

# Pengembangan Model Pendeteksian Gambar Alat Musik dengan Metode Faster R-CNN dengan Library Keras

Ivan Septian<sup>1)</sup>, Henri Septanto<sup>2)</sup>

Infomatika, Fakultas Industri Kreatif, Institut Teknologi dan Bisnis Kalbis  
Jalan Pulomas Selatan Kav. 22, Jakarta 13210

<sup>1)</sup>Email: 2016102301@student.kalbis.ac.id

<sup>2)</sup>Email:henri.septanto@kalbis.ac.id

**Abstract:** This research discusses developing a musical instrument image detection application with a *frCNN* library method. The purpose of this study is to detect types of musical instruments using the fastest R-CNN as a method of detecting objects. The problem with using RCNN is the length of time of computing. It takes about a minute to process the image data, so the training process will take a very long time. Therefore, researchers use the fastest r-cnn method to get the output quickly.

**Keywords:** Faster R-CNN, deep learning, computer vision, machine learning, object detection

**Abstrak:** Penelitian ini membahas mengenai pengembangan aplikasi deteksi gambar alat musik dengan metode keras *frCNN*. Tujuan dari penelitian ini untuk mendeteksi jenis alat musik dengan menggunakan *faster r-cnn* sebagai metode deteksi objek. Masalah dalam menggunakan *rcnn* adalah lamanya waktu komputasi. Untuk memproses satu data gambar dibutuhkan waktu sekitar satu menit, sehingga untuk melakukan proses training akan membutuhkan waktu yang sangat lama. Oleh sebab itu, metode *faster r-cnn* digunakan pada penelitian ini untuk memperoleh output dengan cepat.

**Kata kunci:** Faster R-CNN, deep learning, computer vision, machine learning, object detection

## I. PENDAHULUAN

Alat musik sudah ditemukan sejak jaman nenek moyang dan tidak dapat diketahui pasti kapan alat musik ditemukan. Sejarah musik sangat berikatan erat dengan kebudayaan manusia. Disuatu bangsa yang memiliki kebudayaan yang maju dipastikan memiliki seni musik yang maju juga. Dimasa modern ini, perkembangan musik telah maju dengan sangat pesat. Hal ini disebabkan oleh berkembangnya teknologi dan ilmu pengetahuan. Karena alat musik sudah ada sejak jaman nenek moyang maka dapat dibayangkan banyaknya jenis alat musik yang ada diseluruh dunia. Tetapi dari banyaknya alat musik yang ada, hanya ada beberapa nama alat musik yang populer, sehingga banyak alat musik tidak dikenali oleh

masyarakat [1]. Hal ini menyebabkan sulitnya untuk mencari informasi tentang alat musik tersebut karena banyak yang tidak mengetahui nama alat musik tersebut. Oleh karena itu dibuatlah aplikasi yang dapat mengetahui jenis alat musik tersebut dari gambarnya. Dengan aplikasi ini pengguna hanya perlu mengambil gambar alat musik yang ingin diketahui kemudian dimasukkan kedalam aplikasi dan aplikasi akan memberikan informasi berupa nama alat musik tersebut.

*Computer Vision* merupakan ilmu yang memiliki tujuan untuk membuat komputer dapat meniru kemampuan manusia dalam melihat benda – benda. Untuk mencapai tujuan tersebut, para peneliti *Computer Vision* menggunakan berbagai macam metode

agar komputer dapat memperoleh, memahami, menganalisis, dan mengolah data visual seperti gambar dan video. Implementasi Computer Vision dalam kehidupan sehari – hari dapat dilihat pada kamera smartphone yang dapat menangkap wajah dan menggunakannya sebagai kunci layar, aplikasi manipulasi gambar seperti photoshop, kamera digital yang dapat meningkatkan kualitas gambar, dan masih banyak lagi. Selain implementasi pada kehidupan sehari – hari Computer Vision juga digunakan diberbagai macam bidang pendidikan seperti bidang medis, bidang industri, bidang matematika, bidang Artificial Intelligent, dan bidang otomotif.

Berdasarkan hasil observasi terdapat beberapa metode pendeteksian objek seperti CNN, YOLO, MLP, dan R-CNN. Pada penelitian ini akan diputuskan untuk menggunakan metode Faster R-CNN. Alasan menggunakan metode ini karena Faster R-CNN merupakan pengembangan dari metode R-CNN yang dapat mengolah gambar dengan lebih cepat dibandingkan dengan metode sebelumnya. Dataset yang digunakan pada penelitian ini menggunakan dataset People Playing Musical Instrumen (PPMI) yaitu kumpulan aktivitas interaksi manusia dan objek. Dataset ini berisikan gambar – gambar manusia yang berinteraksi dengan dua belas alat musik yang berbeda. Dua belas alat musik itu yaitu bassoon, cello, clarinet, French horn, gitar, harp, recorder, saxophone, terompet, dan biola. Dataset ini sengaja digunakan untuk menolong peneliti dalam mencari berbagai macam gambar alat musik. Selain itu, gambar pada dataset ini sudah dinormalisasi dan dikumpulkan berdasarkan label dengan jumlah dua ratus untuk setiap label.

