JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2230-2239

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

# Optimization of Historic Buildings Recognition: CNN Model and Supported by Pre-processing Methods

Abdul Haris Rangkuti [a,*], Varyl Hasbi Athala [a], Farrel Haridhi Indallah [a] , Evawaty Tanuar [a], Johan Muliadi Kerta [a]

[a] *Informatics Department, School of Computer Science, Bina Nusantara University, Palmerah, Jakarta, 11480, Indonesia*
*Corresponding author: [*]arangku2000@binus.ac.id*

***Abstract*— Several cities in Indonesia, such as Cirebon, Bandung, and Bogor, have several historical buildings that date back to the Dutch colonial period. Several Dutch colonial heritage buildings can be found in several areas. The existence of historical buildings also would attract foreign or local tourists who visit one of an area. We need a technology or model that would support the recognition and identification of buildings, including their characteristics. However, recognizing and identifying them is a problem in itself, so technology would be needed to help them. The technology or model that would be implemented in this research is the Convolutional Neural Network model, a derivative of Artificial Intelligent technology focused on image processing and pattern recognition. This process consists of several stages. The initial stage uses the Gaussian Blur, SuCK, and CLAHE methods which are useful for image sharpening and recognition. The second process is feature extraction of the image characteristics of buildings. The results of the image process will support the third process, namely the image retrieval process of buildings based on their characteristics. Based on these three main processes, they would facilitate and support local and foreign tourists to recognize historic buildings in the area. In this experiment, the Euclidean distance and Manhattan distance methods were used in the retrieval process. The highest accuracy in the retrieval process for the feature extraction process with the DenseNet 121 model with the initial process is Gaussian Blur of 88.96% and 88.46%, with the SuCK method of 88.3 and 87.8%, and with CLAHE of 87.7%, and 87.6%. We hope that this research can be continued to identify buildings with more complex characteristics and models.***

***Keywords*— Historical building; SuCK; CLAHE; gaussian blur; convolutional neural network; artificial intelligent.**

*Manuscript received 8 Nov. 2022; revised 14 Apr. 2023; accepted 16 May 2023. Date of publication 31 Dec. 2023.*


## I. INTRODUCTION

Several cities in Indonesia, particularly in West Java, such as Bandung, Cirebon, and Bogor, have many Historical buildings that date back to the Dutch East Indies period. As a result, if we walk through the Old Town Area or Heritage Area in Bandung, Cirebon, or Bogor, we will see many old buildings, both Dutch East Indies relics and Cultural Heritage Buildings. Relics from the Dutch East Indies period are considered cultural heritage and called Cultural Heritage Buildings [1]. Cultural heritage is an area where the preservation of people's lives and livelihoods is legally protected from extinction. Cultural heritage is any form of material cultural heritage that is considered as cultural heritage objects, cultural heritage buildings, cultural heritage structures, cultural heritage sites, and cultural heritage areas on land and/or in water that needs to be preserved because they have important values for history, science, education, religion, and/or culture through a determination process [2].

The Cultural Conservation area is a geographical space unit that contains two or more sites close to each other and/or exhibits distinct spatial characteristics. The Regional Museum is an institution that protects, develops, utilizes, and communicates the collection to the public in the City Region. As a result, the Regional Government grants Cultural Conservation status to Objects, Buildings, Structures, Locations, or Geographical Space Units based on the recommendation of the Cultural Conservation Expert Team [3].

In general, cultural heritage is a cultural wealth as a form of thought and behavior of human life, so it must be preserved and managed appropriately in the context of the welfare of the people of Bandung [4]. Development efforts in the context of conservation are defined as increasing the potential value, information, and promotion of cultural heritage, as well as its utilization, through research, revitalization, and adaptation [5]. However, problems arise when it is uninformative and lacks strategic placement of how to find and sign systems that contain information, history, and profiles of Historical Buildings, so many visitors still have difficulty finding information and navigating places in Kota Tua [6]. For this reason, using Artificial Intelligence technology, develop an app that can detect several Historical buildings and their characteristics. Furthermore, this research will reveal a more complete re-discovery of Historical buildings in a specific area, such as Bogor, Bandung, and Cirebon. Historic sites are a form of ancestor and cultural heritage that has value as a source of inspiration for the nation's life today and in the future [7].

This study also focused on introducing the inside of the Historical Building. A building is a physical form of construction work that is integrated with its domicile, most of which is above or in the ground or air, which functions as a place for humans to carry out their activities, either for housing or residence, religious activities, religious activities, or religious activities—social, cultural, and special activities [8].

### *A. Related Study*

Essentially, the object detection process is currently an important research area in the field of computer vision and computer vision artificial intelligence. Detecting and recognizing buildings is one of them [9]. In recent years, some research papers have been published using machine learning and computer vision approaches in the Historical architecture and archaeology section [10]. Another experiment for the application aspect is that in the absence of a tour guide, this application will assist tourists in determining the construction period or era by detecting the features of old spectacular architecture. Another study has focused on the constructional characteristics of old architectural sites using the Canny Edge Detector method [11]. The other proposes an idea to recognize and detect the textures, decorations, and other features of Historical buildings based on machine vision. First, classify many surface texture images of Historical building components manually as a set of samples. The convolution neural network is then used to train the samples to obtain a classification detector.

Finally, check its precision [12]. Another research about the introduction of Stone cultural heritage types is based on weathering using Deep Learning and Artificial Neural Networks. The Stone cultural heritage accuracy rates obtained from the DL and ANN models are 99.4% and 93.95%, respectively. The recall rate (96–100%) in each class of the DL model has been determined to be higher. Based on the results, the lowest precision rates in the testing phase were found in fresh rock (97%) and flaking (98%), while 100% precision rates were obtained in the other classification groups [13]. The other research explores the use of sophisticated image recognition algorithms for home style recognition and its limitations and possibilities. In addition, this paper adopts a convolutional neural network model to classify house styles in the US [14]. In this study, we made observations of the realization of the Kazan sightseeing system developed in the process of scientific research. It is a system for augmented reality location tagging and neural network recognition, which is completely new for the Kazan development market [15].

