNet beserta RMCSAM. Dapat dilihat juga pada konfigurasi RMCSAMx10_pretrain model mengalami kebingungan ditandai dengan ketidakstabilan kehilangan(*loss*) validasi dan *overfitting*, walaupun memiliki akurasi yang tinggi. Namun, untuk konfigurasi ResNetx20_pretrain dengan RMCSAMx20_pretrain memiliki performa hampir identik dengan akurasi mencapai 99 persen.

Tabel 4. Hasil pelatihan

Konfigurasi	Kehilangan (<i>loss</i>)	Akurasi validasi (%)
ResNetx10	0,1145	87,19
ResNetx10 _Pretrain	0,0027	97,69
ResNetx20	0,0345	96,53
ResNetx20 _Pretrain	0,0012	99,88
RMCSAMx10	0,0243	84,72
RMCSAMx10 _Pretrain	0,0007	91,82
RMCSAMx20	0,0061	95,49
RMCSAMx20 _Pretrain	0,0004	99,77

Tabel 5 menunjukkan hasil akhir dari pelatihan model yang dilakukan dalam penelitian ini. Model yang menggunakan arsitektur *Recursive Multi-Scale Channel-Spatial Attention* (RMCSAM) umumnya menunjukkan kehilangan (*loss*) rata-rata yang lebih rendah dibandingkan dengan model yang menggunakan arsitektur ResNet, dengan pengecualian pada model RMCSAMx10. Hal ini menunjukkan bahwa arsitektur RMCSAM memiliki kecepatan belajar yang lebih baik dalam memprediksi data dengan menggunakan dataset yang terbatas. Kehilangan terendah dicapai oleh konfigurasi RMCSAMx20_pretrain, dengan nilai *loss* mencapai 0,0004. Selain itu, jumlah dataset yang digunakan juga memiliki dampak signifikan terhadap hasil pelatihan. Penambahan data hasil augmentasi sebanyak dua kali lipat dari dataset asli dapat meningkatkan akurasi

hingga 9% dalam klasifikasi dataset kardus. Pengujian dilakukan dengan menggunakan dataset yang sama untuk semua model. Dataset ini diperoleh sebelum dilakukan proses augmentasi gambar, yang memungkinkan evaluasi performa model dalam menghadapi data baru yang belum pernah dilihat oleh model sebelumnya. Dengan cara ini, dapat dipastikan bahwa model mampu mempertahankan akurasi yang tinggi meskipun dihadapkan dengan data yang tidak dikenal. Analisis ini juga menekankan pentingnya proses augmentasi dalam meningkatkan generalisasi model, yang merupakan faktor krusial dalam pengembangan model pembelajaran mendalam yang efektif. Penggunaan teknik augmentasi yang tepat dapat memperluas cakupan data latih dan mengurangi risiko *overfitting* pada model.

Tabel 5. Akurasi hasil pengujian

Konfigurasi	Akurasi Uji (%)
ResNetx10	87,21
ResNetx10_Pretrain	97,67
ResNetx20	96,51
ResNetx20_Pretrain	97,67
RMCSAMx10	90,12
RMCSAMx10_Pretrain	90,12
RMCSAMx20	96,51
RMCSAMx20_Pretrain	99,42

Dalam Tabel hasil pengujian terlihat RMCSAM lebih unggul dibandingkan dengan ResNet dengan menggunakan data yang belum pernah diprediksi. Terdapat juga perbedaan hasil latih dengan hasil uji. ResNet10 memiliki akurasi paling rendah di angka 87.21%, hasil ini hampir sama terjadi pada pelatihan yaitu 87.19%. RMCSAMx10 memiliki peningkatan yang tadinya memiliki akurasi 84,72% naik hingga 90,12%. Capaian tertinggi terlihat pada RMCSAMx20_pretrain yang memiliki akurasi sebesar 99,42%.