## II. METODE PENELITIAN

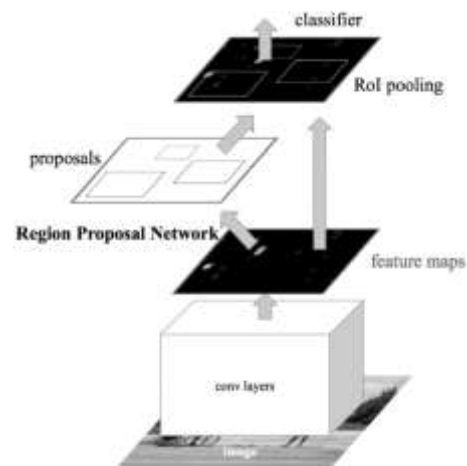
### A. Metode

Penelitian ini menggunakan beberapa teori – teori pendukung yang

berkaitan dalam pengembangan algoritma deteksi gambar alat musik. Guna dari teori – teori ini bertujuan sebagai arahan dalam proses penelitian agar tujuan dari penelitian ini dapat tercapai.

### 1. Faster R-CNN

Faster R-CNN merupakan metode deteksi objek, modifikasi dari Fast R-CNN. Metode ini dibagi menjadi 3 bagian yaitu convolution layer, Region Proposal Network (RPN), dan calsses. Pada tahap pertama, akan melatih filter untuk mendapatkan fitur yang sesuai dari gambar. Sebagai contoh, jika data yang digunakan adalah gambar kucing maka filter akan belajar melalui proses training dari bentuk dan warna yang hanya dimiliki oleh kucing. Tahap selanjutnya, gambar yang telah melalui filter akan dimasukkan kedalam RPN untuk memprediksi bounding box dari gambar tersebut. Setelah melalui tahap pertama dan kedua, selanjutnya akan dilakukan prediksi klasifikasi dari input yang dimasukkan [2].



Gambar 1 Arsitektur Faster R-CNN

### 2. Region Proposal Network

*Region Proposal Network (RPN)* merupakan bagian dari metode *Faster R-CNN* yang digunakan untuk menentukan dimana bagian gambar akan digunakan agar mengurangi lama proses komputasi. RPN dapat dengan cepat dan efisien memindai setiap lokasi pada gambar dan

menentukan wilayah mana yang nanti akan digunakan pada proses selanjutnya. [2]

$$L(p_i, t_i) = (1/N_{cls}) \times \sum L_{cls}(p_i, p_i^*) + (\lambda/N_{reg}) \times \sum p_i^* L_{reg}(t_i, t_i^*)$$

Variable  $i$  merupakan index dari anchor box,  $p$  sebagai nilai probabilitas objek,  $t$  sebagai vector dengan empat parameter yang merupakan lokasi bounding box, \* sebagai Ground-truth.

### 3. Dataset

Dataset yang digunakan dalam penelitian ini diambil dari People Playing Musical Instrument. People Playing Musical Instrument adalah dataset yang digunakan pada penelitian "Grouplet: A Structured Image Representation for Recognizing Human and Object Interactions". Dataset tersebut terdiri dari dua belas jenis alat musik yaitu bassoon, cello, clarinet, erhu, French horn, guitar, harp, recorder, saxophone, trumpet, dan violin. Untuk setiap jenis alat musik dibagi menjadi dua jenis yaitu with instrument dan play instrument yang masing – masing terbagi lagi menjadi kumpulan gambar test dan train. Pada penelitian ini akan digunakan lima jenis alat musik, lima jenis alat musik itu adalah bassoon, French horn, guitar, recorder, dan biola dengan total gambar lima ratus. Alasan menggunakan lima alat musik tersebut kedalam penelitian ini dikarenakan untuk tidak mempersulit model dalam mengenali objek, hal ini karena beberapa gambar alat musik seperti saxophone, terompet, dan french horn memiliki bentuk dan tekstur yang serupa terutama pada bagian mulut alat musik, sehingga harus dipilih mana jenis alat musik yang akan digunakan. Kemudian ditemukan juga gambar – gambar yang memiliki kualitas yang buruk atau tidak layak untuk digunakan sebagai dataset, sehingga didapatkan lima jenis alat musik berdasarkan hasil pertimbangan tersebut.

### 4. ResNet50

ResNet50 merupakan bagian dari Convolutional Neural Network

(CNN) yang memiliki lima puluh layer. Convolutional Neural Network adalah metode yang selalu digunakan dalam kasus deteksi dan klasifikasi gambar. Banyak kasus pengenalan objek lainnya diuntungkan dengan menggunakan metode ini. Lama waktu berlalu menyebabkan kasus deteksi dan klasifikasi menjadi semakin rumit. Namun ketika mendapatkan kasus yang lebih rumit, metode ini menjadi tidak relevan atau tidak mampu menyelesaikan kasus tersebut. Oleh karena itu Resnet atau kependekan dari Residual Learning mencoba untuk menyelesaikan masalah tersebut. Secara umum Convolutional Neural Network memiliki beberapa tumpukan layer yang digunakan dalam melakukan training. Arsitektur ini mempelajari fitur – fitur diakhir dari layer. Dengan menggunakan ResNet, fitur dapat dipahami dengan mudah. Hal ini disebabkan karena ResNet Menggunakan koneksi yang dapat menghubungkan input layer dari layer  $n$  kebeberapa layer lainnya. Hal ini terbukti menjadikan proses training menjadi lebih mudah dibandingkan menggunakan metode yang biasa dan juga mengatasi masalah pengurangan nilai akurasi ketika tanpa menggunakan ResNet. [3]