The Other research presents a method of recognizing historical buildings with deep learning for UAV remote sensing technology. In the experiment, the Faster RCNN model was used to identify UAV remote sensing images, which showed that recognition accuracy reached 93.2% for this dataset with an average processing time of 74ms on image recognition. The results illustrate the effectiveness and efficiency of building recognition applications from UAV remote sensing images by deep learning networks [16]. In research on remote sensing, image features are extracted through convolution, pooling, and classification. The data collection model is used for comparative analysis and to verify model performance. The results show that when using CNN to recognize remotely sensed images, the recognition accuracy is much higher than that of traditional image recognition models, which can reach 95.3%. Compared with the newly researched model, the performance has increased by 15%, and the recognition speed is increased by 20% [17]. In another study, describing height information is exploited as an additional feature derived from applying a dense image-matching algorithm. As test sites, several complex urban areas of different building types, pixel resolutions, and data types were used in Vaihingen in Germany and Perissa in Greece. Our method was evaluated using completeness, correctness, and quality levels and compared with other conventional and "shallow" learning paradigms, such as support vector machines [18].

Another research for using vision-based manual inspection technology in identifying and assessing superficial damage to historic buildings is time-consuming and laborious. To overcome this limitation, this paper proposes a new automatic damage detection technique using the Faster R-CNN model based on the ResNet101 framework to detect two categories of damage (efflorescence and flaking) for historic stone structures [19]. Besides, Stone's cultural heritages provide meaningful value and information about the culture, religion, economics, and esthetics of the period in which they were built. This study developed recognition models based on deep learning (DL) and Artificial Neural Network (ANN) to eliminate human errors that may arise in weathering recognition. Although the accuracy rates obtained from the DL and ANN models are 99.4% and 93.95%, respectively, the recall rate (96–100%) in each class of the DL model has been determined to be higher [20]. However, the current visual inspection method for identifying and assessing superficial damage on historic buildings is time and labor intensive. To verify the performance of the proposed method, a comparative study was conducted with Mask R-CNN and a fully convolutional network. This is the first attempt at employing a two-level strategy to automatically detect, segment, and measure large-scale superficial damage on historic buildings based on deep learning, and it achieved good results [21].

In this research, we propose a new deep transfer learning approach based on aerial photography to automatically detect Hakka Weilong Houses (HWHs), a famous type of historical residence and an important cultural symbol of Hakka, which is called the Hakka capital of the world. The model approach used ResNet50 as the backbone transfer network and YOLO v2 as a training framework. Experimental results showed that the average precision was $0.9599 \pm 0.0150$, the loss rate was 0.0250, the Root Mean Square Error (RMSE) for training was 0.1580, and the average detecting time per image clip was $0.0383 \pm 0.0150$ second, suggesting that our model has a high accuracy and an excellent performance for the HWH detection task [22], [23].

## II. Material and Method

This research focused on 180 images of historical buildings detected by the application from three different cities in Indonesia. So that the variations of historical buildings in this research can be conducted, this research also does not concentrate on color characteristics. As a result, all images must be pre-processed to reduce noise or increase image uniqueness. In this experiment, three image processing methods are used: blur, CLAHE, and SUCK. The Gaussian blur is a convolution technique used as a pre-processing stage of many computer vision algorithms. Gaussian Blur reduces noise by blurring, smoothing, and eliminating noise in an image [24], [25]. CLAHE stands for Contrast Limited Adaptive Histogram Equalization, and it equalizes the image's value by reducing contrast amplification. To address the issue of noise amplification, the CLAHE implemented a clipping limit [26, 27]. Finally, SUCK, which stands for Sharpening Using Custom Kernel, is the opposite of the previous two. SUCK sharpens the image to enhance its characteristics. The potential disadvantage is that image noise may be worse than with the other two.

Furthermore, all images will be processed using various Convolutional Neural Network models or CNN models, including InceptionV3, InceptionResNetV2, ResNet50V2, VGG19, and DenseNet201 [28]. These CNN models serve as the feature extractor in this research. The extracted features were compared with two distance metrics: Manhattan Distance and Euclidean Distance. Distance metrics provide algorithms capable of searching for distances that capture features or relationships hidden in our data [29, 30]. Euclidean distance is the distance between points in a straight line. This distance method uses the Pythagorean theorem.

Moreover, Euclidean distance is the distance calculation that is most commonly used in the machine learning process. On the other hand, the Manhattan distance is the sum of the distances from all attributes [31, 32, 33]. In this experiment, there are four main activities in image retrieval research: collecting images, pre-processing images, training by feature extraction, and testing by finding similarities between datasets. These main activities are completed in order to support maximum accuracy. This research would find out more about the performance. The training dataset will be rotated and scaled, increasing the dataset and the validity of the relative performance. Thus, this research focuses on the performance comparison between these image processing methods, CNN Models, and Distance Metrics. A more thorough explanation can be seen in the following sections.

### *A. Collecting Images of Historical Building*

This experiment has three cities, each with 17 to 24 historical buildings. Images were gathered by visiting the building and taking some pictures or by getting the images from the internet. Each image has unique lighting, size, setting, or angle. These distinctions are important in feature extraction. Image retrieval aims to teach the computer to recognize similar images [34]. As a result, the image must be pre-processed to reduce noise, lighting, and any other unnecessary differences.

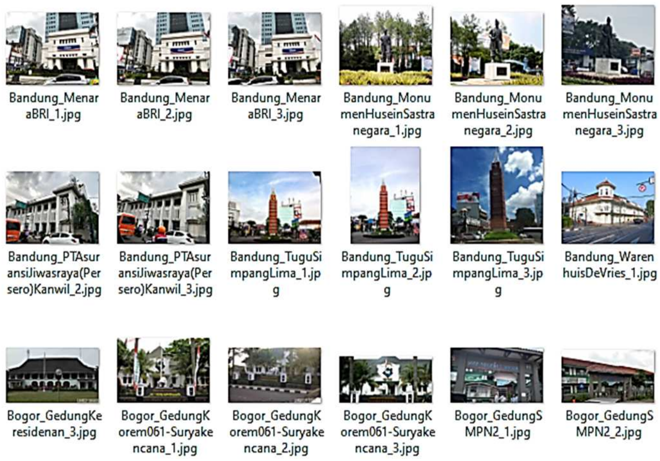


Fig. 1  Collected Images of Historical Building

Fig 1 Inform to calculate accuracy by name become better, name formatting is required for this experiment. These images are formatted into the special naming format with 3 information combined without space but with underscore symbols.