Hampir semua model RMCSAM mengalami kenaikan akurasi pada saat tahap uji sedangkan beberapa model ResNet mengalami penurunan akurasi. Dikatakan bahwa arsitektur ResNet dapat mengalami masalah yang rumit ketika terdapat beberapa poin perbedaan kondisi antara data pelatihan dan pengujian yang menimbulkan *Overfitting*. Beberapa cara dilakukan dengan melakukan modifikasi struktur dasar, tetapi masih mempertahankan ukuran model untuk dapat dikelola[15]. RMCSAM yang menggunakan teknik Atensi dengan mengambil hasil ekstraksi dari ResNet yang merupakan salah satu teknik dalam pendekatan FGIC terbukti dalam penggunaan pada *dataset* kardus yang memiliki kemiripan visual tinggi yang memerlukan model klasifikasi khusus untuk dapat memberikan akurasi yang tinggi dan performa yang diinginkan.

4. Kesimpulan

Penelitian ini melakukan berbagai proses dalam membuat model klasifikasi kardus pada PT. XYZ. Pengumpulan data dengan menggunakan teknik *Purposive Sampling* dengan mengambil data dengan

ketentuan tertentu perlu pemahaman yang lebih dalam wawasan terhadap data yang akan di ambil dalam kasus ini adalah kardus barang yang ada pada gudang PT.XYZ. Penentuan banyaknya kelas didasarkan pada jumlah banyaknya jenis produk dalam gudang. Dalam setiap kardus tidak semua sisi kardus memiliki pembeda dengan varian lainnya dalam satu kategori produk sehingga lebih teliti dalam penentuan pelabelan dan pemisahan kelas ke ruang lingkup yang lebih kecil untuk mempermudah model dalam mengklasifikasi.

Pemilihan teknik pengambilan data menggunakan teknik *Purposive Sampling* merupakan teknik paling tepat dalam penelitian ini. Akan tetapi, banyaknya data dan jenis kardus yang terlalu banyak mengakibatkan pengambilan data memakan waktu yang cukup banyak. Selain itu, terdapat juga kemungkinan penambahan produk jenis baru dan juga produk yang sudah tidak diproduksi lagi menambah masalah dalam penelitian ini.

Penggunaan model arsitektur dengan *Transfer Learning* model ResNet dengan penggabungan menggunakan RMCSAM diusulkan dalam mengklasifikasikan gambar kardus. Mengombinasikan konfigurasi antara pemilihan model, *dataset*, dan bobot dilakukan untuk memperoleh hasil yang diinginkan. Berdasarkan hasil pengujian penggunaan model ResNet yang digabung dengan RMCSAM yang diinisialisasi bobot ImageNet menggunakan *dataset* Augmentasi x20 lebih unggul untuk mengklasifikasikan kardus memiliki akurasi mencapai 99,42%. Disusul dengan model menggunakan konfigurasi ResNet *dataset* Augmentasi x20 dan augmentasi x10 dengan inisialisasi keduanya menggunakan ImageNet memiliki akurasi sebesar 97,67%. Keduanya memiliki performa yang sama hanya dengan perbedaan *dataset* yang digunakan.

Menandakan dengan pendekatan FGIC model dapat menghasilkan akurasi lebih tinggi daripada menggunakan model klasifikasi konvensional. Lalu untuk model dengan inisialisasi bobot secara random dipegang oleh 2 model yaitu dengan konfigurasi ResNetx20 dan RMCSAMx20 dengan akurasi hampir sama sebesar 96,51%.

Pemilihan model pengujian menggunakan pendekatan FGIC sudah terbukti lebih baik menggunakan arsitektur RMCSAM. Akan tetapi, RMCSAM masih memiliki beberapa kekurangan antara lain penambahan beberapa sub-modul ke dalam model serta dilakukan secara rekursif mengakibatkan model semakin dalam dan besar yang akan mempengaruhi kecepatan prediksi pada pelatihan maupun pengujian tergantung pada model fitur ekstraksi yang digunakan. Arsitektur lain masih perlu dipertimbangkan untuk penelitian di masa berikutnya.