### 5. Artificial Intelligent

*Artificial Intelligent (AI)* jika didefinisikan oleh para ahli dapat diartikan menjadi berbagai macam definisi. Stuart Russel dan Peter Norvig mengumpulkan defines tersebut kedalam 4 kategori yaitu *Thinking humanly*, *Acting humanly*, *Thinking rationally*, dan *Acting rationally*. Berdasarkan 4 kategori tersebut *Thinking humanly* dan *Acting humanly* adalah definisi dengan arti yang sangat luas, karena sampai saat ini pemikiran manusia yang belum dapat ditiru sepenuhnya kedalam mesin adalah reflex dan intuitif. [4]

### 6. Computer Vision

Computer Vision adalah sebuah penelitian untuk membuat komputer memiliki kemampuan persepsi seperti

halnya manusia dalam memahami lingkungan sekitar, memahami suatu data, mengambil keputusan yang tepat, dan belajar dari pengalaman untuk performa yang lebih baik. Seiring berjalannya waktu, bidang ini telah berkembang dari pemahaman suatu pola dan metode image processing menuju teknik yang lebih rumit dalam memahami gambar seperti model-based dan knowledge-based vision. [5]

### **7. Machine Learning**

*Machine Learning* merupakan pengaplikasian dari artificial intelligent (AI) kedalam mesin atau sistem agar dapat menemukan pola dari suatu data dan mengambil kesimpulan tanpa ada campur tangan manusia. Metode pada machine learning dibagi menjadi dua yaitu supervised learning dan unsupervised learning. Perbedaan yang paling terlihat dari kedua metode tersebut adalah pada pelabelan data yang digunakan untuk melakukan proses learning. Data dikatakan memiliki label jika data tersebut sudah dikumpulkan kedalam suatu klas dan sudah diketui nama dari klas tersebut, misalnya data foto kucing dan anjing adalah data yang sudah memiliki label. [6]

### **8. Deep Learning**

Deep learning merupakan implementasi dari AI yang menirukan cara kerja otak manusia dalam memproses data dan menemukan pola untuk mengambil keputusan. Deep learning termasuk kedalam bagian machine learning didalam bidang AI yang mampu melakukan pembelajaran unsupervised dari data yang tidak terstruktur dan tidak berlabel. Perkembangan deep learning diikuti dengan terjadinya ledakan penyebaran data dalam berbagai bentuk ke seluruh bagian dunia, kumpulan data ini kemudian dinamakan dengan bigdata. Untuk mengolah banyaknya data tersebut tidak mungkin dilakukan manusia, oleh karena itu peran deep learning sangat

dibutuhkan dalam menangani masalah ini. [7]

### **9. Epoch**

Epoch adalah istilah dalam machine learning yang mengacu pada perulangan dalam proses training dataset. Proses perulangan dalam satu epoch dilakukan dengan mengolah dataset dari data pertama sampai data terakhir yang kemudian dilanjutkan kembali dari data terakhir ke data pertama, jika digambarkan maka polanya seperti  $A - B - C - B - A$ . Biasanya dalam proses training dibutuhkan beberapa epoch, hal ini dilakukan agar model algoritma mendapatkan kombinasi pola yang rumit. Harapan dari penggunaan epoch yaitu agar model dapat mengambil keputusan yang baik jika diberikan data baru atau data diluar dari lingkungan dataset. [8]

### **10. Image Preprocessing**

Image preprocessing merupakan metode untuk mengubah dan mengekstrak data gambar agar mendapatkan informasi yang dapat digunakan. Mendapatkan informasi gambar, bisa dilakukan dengan memotong gambar dengan ukuran tertentu, meningkatkan kualitas gambar, dan ekstrasi atribut seperti nilai RGB. [8]

### **11. Ground-truth**

Ground truth adalah istilah yang digunakan dalam pengertian machine learning untuk melakukan pengukuran target training dan testing sehingga bisa disebut sebagai label pada dataset. Dalam beberapa kasus, ground truth tidak selalu diartikan sebagai label. Misalnya, ketika melakukan pengukuran untuk dataset, nilai yang diberikan bisa berbeda dengan nilai yang sudah ditentukan sebagai label dari dataset. Hal ini, bukan menjadi masalah yang besar karena dalam melakukan pengukuran biasa terjadi kesalahan perhitungan. [6]