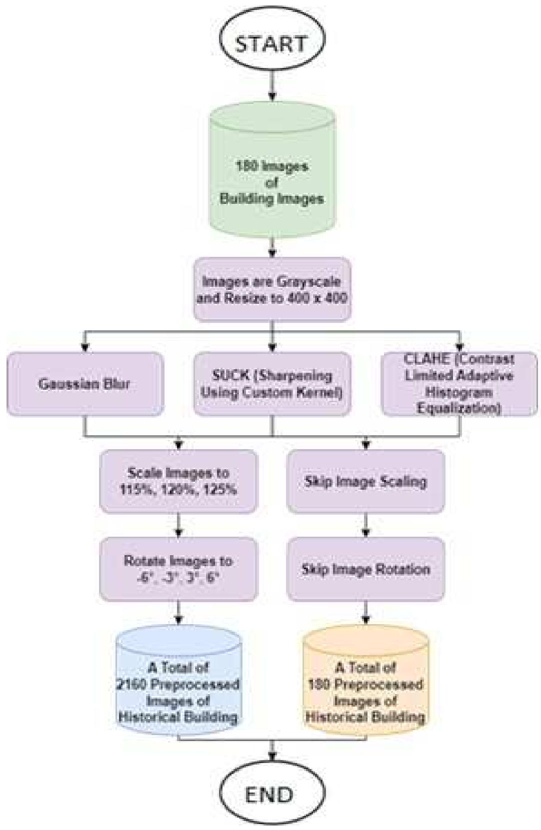


Fig. 2  Image building carry out a Pre-processing

Fig. 2 informs the pre-processing images. As previously stated, each image differs in lighting and resolution size. For this experiment, the same image size with a 1:1 ratio is required for CNN to accept it as input. Input in such aspect ratio is a common practice in computer vision because the ImageNet data in pre-trained weights use resolutions in either 224x224 or 227x227 [35], [36]. Therefore, this procedure begins by converting the image to grayscale and resizing it to

400x400. The image will then be processed using one of the image processing methods. Following that, there are two types of datasets: training dataset (shown the purple color in Fig. 2) and testing dataset (shown in blue in Fig. 2.). The training dataset was used by an image retrieval algorithm database, while the testing dataset was used to evaluate the performance accuracy [37] [38]. This process generates datasets that differ only in scale and rotation for the training dataset.

### B. Dataset Setup

Fig. 3 informs the three types of pre-processing image results. Each chosen building contains a different number of historical buildings. In this case, Bandung has 24 historic buildings, Bogor has 17 historic buildings, and Cirebon has 20 historic buildings. Each historical building contains 2-3 images at a similar angle—all these combined result in 180 images of historical buildings. The number of images is too few for the machine to learn. Expanding the dataset might improve the robustness of the detector due to increased samples [39]. Therefore, data augmentation by pre-processing the image is done to both datasets with traditional transformations on the training dataset.


Fig. 3 (a) Sample of Blur Pre-processed Images (b) Sample Of "CLAHE" Pre-processed Images (c) Sample Of "SUCK" Pre-processed Images

Pre-processing causes the amount of the image to increase for the training dataset. This is because of the rotation and scaling phase in the pre-processing. For the testing dataset, the amount is the same as before because of no rotation and scaling involved in the process. In the training dataset's case, the process turns images into 115%, 120%, and 125% in scaling percentage. This means there are 3 different scaling, which causes the amount of training dataset to increase to 549 images or 9 images for each building. Then, the images get rotated into 4 variations of rotation, which are -6°, -3°, 3°, and 6°. Therefore, 549 images multiplied by 4 equals 2196 images of the combined training dataset or 36 for each building.

In this study, several pre-processing methods were used such as Blur, CLAHE and SuCK to support the feature extraction process for building images. The novelty of this research is that after pre-processing using several methods, all building images per pixel will be converted into array numbers and values for the image normalization process. Each array will be expanded into the image width, height, and channel parameters. This process will support the classification process using several CNN models (all of these processes can be seen in Figure 4.0.

### C. Training by Feature Extract

Fig. 4 describes the training process for the machine to calculate image similarity. Image features are extracted from the training database.

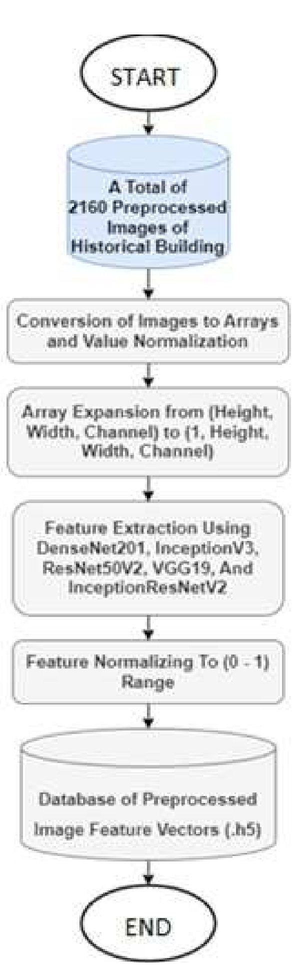


Fig. 4 Diagram of database Training by Feature Extraction

The process starts with image-to-array conversion and value normalization. The image array is then expanded to (1, H, W, C). Finally, this experiment uses five different models for feature extraction: InceptionV3, VGG19, DenseNet201, ResNet50V2, and InceptionResNetV2. The extracted feature values are normalized to the (0-1) range. Normalization is performed to ensure the machine's similarity calculation is simplified due to the normalized value. These characteristics from various images are compiled and saved in a database. A database might have different amounts of features due to the CNN model that is used for feature extraction [40]. For example, ResNet50V2 outputs 2048 features for each image. In this experiment, 2196 images were extracted using ResNet50V2. Therefore, the array of the database is (2196, 2048).