5. Daftar Pustaka

- [1] Rahardjo, B. (2017). Perancangan sistem manajemen gudang material penunjang di PT XYZ. *J@ ti Undip: Jurnal Teknik Industri*, 12(2), 127-136.
- [2] Senduk, H. Y., & Sitokdana, M. N. (2022). Perancangan Sistem Informasi Pencatatan Gudang Berbasis Website (Studi Kasus Slingbag Salatiga). *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(1), 373-383. DOI: <https://doi.org/10.35957/jatisi.v9i1.1503>.
- [3] Liu, D., Wang, Y., Mase, K., & Kato, J. (2022). Recursive multi-scale channel-spatial attention for fine-grained image classification. *IEICE TRANSACTIONS on Information and Systems*, 105(3), 713-726. DOI: 10.1587/transinf.2021EDP7166.
- [4] Radikto, R., Mulyana, D. I., Rofik, M. A., & Zakaria, M. O. Z. (2022). Klasifikasi Kendaraan pada Jalan Raya menggunakan Algoritma Convolutional Neural Network (CNN). *Jurnal Pendidikan Tambusai*, 6(1), 1668-1679. DOI: <https://doi.org/10.31004/jptam.v6i1.3179>.
- [5] Sholeh, I., & Wiyono, D. A. (2023). Model Klasifikasi Terumbu Karang Menggunakan Metode Convolutional Neural Network. *DoubleClick: Journal of Computer and Information Technology*, 6(2), 77-81. DOI: <http://doi.org/10.25273/doubleclick.v6i2.13901>.
- [6] Nata, I. D., Sulistianingsih, N., & Yusuf, S. A. A. (2023). Implementasi Deep Learning Dalam Klasifikasi Citra Gambar Dengan Menggunakan Metode CNN.
- [7] Harahap, M., Laia, E. M., Sitanggang, L. S., Sinaga, M., Sihombing, D. F., & Husein, A. M. (2022). Deteksi penyakit covid-19 pada citra x-ray dengan pendekatan convolutional neural network (cnn). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(1), 70-77. DOI: <https://doi.org/10.29207/resti.v6i1.3373>.
- [8] Wei, X. S., Song, Y. Z., Mac Aodha, O., Wu, J., Peng, Y., Tang, J., ... & Belongie, S. (2021). Fine-grained image analysis with deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(12), 8927-8948. DOI: 10.1109/TPAMI.2021.3126648.
- [9] Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian journal of psychological medicine*, 43(1), 86-88. DOI: <https://doi.org/10.1177/0253717620977000>.
- [10] Sanjaya, J., & Ayub, M. (2020). Augmentasi Data Pengenalan Citra Mobil Menggunakan Pendekatan Random Crop, Rotate, dan Mixup. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(2). DOI: <https://doi.org/10.28932/jutisi.v6i2.2688>.
- [11] Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(5S4), 427-432.

- [12] Wightman, R., Touvron, H., & Jégou, H. (2021). Resnet strikes back: An improved training procedure in timm. *arXiv preprint arXiv:2110.00476*.
- [13] Arieska, P. K., & Herdiani, N. (2018). Pemilihan teknik sampling berdasarkan perhitungan efisiensi relatif. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 6(2). DOI: <https://doi.org/10.26714/jsunimus.6.2.2018.%25p>.
- [14] Zuo, X., Yang, X., Dou, Z., & Wen, J. R. (2019). RUCIR at TREC 2019: Conversational Assistance Track. In *TREC*.
- [15] Zhou, T., Zhao, Y., & Wu, J. (2021, January). Resnext and res2net structures for speaker verification. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 301-307). IEEE. DOI: 10.1109/SLT48900.2021.9383531.