### **12. Anotasi**

Anotasi adalah proses pelabelan data yang biasa dilakukan dengan

menggunakan tools. Proses pelabelan ini dilakukan dengan membuat kotak atau Bounding Box di sekeliling objek yang akan digukan sebagai dataset untuk diteksi objek dan pengenalan citra. Untuk melakukan hal tersebut bisa dilakukan dengan menggunakan tools siap pakai. Tools ini akan menghasilkan file dalam format XML yang berisikan data – data dari gambar seperti lokasi gambar pada komputer, nama file, dan lokasi Bounding Box dalam koordinat xmin, xmax, ymin, dan ymax. [9]

### 13. Black Box Testing

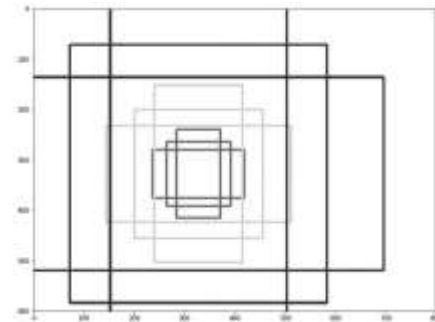
Black box merupakan pengujian yang dilakukan untuk mengamati hasil eksekusi dengan menggunakan data uji guna untuk memeriksa fungsi dari aplikasi yang dikembangkan. Ketika melakukan pengujian, black box bagaikan sebuah kotak yang tidak dapat dilihat isinya tetapi bisa dilihat luarnya. Hal ini disebabkan ketika melakukan pengujian fungsi aplikasi, isi dari fungsi tersebut tidak diperlihatkan karena hanya digunakan sebagai pengecekan fungsional aplikasi. [10]

### 14. Konversi Data

Konversi data adalah proses mengubah format sebuah data menjadi format yang berbeda. Konversi data dilakukan ketika format data yang ingin digunakan tidak bisa dikelani oleh perangkat lunak. Konversi bisa dilakukan dengan menggunakan aplikasis khusus untuk melakukan tugas tersebut. [11]

### 15. Anchors

Anchors merupakan area dengan bentuk kotak pada suatu area tertentu yang digunakan untuk menandai area gambar. Pengaturan standar yang digunkana Faster R-CNN, menggunakan sembilan gambar pada posisi gambar. [2] Box anchors dapat dilihat pada Gambar 2.



Gambar 2 Anchors box

### 16. Python

Python merupakan bahasa pemograman yang biasa digunakan dalam membangun algoritma machine learning. Selain itu, bahas pemograman ini sangat mudah untuk digunakan karena menggunakan struktur sintak yang rapih dan sangat mudah untuk dipahami dibandingkan dengan bahasa pemorgraman lainnya seperti C++ dan Java [12]. Karena Python merupakan bahasa pemograman, maka, harus digunakan text editor untuk membangun algoritma dengan bahasa pemograman ini.

### B. Pengumpulan data

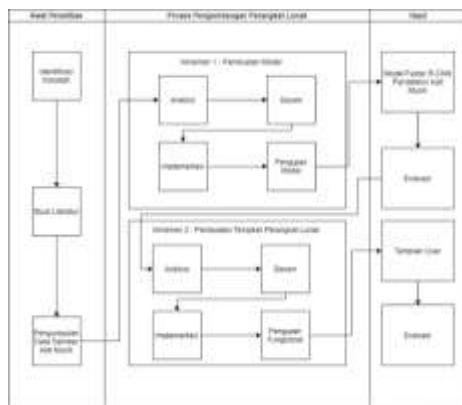
Dataset yang digunakan pada penelitian ini perlu dilakukan ekstrasi informasi yang akan digunakan saat melakukan proses training dan testing model Faster R-CNN. Sebelum melakukan ekstrasi, perlu dilakukan anotasi data. Proses anotasi data dilakukan untuk membuat bounding box yang nanti digunakan sebagai informasi untuk diekstrasi.



Gambar 3 Gambar proses anotasi dengan bounding box

Untuk membuat bounding box dapat menggunakan LabelImg sebagai tools dalam melakukan anotasi, tampilan yang sederhana dari tools ini sangat membantu proses anotasi sehingga proses ini dapat terselesaikan dengan cepat. Tools ini akan menyimpan hasil anotasi kedalam beberapa file yang berisikan label objek, lokasi gambar, lokasi bounding box dalam koordinat xmin, xmax, ymin, dan ymax. Kemudian dari file tersebut akan dilakukan proses ekstraksi untuk keempat atribut. Data kemudian disimpan kedalam satu buah file, file ini yang akan digunakan pada proses training dan testing model.

**C. Implementasi**



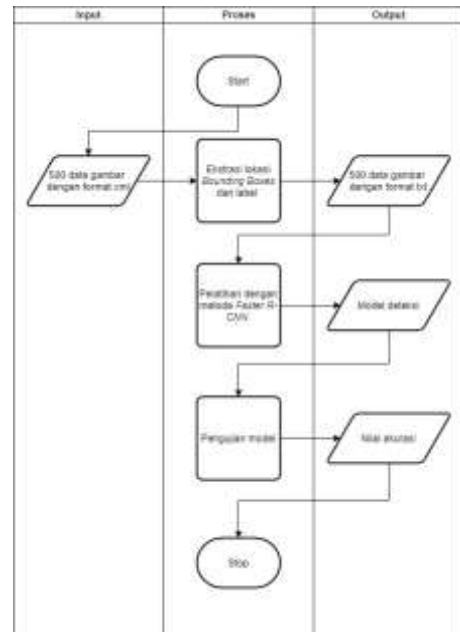
Gambar 4 Alur proses penelitian

Proses penelitian ini akan menggunakan dua inkremen. Untuk inkremen satu akan fokus dalam melakukan pembangunan model algoritma pendeteksi gambar alat musik, sedangkan untuk inkremen dua akan fokus membangun tampilan aplikasi berbasis desktop.