### D. Image Query and Result Calculation

Fig. 5 Inform the process of image query and the calculation. This process queries every single dataset in the testing to find 39 matching images. These 39 images are from 36 images from training and the images from the testing dataset to test the ability to recall itself. Like the training phase, the query image undergoes the same phase as depicted in green color. The yellow color illustrates the result calculation phase. In the experiment, Manhattan and Euclidean distances are used for measuring the similarity distance between images.

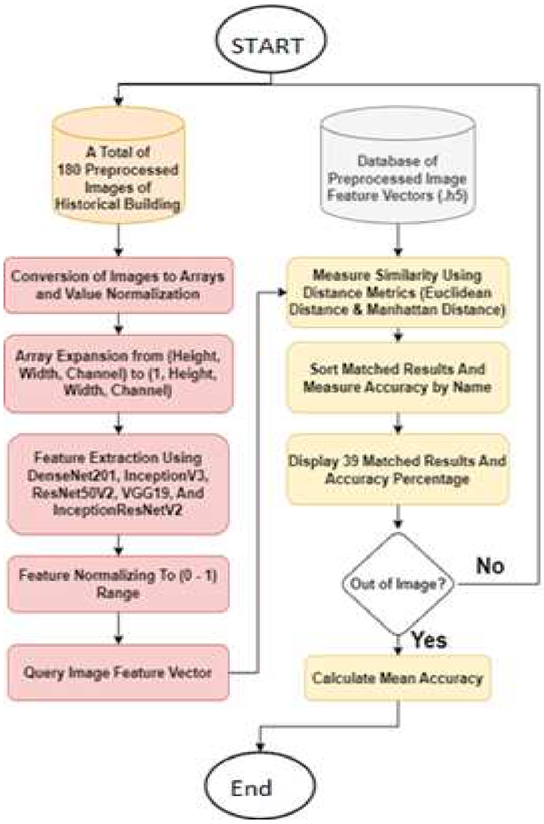


Fig. 5 Diagram of Image Query and Result Calculation

## III. Results and Discussion

Based on the experiments that have been completed, the results of the CNN model testing query with the Euclidean Distance and Manhattan Distance will display historical buildings that tourists wish to recognize further. All historic building data that has been collected will be analyzed to determine which CNN model is most suitable for detecting historic buildings in three different cities (Bandung, Bogor, and Cirebon). In addition, the Manhattan and Euclidean methods will support the feature extraction process to carry out the retrieval process of historic buildings. All the processes would be elaborated such as:

### A. CNN Results

In this experiment, five different CNN models with Euclidean Distance are used in the testing query, and the same CNN models are used in a second test with Manhattan Distance. Regardless of whether a model uses the Manhattan Distance or the Euclidean Distance, the results show that all models have a high accuracy of picture retrieval. The first testing result will be shown is the CNN model testing with Euclidean Distance. Fig. 6. Shows a more thorough detail of the total query above 70% between CNN models using the Euclidean Distance.

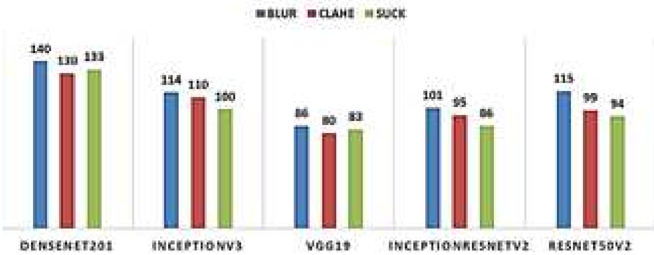


Fig. 6 Comparison of Total Query Above 70% between CNN Models with Euclidean Distance

DenseNet201 is the CNN model with the most images with retrieval accuracy greater than 70%, with 140 images for the pre-process images blur, 130 images for the pre-process images CLAHE, and 133 images for the pre-process images SUCK. The other four CNN models did not come close to DenseNet201. The InceptionV3 model finished second because it had over 100 images for each of the three types of pre-processing images. While VGG19 is the CNN model with the lowest value, there is no value in any of the three types of pre-processing images that exceed 100 images for the VGG19 model.

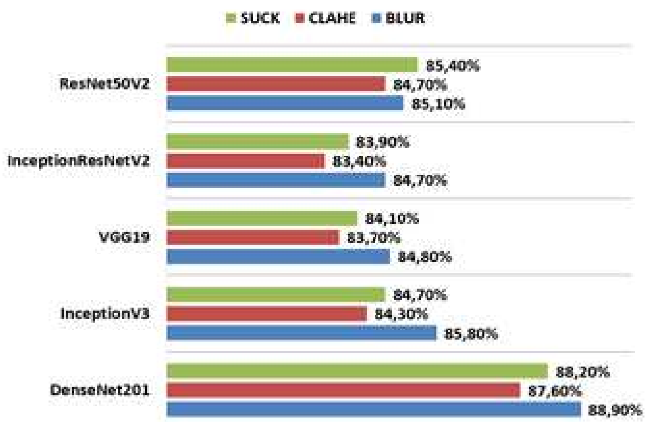


Fig. 7 Mean Accuracy of Image Retrieval CNN Comparison Graph with Euclidean Distance

Fig. 7 compares the mean image retrieval accuracy for each of the five CNN models with Euclidean Distance. Each model has three types of mean accuracy data due to the three types of pre-processing images. DenseNet201 has the highest mean accuracy in the blur pre-process images, with a value of 88.9%. DenseNet201 has a mean accuracy of 88.2% when pre-processing SUCK images and 87.6% when pre-processing CLAHE images. When attempting to detect CLAHE pre-process images, the InceptionResNetV2 model has the lowest mean image retrieval accuracy. DenseNet201 has the highest overall mean accuracy for all pre-processing images.