Pada Gambar 4, rancangan pengembangan aplikasi dimulai dari indentifikasi masalah sampai hasil evaluasi. Setelah melakukan indentifikasi masalah, dilakukan studi literatur oleh peneliti untuk mendukung proses penelitian yang dilakukan. Setelah dua tahap tersebut selesai dilakukan maka peneliti mulai mengumpulkan data untuk digunakan dalam tahap inkremen satu, kemudian dilanjutkan ketahap inkremen dua untuk membuat tampilan aplikasi

untuk pengguna. Pada tahap akhir peneliti akan melakukan evaluasi dari setiap inkremen guna memperoleh kekurangan dan kelebihan dari aplikasi ini.

**1. Inkremen satu**



Gambar 5 Alur inkremen satu

Proses ini melakukan analisis kebutuhan-kebutuhan untuk membangun logika program deteksi alat musik dengan Faster R-CNN. Hal pertama yang dibutuhkan adalah dataset gambar alat musik beserta anotasi data gambar. Hasil data dari anotasi berformat pascal VOC.

Agar format tersebut dapat dibaca metode keras frCNN maka perlu ekstraksi data yang dibutuhkan dan disimpan kedalam file txt. Oleh karena itu, diperlukan code untuk melakukan ekstraksi tersebut. Data yang digunakan adalah lokasi. Setelah melakukan ekstraksi data, data tersebut sudah bisa digunakan untuk melatih model deteksi alat musik.

Berdasarkan studi literatur yang telah dilakukan, peneliti memutuskan memilih menggunakan metode Faster R-CNN untuk melakukan deteksi alat musik. Alasan memilih metode ini karena menggantikan Selective Search Method pada versi sebelumnya dengan

Region Proposal Network (RPN) yang dikatakan pada jurnal acuan membuat algoritma menjadi lebih cepat. Dalam melakukan proses training peneliti menggunakan komputer pribadi sehingga perlu memasang perangkat lunak CUDA dan cuDNN sebelumnya. Agar kedua perangkat lunak tersebut dapat berjalan dengan baik maka harus disesuaikan dengan versi Library Tensorflow-gpu yang sesuai.

Tabel 1 Spesifikasi perangkat yang digunakan

Perangkat Keras	Spesifikasi	Perangkat Lunak
Komputer	Processor : AMD Ryzen 3 2200G Ram : 8 GB Vga : GTX 1660 Ti Storage : 1 TB	CMD.exe Prompt 0.1.1 Spyder 4.0.1 LabelImg Cuda 10 cuDNN 7.4

Tabel 2 Library yang digunakan

Library	Versi
Keras	2.3.1
Tensorflow-gpu	1.14.0
PIL	7.1.2
PyQt5	5.13.0

Tahap selanjutnya dilakukan ekstrasi lokasi *Bounding Boxes* dan label dari file (.XML). Data lokasi tersebut terdiri dari nilai xmin, xmax, ymin, dan ymax. Setelah melakukan ekstrasi, data yang sudah dikumpulkan disimpan kedalam file dengan format (.TXT). Jika sudah melakukan tahap ini maka bisa dilanjutkan ke tahap berikutnya. Data yang sudah diolah siap digunakan untuk proses *training* model. Proses training ini menggunakan arsitektur resnet50 dan melakukan augmentasi dengan membalik gambar secara horisontal dan vertikal, dan memutar gambar sebesar sembilan puluh derajat. Output dari proses training adalah *file* (.PICKLE) dan (.HDF5) yang

nanti digunakan untuk melakukan testing. Selesai melakukan proses training, nilai bobot yang tersimpan pada model sudah siap dilakukan testing. Hasil dari testing ini adalah nilai akurasi, dan hasil deteksi alat musik. Nilai akurasi digunakan untuk mengetahui seberapa akurat model mendeteksi gambar alat musik. Sedangkan hasil diteksi digunakan untuk mengetahui jika model sudah dapat mendeteksi objek dengan benar. Maka pada akhir tahapan inkremen satu, akan dilakukan pengujian dengan tujuan untuk mendapatkan informasi kelemahan dan kekurangan dari model ini.

Kemudian pada tahap terakhir akan dilakukan pengujian model. Tahap pengujian yang dilakukan pada inkremen satu dapat dilihat pada Tabel 1.