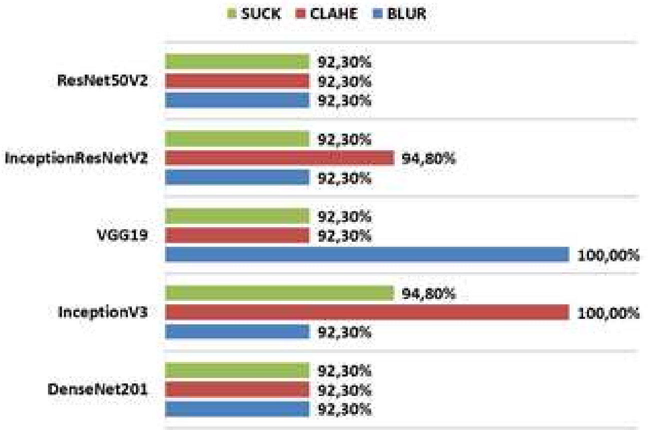


Fig. 8 Max Accuracy of Image Retrieval CNN Comparison Graph with Euclidean Distance

In Fig. 8, InceptionV3 for CLAHE pre-process image and VGG19 for blur pre-process image have the highest accuracy when detecting three pre-processing images with Euclidean Distance. Both CNN models have a 100% accuracy value for their respective pre-process image types. The majority of the CNN models have a maximum accuracy of 92.3%. The CNN model testing results with Manhattan Distance are shown in the second set of results. The number of images with an accuracy value greater than 70% on the CNN testing model that uses the Manhattan distance is not significantly different from those with an accuracy value greater than 70% on the Euclidean distance model. Fig. 9 illustrates a comparison of each CNN model.

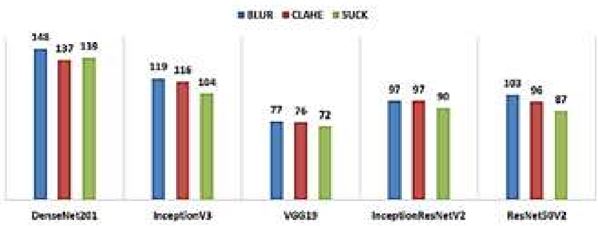


Fig. 9 Comparison of Total Query Above 70% Between CNN Models with Manhattan Distance

Fig. 9 shows the comparison of the total query above 70%. The denseNet201 model has the most images with an accuracy value greater than 86%. The DenseNet201 CNN model includes 148 blur pre-processing images, 137 CLAHE pre-processing images, and 139 SUCK pre-processing images. The results of the other four models are not significantly different from those obtained when Euclidean Distance is used. The CNN model with the fewest total images is the VGG19, with fewer than 80 images in each type of pre-processing image. Fig. 10 informs that the mean image retrieval accuracy for each of the five CNN models with Manhattan Distance is shown in Fig. 10, with each model having three types of mean accuracy data due to the three types of pre-processing images.

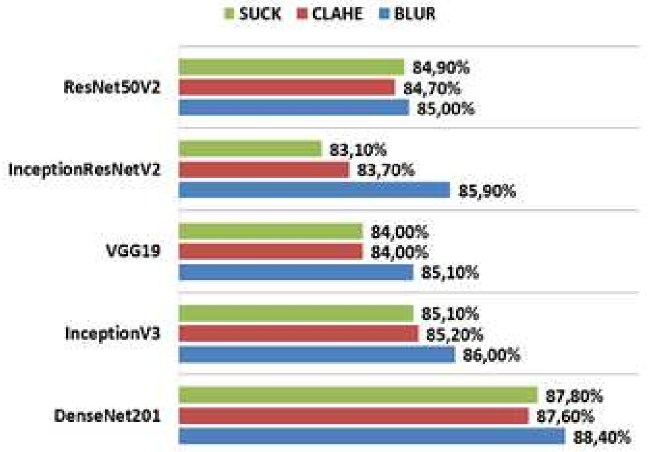


Fig. 10 Mean Accuracy of Image Retrieval CNN Comparison Graph with Manhattan Distance

The highest mean accuracy is DenseNet201 with the blur pre-process images type, with a value of 88.4% mean accuracy. When attempting to detect SUCK pre-process images, the InceptionResNetV2 model has the lowest mean accuracy of image retrieval. DenseNet201 has a mean accuracy of 87.8% for SUCK pre-processing images and 87.6% for CLAHE pre-processing images. This makes DenseNet201 the highest overall mean accuracy for all pre-processing images with Manhattan Distance.

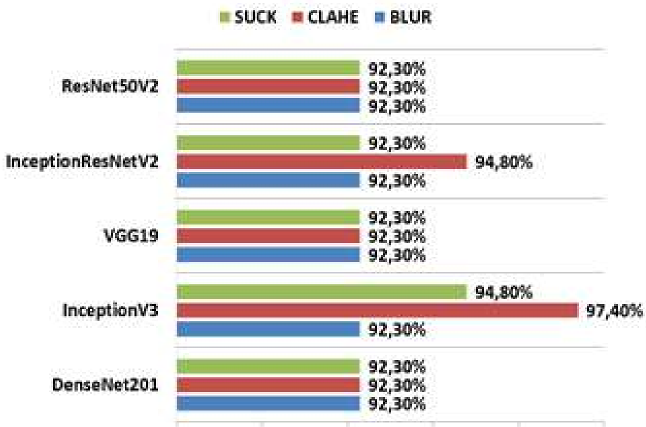


Fig. 11 Max Accuracy of Image Retrieval CNN Comparison Graph with Manhattan Distance

Fig. 11 indicates the highest accuracy a CNN model achieves when detecting three types of pre-process images with Manhattan Distance is InceptionV3 for CLAHE pre-process images, with a value of 97.4%. When detecting SUCK pre-process images, InceptionV3 has the highest max accuracy value, followed by InceptionResNetV2 when detecting CLAHE pre-process images. Both models have a maximum accuracy of 94.8%. The remaining models have the lowest maximum accuracy of 92.3%.

### *B. Implementation Result*

After training and testing all five CNN models, each CNN model can be implemented in the application. The application is made with Tkinter, which is a Python GUI toolkit. The application has three main features: displaying all 61 historical buildings, scanning real-time historical building that is registered, then redirecting them to their history page, and choosing a picture from local data storage to then be scanned then redirected to their history page.


Fig. 12 First/Main Page of the Historical Building Scanning Application

Fig. 12 shows the main page or first page of the application. This main page gives two options. The first is the list of all historical buildings that have been trained to recognize and the second is the scanning page where the application could scan real-time footage of the buildings or choose a picture from previously token historical building photos.