Tabel 3 Pengujian Inkremen Satu

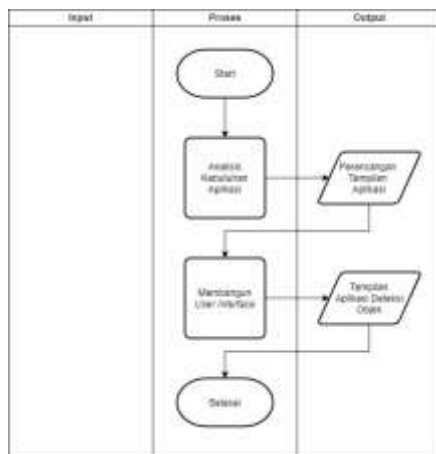
Tahap Uji	Data Uji	Harapan
Ekstrasi label dan <i>Bounding Boxes</i>	Data (.XML) pada folder annotations	Menghasilkan file annotation.txt
Trainig model	Data annotation.txt	Menghasilkan file config.pickle dan model_frcnn.hdf5
Testing model	File config.pickle dan model_frcnn.hdf5	Melakukan deteksi dengan <i>Faster R-CNN</i>

Pada pengujian inkremen satu, terdapat tiga tahap uji yang akan dilakukan. Pertama akan dilakukan uji ekstrasi label dan bounding box, data yang akan digunakan adalah data pada file (.XML), hasil akhir yang diharapkan dari tahap ini yaitu menghasilkan output berupa data anotasi. Kemudian tahap uji selanjutnya dilakukan pengujian pada proses training model. Proses ini akan menggunakan data anotasi hasil dari tahap uji pertama, dari tahap uji ini diharapkan menghasilkan output file yang menyimpan bobot model. Untuk tahap uji selanjutnya akan dilakukan pengujian pada proses testing model. Tahap uji ini menggunakan data hasil

dari tahap proses training model. Diharapkan tahap ini dapat melakukan deteksi gambar alat musik dengan metode Faster R-CNN.

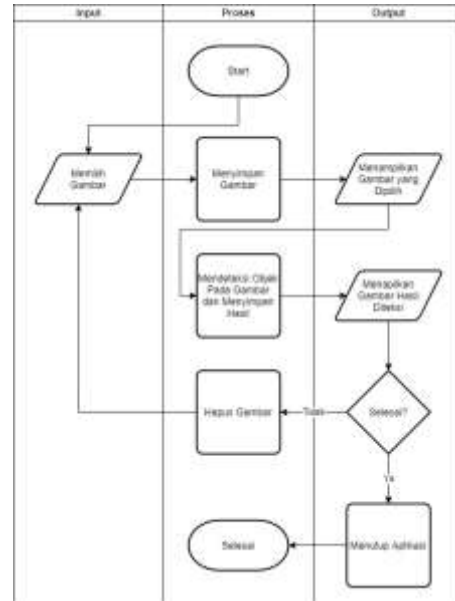
## 2. Inkremen dua

Inkremen dua adalah tahap terakhir dalam melanjutkan proses pada tahap inkremental satu. Didalam tahap akhir ini, akan dibuatkan tampilan aplikasi untuk mempermudah pengguna dalam menggunakan aplikasi ini. Pada tahap ini, penelitian akan fokus membuat tampilan aplikasi yang sesuai dengan kebutuhan penelitian. Pada inkremen dua, ada beberapa objek yang akan digunakan pada aplikasi ini. Aplikasi ini hanya menampilkan satu tampilan untuk melakukan proses deteksi alat musik. Objek – objek ini akan dimasukkan kedalam tampilan aplikasi ini. Pada Gambar 4 ditampilkan tahapan yang dilakukan pada penelitian dalam membuat tampilan aplikasi deteksi alat musik.



Gambar 6 Tahapan pengembangan tampilan

Kemudian ditampilkan juga gambar alur penggunaan aplikasi ini agar tidak terjadi kesalahan input ketika digunakan oleh pengguna. Alur penggunaan aplikasi dapat dilihat pada Gambar 7 berikut ini.



Gambar 7 Alur penggunaan aplikasi

Langkah–langkah penggunaan aplikasi ini, yaitu, pertama pengguna memasukkan data gambar dengan tombol Open File, kemudian menggunakan tombol Proses untuk melakukan deteksi pada data yang sudah dimasukkan. Hasil deteksi dan data yang dimasukkan ditampilkan pada dua buah text field, yaitu field Original Image dan Image Found. Setelah selesai melakukan deteksi, pengguna bisa menekan tombol Hapus atau tombol Keluar. Tombol hapus akan menghapus gambar pada *field Original Image* dan *Image Found* sehingga siap untuk digunakan Kembali untuk deteksi.