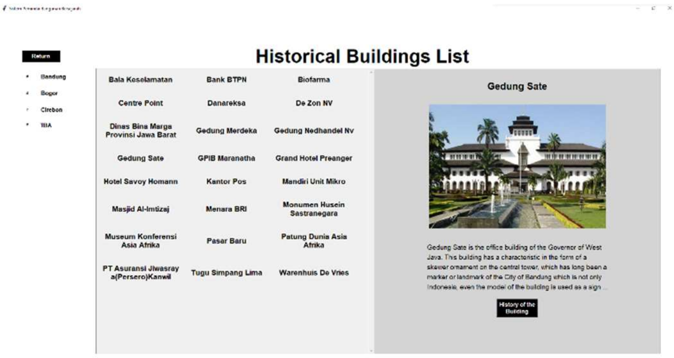


Fig. 13 List of all Registered Historical Buildings in the Historical Building Scanning Application Comparison

Figure 13 shows that the historical buildings list page separates the buildings into their respective city. After choosing which of the three cities, the app will show every historical building in that chosen city. After choosing the desired historical buildings, the app will also display the buildings with a brief history. It will also provide a button to view a more in-depth history of the buildings (As shown in Fig. 14).


Fig. 14 History of the Building in the Historical Building Scanning Application

Figure 14 informs that the history page of the chosen building could only be accessible through the buildings list page, after scanning the real-time footage or choosing a file to scan from the device storage. The scanning process uses the database of pre-processed image feature vectors (a .h5 type file), one of the five CNN models, and one of the two Distance Metrics (Euclidean Distance and Manhattan Distance) to predict which historical building is being scanned. The prediction result is sent to the UI, where it will show the history page of the scanned building. When the building is in the frame of the camera, then click the button above the camera frame to start scanning (See Fig. 15).


Fig. 15 Building Scan Page in the Historical Building Scanning Application

Fig. 15 informs the result from other methods than scanning real-life historical building by choosing a picture that contains the historical buildings in the application. After choosing the picture, the application will perform the scanning process and show the result by displaying the history page of the scanned building. The picture must be a .jpg-type file.

Fig. 16 Choosing a .jpg Picture Type File in the Historical Building Scanning Application.

## C. *Evaluation*

The testing query produced a graph or table displaying the mean accuracy of image retrieval CNN, the maximum accuracy of image retrieval CNN, the total number of queries with greater than 70% accuracy retrieval, and which historical building has the highest accuracy. The previous graph and table are used twice to demonstrate the testing results when using Euclidean Distance and another time when using Manhattan Distance. The resulting data will be used to determine which CNN model is best for detecting historical buildings using either Euclidean Distance or Manhattan Distance. Most models show that using Euclidean Distance improves image retrieval accuracy more than using Manhattan Distance. For example, ResNet50V2 and

DenseNet201 have a better mean accuracy value on all pre-process image types when using Euclidean Distance than Manhattan Distance. InceptionV3 has a better result when using the Manhattan method, while InceptionResNetV2 and VGG19 have a better result when detecting blur pre-process image type, and CLAHE pre-process image type using the Manhattan method, but a better result when detecting SUCK pre-process image type using Euclidean method. DenseNet201 has the highest mean accuracy value in all pre-process image types. DenseNet201 has a better result when using the Euclidean method. Table 1. shows a more detailed comparison of mean accuracy image retrieval using Euclidean Distance and Manhattan Distance.

TABLE I
COMPARISON OF MEAN ACCURACY IMAGE RETRIEVAL CNN MODELS USING EUCLIDEAN METHOD AND MANHATTAN METHOD

| **CNN Model** | **Euclidean Method** | | | **Manhattan Method** | | |
|---|---|---|---|---|---|---|
| | **BLUR** | **CLAHE** | **SUCK** | **BLUR** | **CLAHE** | **SUCK** |
| ResNet50V2 | 85.1% | 84.7% | 85.4% | 85% | 84.7% | 84.9% |
| InceptionResNetV2 | 84.7% | 83.4% | 83.9% | 85.9% | 83.8% | 83.1% |
| VGG19 | 84.8% | 83.7% | 84.1% | 85.1% | 84% | 84% |
| InceptionV3 | 85.8% | 84.3% | 84.7% | 86% | 85.2% | 85.1% |
| DenseNet201 | 88.9% | 87.6% | 88.2% | 88.4% | 87.6% | 87.8% |

Table 1 shows the maximum accuracy of image retrieval CNN models data when it used the Euclidean Distance and Manhattan Distance for the testing query data. The table shows that most of the maximum accuracy between the Euclidean method and the Manhattan method is the same, except for VGG19 when detecting blur pre-process images and InceptionV3 when detecting CLAHE pre-process images. The Euclidean method outperforms the Manhattan method, which has a maximum accuracy of 100%.

TABLE II
COMPARISON OF MAX ACCURACY IMAGE RETRIEVAL CNN MODELS USING EUCLIDEAN METHOD AND MANHATTAN METHOD

| CNN Model | Euclidean Method | | | Manhattan Method | | |
|---|---|---|---|---|---|---|
| | BLUR | CLAHE | SUCK | BLUR | CLAHE | SUCK |
| ResNet50V2 | 92.3% | 92.3% | 92.3% | 92.3% | 92.3% | 92.3% |
| InceptionResNetV2 | 92.3% | 94.8% | 92.3% | 92.3% | 94.8% | 92.3% |
| VGG19 | 100.0% | 92.3% | 92.3% | 92.3% | 92.3% | 92.3% |
| InceptionV3 | 92.3% | 100.0% | 94.8% | 92.3% | 97.4% | 94.8% |
| DenseNet201 | 92.3% | 92.3% | 92.3% | 92.3% | 92.3% | 92.3% |

Table 2 informs that the highest number of total queries that exceed 70% accuracy in image retrieval is 148 images, owned by DenseNet201 when detecting blurred pre-process images using the Manhattan method. When using the Manhattan method, most CNN models have a better result than the Euclidean method. DenseNet201 has the highest overall number of total queries above 70% accuracy when detecting all pre-process image types. DenseNet201 excels in both the Euclidean and the Manhattan methods compared to the other CNN models, while VGG19 has the lowest number of total queries that are above 70% accurate in both methods.

TABLE III
COMPARISON OF TOTAL QUERY ABOVE 70% ACCURACY BETWEEN CNN MODELS USING EUCLIDEAN METHOD AND MANHATTAN METHOD

| **CNN Model** | **Euclidean Method** | | | **Manhattan Method** | | |
|---|---|---|---|---|---|---|
| | **BLUR** | **CLAHE** | **SUCK** | **BLUR** | **CLAHE** | **SUCK** |
| ResNet50V2 | 115 | 99 | 94 | 103 | 96 | 87 |
| InceptionResNetV2 | 101 | 95 | 86 | 97 | 97 | 90 |
| VGG19 | 86 | 80 | 83 | 77 | 76 | 72 |
| InceptionV3 | 114 | 110 | 100 | 119 | 116 | 104 |
| DenseNet201 | 140 | 130 | 133 | 148 | 137 | 139 |

Table 3 shows that the historical buildings that have the highest image retrieval accuracy are all from Bandung City. Whether the testing query uses the Euclidean or the Manhattan method, the historical buildings in the top ten highest accuracies are the historical buildings in Bandung City. The buildings from ranking one to ten have the same accuracy value. The test query results show that DenseNet201 is the most suitable CNN model to detect historical buildings. DenseNet201 has the highest image retrieval mean accuracy in all three pre-process image types, whether it uses Euclidean Distance or Manhattan Distance. DenseNet201 also has the highest number of total queries that exceed 70% accuracy with Euclidean Distance and Manhattan Distance. However, InceptionV3 has the highest max accuracy value in all pre-process image types, especially when using the Euclidean method, which reaches 100% max accuracy when detecting CLAHE pre-process images.

## IV. CONCLUSION

This study represents a significant advancement in recognizing and detecting historical buildings in the Bandung, Bogor, and Cirebon areas using deep learning via several CNN models supported by retrieval methods, namely Euclidean and Manhattan Distance. In recognizing historical buildings, an initial process is performed to make the image clearer, brighter, and free of image noise. The methods used are Gaussian Blur, SuCK, and CLAHE.

According to the results of the experiment, the highest average retrieval accuracy carried out using the Blur method for the pre-process and the feature extraction process with the Densenet121 and Inception V3 model was supported by the retrieval process using the Euclidean Method, namely 88.9%, and 85.8%. The results of the tested CNN models could be used in a Tkinter-based User Interface to predict a real-time capture picture or any picture containing one of the 61 historical buildings trained to recognize them.

## REFERENCES


[1] E. Andiyan, A., & Budianto, "Penerapan Konsep Arsitektur Kontemporer pada Penataan Cagar Budaya Situ Tasikardi," vol. 6, no. 6, p. 6, 2021, [Online.]

[2] K. P. dan Kebudayaan., "Undang-undang Republik Indonesia Nomor 11 Tahun 2010 Tentang Cagar Budaya," *Undang. Republik Indones. Nomor 11 Tahun 2010 Tentang Cagar Budaya*, p. 13, 2011.

[3] K. D. Rahmat, "Pelestarian Cagar Budaya Melalui Pemanfaatan Pariwisata Berkelanjutan," Jurnal Pariwisata Terapan, vol. 5, no. 1, p. 26, Nov. 2021, doi: 10.22146/jpt.58505.

[4] S. Rahardjo, "Beberapa Permasalahan Pelestarian Kawasan Cagar Budaya Dan Strategi Solusinya," Jurnal Konservasi Cagar Budaya, vol. 7, no. 2, pp. 4–17, Mar. 2013, doi:10.33374/jurnalkonservasicagarbudaya.v7i2.109.

[5] R. Parhani, "Manajemen Pengelolaan Objek Wisata Kota Tua Jakarta Berbasis Masyarakat," p. 6, 2016.

[6] M. F. Rizqullah and W. Swasty, "Perancangan Media Informasi Kota Tua Jakarta Utara Melalui Sign System Yang Terintegrasi Website," Andharupa: Jurnal Desain Komunikasi Visual & Multimedia, vol. 5, no. 02, pp. 210–225, Sep. 2019, doi: 10.33633/andharupa.v5i2.1957.

[7] d. R. Aditia, "klasterisasi pengenalan situs bersejarah kota magelang berbasis web dengan menggunakan algoritma k- medoid," 2019.

[8] Y. N. Tonapa, S. M. , Dwight M. Rondonuwu, and S. M. , Dr. Aristotulus E. Tungka, "Kajian Konservasi Bangunan Kuno Dan Kawasan Bersejarah Di Pusat Kota Lama Manado," *Spasial*, vol. 2, no. 3, pp. 121–130, 2015.

[9] L. Moya, F. Yamazaki, W. Liu, and M. Yamada, "Detection of collapsed buildings from lidar data due to the 2016 Kumamoto earthquake in Japan," Natural Hazards and Earth System Sciences, vol. 18, no. 1, pp. 65–78, Jan. 2018, doi: 10.5194/nhess-18-65-2018.

[10] C. Hörr, E. Lindinger, and G. Brunnett, "Machine learning based typology development in archaeology," Journal on Computing and Cultural Heritage, vol. 7, no. 1, pp. 1–23, Feb. 2014, doi:10.1145/2533988.

[11] Md. Samaun Hasan et al., "Heritage Building Era Detection using CNN," IOP Conference Series: Materials Science and Engineering, vol. 617, no. 1, p. 012016, Sep. 2019, doi: 10.1088/1757-899x/617/1/012016.

[12] X. Zhao, Z. Zou, N. Wang, and P. Zhao, "Feature recognition and detection for ancient architecture based on machine vision," Smart Structures and NDE for Industry 4.0, Mar. 2018, doi:10.1117/12.2296543.

[13] M. E. Hatir, M. Barstuğan, and İ. İnce, "Deep learning-based weathering type recognition in historical stone monuments," Journal of Cultural Heritage, vol. 45, pp. 193–203, Sep. 2020, doi:10.1016/j.culher.2020.04.008.

[14] Y. K. Yi, Y. Zhang, and J. Myung, "House style recognition using deep convolutional neural network," Automation in Construction, vol. 118, p. 103307, Oct. 2020, doi: 10.1016/j.autcon.2020.103307.