Kedua gambar alur tersebut adalah hasil analisis kebutuhan fungsi yang akan digunakan untuk aplikasi deteksi alat musik. Aplikasi ini hanya digunakan untuk melakukan deteksi alat musik. Aplikasi ini akan menampilkan gambar yang digunakan pengguna kemudian menampilkan gambar hasil dari deteksi objek. Untuk mempermudah pengguna dalam menggunakan aplikasi ini, peneliti menyediakan empat tombol yaitu tombol Open File, tombol Close, tombol Proses, tombol Hapus. Tombol – tombol ini diletakkan pada bagian kanan hasil deteksi dan disusun secara vertikal. Untuk tombol Open File berfungsi untuk



membuka file gambar dan disimpan kedalam sebuah folder gambar, kemudian ditampilkan pada aplikasi. Tombol Close berfungsi menampilkan menu pop up untuk menayakan pengguna apakah ingin keluar aplikasi atau tidak. Tombol Proses berfungsi menjalankan code test dan membaca file didalam folder gambar. Setelah selesai memproses gambar, hasilnya akan disimpan pada folder *result\_img*, kemudian gambar yang sudah disimpan ditampilkan pada aplikasi ini. Tombol Hapus berfungsi untuk menghapus seluruh gambar yang ditampilkan pada aplikasi ini.

Kemudian pada tahap terakhir akan dilakukan pengujian model. Tahap pengujian yang dilakukan pada inkremen dua dapat dilihat pada Tabel 4

Tabel 4 Pengujian Inkremen Dua

Tahap Uji	Fungsi	Harapan
Menekan tombol "Open File"	Membuka folder untuk mencari <i>file</i> gambar kemudian ditampilkan dan disimpan	Dapat membuka folder dan menampilkan <i>file</i> dengan format (.PNG), (.JPG), dan (.JPEG) kemudian ditampilkan dan disimpan kedalam folder <i>image</i>
Menekan tombol "Proses"	Menjalankan model deteksi objek kemudian menyimpan hasil lalu ditampilkan	Melakukan deteksi objek yang hasilnya disimpan kedalam folder <i>result_img</i> kemudian ditampilkan pada <i>text box</i>
Menekan tombol "Hapus"	Menghapus gambar	Dapat mengganti nilai pada objek <i>Original Image</i> dan <i>Object Image</i> sehingga gambar yang sudah ditampilkan menjadi kosong

Menekan tombol "Keluar"	Keluar dari aplikasi	Menjalankan <i>code</i> untuk menutup aplikasi
-------------------------	----------------------	--

Pengujian untuk inkremen dua akan melakukan uji coba tombol pada tampilan aplikasi. Pada tahap uji pertama akan dilakukan pengujian untuk tombol "Open File", tujuan pengujian tombol ini yaitu untuk mengetahui apakah tombol dapat membuka file dialog dan menampilkan gambar pada tempat yang disediakan dalam tampilan aplikasi. Kemudian tahap uji kedua akan menguji tombol "Proses" untuk mengetahui apakah tombol ini dapat memproses gambar yang sudah dibuka melalui file dialog atau tidak kemudian hasilnya akan ditampilkan pada tampilan aplikasi. Selanjutnya pada tahap uji tombol "Hapus" akan diperhatikan apakah tombol dapat menghapus gambar yang sudah ditampilkan pada tampilan aplikasi. Kemudian tahap uji terakhir akan melakukan uji tombol "Keluar". Hasil yang diharapkan dari tombol ini yaitu dapat menutup atau memberhentikan aplikasi. Untuk hasil pengujian akan dibahas pada pembahasan.

### III. HASIL DAN PEMBAHASAN

#### A. Inkremen satu



Gambar 8 Hasil deteksi dengan metode Faster R-CNN

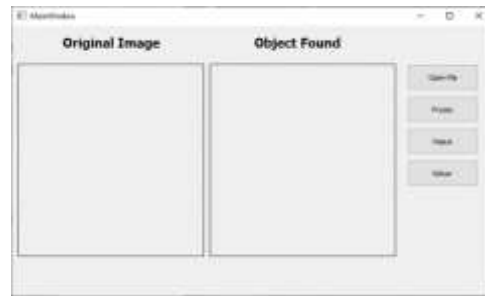
Proses pengembangan algoritma pendeteksi alat musik dimulai dari ekstraksi file (.CSV) untuk mendapatkan data label dan lokasi Bounding Boxes yang kemudian disimpan kedalam file `annotate.txt`. Selanjutnya file `annotate.txt` dimasukkan kedalam code training. Dalam melakukan proses training, peneliti menggunakan epoch sebanyak dua ratus dan Batch Size sebanyak lima ratus. Setelah proses training selesai, code training menghasilkan dua buah file baru, yaitu, `config.pickle` dan `model_frcnn.hdf5`. Kedua file tersebut akan digunakan dalam melakukan testing. Hasil dari proses testing menampilkan gambar pada jendela baru dengan Bounding Box dan label yang sesuai dengan objek yang telah ditemukan. Nilai akurasi juga ditampilkan pada gambar dan dicetak pada output Command Prompt. pada jendela Command Prompt ditampilkan juga lama waktu model ini memproses gambar yang dituliskan pada Elapsed Time.

```
Elapsed time = 11.721171379089355
[{'bassoon', 99.80251789093018}]
```