[15] O. S. Laptev and I. I. Bikmullina, "Sightseeing Application Based on Location Marking and Convolutional Neural Network Building Recognition," 2020 International Russian Automation Conference (RusAutoCon), Sep. 2020, doi:10.1109/rusautocon49822.2020.9208062.

[16] L. Zheng, P. Ai, and Y. Wu, "Building Recognition of UAV Remote Sensing Images by Deep Learning," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Sep. 2020, doi: 10.1109/igarss39084.2020.9323322.

[17] Y. Zeng, Y. Guo, and J. Li, "Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning," Neural Computing and Applications, vol. 34, no. 4, pp. 2691–2706, May 2021, doi: 10.1007/s00521-021-06027-1.

[18] E. Maltezos, N. Doulamis, A. Doulamis, and C. Ioannidis, "Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds," Journal of Applied Remote Sensing, vol. 11, no. 04, p. 1, Dec. 2017, doi: 10.1117/1.jrs.11.042620.

[19] N. Wang, X. Zhao, P. Zhao, Y. Zhang, Z. Zou, and J. Ou, "Automatic damage detection of historic masonry buildings based on mobile deep learning," Automation in Construction, vol. 103, pp. 53–66, Jul. 2019, doi: 10.1016/j.autcon.2019.03.003.

[20] M. E. Hatir, M. Barstuğan, and İ. İnce, "Deep learning-based weathering type recognition in historical stone monuments," Journal of Cultural Heritage, vol. 45, pp. 193–203, Sep. 2020, doi: 10.1016/j.culher.2020.04.008.

[21] N. Wang, X. Zhao, Z. Zou, P. Zhao, and F. Qi, "Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning," Computer-Aided Civil and Infrastructure Engineering, vol. 35, no. 3, pp. 277–291, Aug. 2019, doi: 10.1111/mice.12488.

[22] Y. Xiong, Q. Chen, M. Zhu, Y. Zhang, and K. Huang, "Accurate Detection of Historical Buildings Using Aerial Photographs and Deep Transfer Learning," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Sep. 2020, doi: 10.1109/igarss39084.2020.9323541.

[23] M. Sun, F. Zhang, F. Duarte, and C. Ratti, "Understanding architecture age and style through deep learning," Cities, vol. 128, p. 103787, Sep. 2022, doi: 10.1016/j.cities.2022.103787.

[24] N. Ibrahim, A. ElFarag, and R. Kadry, "Gaussian Blur through Parallel Computing," Proceedings of the International Conference on Image Processing and Vision Engineering, 2021, doi: 10.5220/0010513301750179.

[25] Munesh Singh Chauhan, "Optimizing Gaussian Blur Filter using CUDA Parallel Framework," no. May, p. 442, 2018.

[26] J. Ma, X. Fan, S. X. Yang, X. Zhang, and X. Zhu, "Contrast Limited Adaptive Histogram Equalization-Based Fusion in YIQ and HSI Color Spaces for Underwater Image Enhancement," International Journal of Pattern Recognition and Artificial Intelligence, vol. 32, no. 07, p. 1854018, Mar. 2018, doi: 10.1142/s0218001418540186.

[27] P. Lu and Q. Huang, "Robotic Weld Image Enhancement Based on Improved Bilateral Filtering and CLAHE Algorithm," Electronics, vol. 11, no. 21, p. 3629, Nov. 2022, doi: 10.3390/electronics11213629.

[28] N. Kaveri and N. B. Mustare, "Modified CNN Model for Malaria Diagnosis," *J. Posit. Sch. Psychol.*, no. July, 2022, [Online]. Available: https://www.researchgate.net/publication/361842000_Modified_CNN_Model_for_Malaria_Diagnosis.

[29] J. L. Suárez-Díaz, S. García, and F. Herrera, "A Tutorial on Distance Metric Learning: Mathematical Foundations, Algorithms, Experimental Analysis, Prospects and Challenges (with Appendices on Mathematical Background and Detailed Algorithms Explanation)," 2018, [Online]. Available: http://arxiv.org/abs/1812.05944.

[30] M. K. Gupta and P. Chandra, "Effects of similarity/distance metrics on k-means algorithm with respect to its applications in IoT and multimedia: a review," Multimedia Tools and Applications, vol. 81, no. 26, pp. 37007–37032, Sep. 2021, doi: 10.1007/s11042-021-11255-7.

[31] R. Suwanda, Z. Syahputra, and E. M. Zamzami, "Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K," Journal of Physics: Conference Series, vol. 1566, no. 1, p. 012058, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012058.

[32] A. Hafidz, D. I. Mulyana, D. B. Sumantri, and K. S. Nugroho, "Identification of Buni Fruit Image Using Euclidean Distance Method," SinkrOn, vol. 7, no. 2, pp. 392–398, Mar. 2022, doi: 10.33395/sinkron.v7i2.11333.

[33] T. M. Ghazal et al., "Performances of K-Means Clustering Algorithm with Different Distance Metrics," Intelligent Automation & Soft Computing, vol. 29, no. 3, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[34] G. Feng, Z. Jiang, X. Tan, and F. Cheng, "Hierarchical Clustering-Based Image Retrieval for Indoor Visual Localization," Electronics, vol. 11, no. 21, p. 3609, Nov. 2022, doi: 10.3390/electronics11213609.

[35] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person Re-identification: Past, Present and Future."

[36] P. Marchwica, M. Jamieson, and P. Siva, "An Evaluation of Deep CNN Baselines for Scene-Independent Person Re-Identification."

[37] F. A. Irawan and D. A. Rochmah, "Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19," Jurnal Informatika, vol. 9, no. 2, pp. 148–158, Oct. 2022, doi:

10.31294/inf.v9i2.13257.
[38] V. R. Joseph and A. Vakayil, "SPlit: An Optimal Method for Data Splitting," Technometrics, vol. 64, no. 2, pp. 166–176, Jun. 2021, doi: 10.1080/00401706.2021.1921037.
[39] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation," Expert Systems with Applications, vol. 95, pp. 43–56, Apr. 2018, doi: 10.1016/j.eswa.2017.11.028.
[40] A. H. Rangkuti, V. H. Atthala, E. Tanuar, and J. M. Kerta, "Performance Evaluation of traditional Clothes pattern retrieval with CNN Model and Distance Matrices," pp. 1–9, 2020.