Gambar 9 Nilai akurasi dan elapsed time

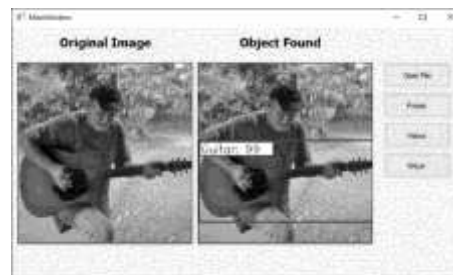
Ketika dalam proses pengujian untuk inkremen satu, ditemukan kelemahan pada model ini dalam melakukan pendeteksian gambar alat musik. Saat menggunakan data baru, model terkadang tidak dapat mengenali gambar alat musik tersebut bahkan terjadi pendeteksian yang salah. Masalah ini dapat disebabkan oleh gambar – gambar yang kurang bervariasi karena gambar alat musik diambil dari posisi depan dan samping saja, sehingga ketika menggunakan gambar alat musik yang diambil dari atas maka dapat dipastikan model ini tidak dapat mengenali alat musik tersebut. Selain itu perlu juga diperhatikan jumlah epoch yang digunakan, hal ini bisa dilakukan dengan melakukan pengurangan dan penambahan nilai epoch.

## B. Inkremen dua



Gambar 10 Hasil akhir tampilan aplikasi

Pengembangan aplikasi perlu memperhatikan fungsi dan tampilan aplikasi. Fungsi yang bekerja sesuai dengan tujuannya dapat mempermudah pengguna menggunakan aplikasi, selain itu pemberian batasan pada fungsi juga bermanfaat agar pengguna tidak menggunakan aplikasi diluar kemampuannya. Tampilan yang dibuat sederhana harus dapat menyampaikan informasi kepada pengguna sehingga pengguna tidak kesulitan menggunakan aplikasi yang sudah dibangun. Pyq5 dan Pyqt5 Tools merupakan alat untuk membuat tampilan yang mudah dan cepat.



Gambar 11 Hasil pengujian aplikasi

Berdasarkan hasil pengujian pada Tabel 4 tampilan aplikasi ini masih memiliki bug pada tombol “Open File” dan untuk tombol yang masih belum berfungsi dengan baik yaitu tombol “Keluar”. Permasalahan pada tombol “Open File” terjadi ketika setelah menekan tombol maka akan muncul jendela baru yang kosong setiap kali tombol ini ditekan. Masalah ini tidak muncul ketika aplikasi masih dalam proses pembangunan pada tombol “Open File” dan “Hapus”. Kemudian permasalahan

pada tombol “Keluar” aplikasi tidak tertutup melainkan terjadi crash sehingga harus dihentikan secara paksa melalui Task Manager.

#### IV. SIMPULAN

Berdasarkan pembahasan sebelumnya maka dapat ditarik beberapa kesimpulan dari hasil penelitian sebagai berikut:

- Metode Faster R-CNN dapat digunakan untuk mendeteksi gambar alat musik.
- Aplikasi dapat bekerja dengan baik jika menggunakan gambar dengan kualitas yang bagus
- Model dapat mendeteksi gambar dengan baik jika objek pada gambar berukuran kecil dan diambil dalam posisi alat musik sedang dimainkan

#### DAFTAR RUJUKAN

- [1] S. Hadi, “Sejarah Musik,” no. November 2015, p. 591, 2009.
- [2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.
- [4] Ms. Suyanto, ST, *Artificial Intelligence*, 2nd ed. 2014.
- [5] N. Sebe, *Machine Learning in Computer Vision*. 2005.
- [6] Y. Baştanlar and M. Özuysal, “Introduction to machine learning,” *Methods Mol. Biol.*, vol. 1107, pp. 105–128, 2014.
- [7] D. D. Model, “applied sciences Efficient Deep Learning for Gradient-Enhanced Stress Dependent Damage Model.”
- [8] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Inc., 2019.
- [9] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes Challenge: A Retrospective,” *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2014.
- [10] M. Larrea, “Black-Box Testing Technique for Information Visualization. Sequencing Constraints with Low-Level Interactions,” *J. Comput. Sci. Technol.*, vol. 17, no. 1, pp. 37–48, 2017.
- [11] V. Y. Lum, N. C. Shu, and B. C. Housel, “General Methodology for Data Conversion and Restructuring,” *IBM J. Res. Dev.*, vol. 20, no. 5, pp. 483–497, 1976.
- [12] S. Maruch and A. Maruch, *Python For Dummies*. Wiley Publishing, Inc., 2006.
- [13] J. U. Kim and Y. M. Ro, “ATTENTIVE LAYER SEPARATION FOR OBJECT CLASSIFICATION AND OBJECT LOCALIZATION IN OBJECT DETECTION Jung Uk Kim and Yong Man Ro,” 2019 *IEEE Int. Conf. Image Process.*, vol. 1, pp. 3995–3999, 2019.
- [14] Z. Zhao and P. Zheng, “Object Detection with Deep Learning : A Review,” pp. 1–21, 2012.
- [15] D. Alamsyah and D. Pratama, “Deteksi Ujung Jari menggunakan Faster-RCNN dengan Arsitektur Inception v2 pada Citra Derau,” *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 2, no. 1, p. 1, 2019.
- [16] J. S. Asri and G. Firmansyah, “Implementasi objek detection dan tracking menggunakan deep learning untuk pengolahan citra digital,” pp. 8–9, 2018.