Applying Deep Learning and Convolutional Neural Network System to Identity Historic Buildings: The "Little China" Building in Central Java, Indonesia

Sumaiyah Fitrian Dini[1], Eri Prasetyo Wibowo[2], Mohammad Iqbal[3], Yudi Nugraha Bahar[4], Affa Alfiandy[5]

[1] Faculty of Civil Engineering and Planning, Department of Architecture, Gunadarma University, Depok, Indonesia.

[2&3] Computer Science Faculty, Gunadarma University, Jakarta, 16424, Indonesia

[4] Faculty of Civil Engineering and Planning, Department of Architecture, Gunadarma University, Depok, Indonesia.

[5] Computer Science Faculty, Gunadarma University, Jakarta, 16424, Indonesia

Email: sumaandini@gmail.com

Abstract

Current technological improvements enable the design of systems to identify a building facade from a digital image through image processing. Indonesia has many historic buildings, one of which is the Lasem District, which has the nickname "Little China". Presently however, the beauty of the architecture of this building is under threat. It is known that many Chinese houses and other historic buildings also have damaged building facades. With these happening, there is a serious issue that the historic buildings could be lost and there may be no records of them.

This research produces a system to identify the condition of the facade of a historic building using a system known as the Convolutional Neural Network (CNN). Transfer learning method was applied to overcome the limitation of image data collection, which were only 260 images of building facades that became the data set. The images were obtained by taking photos directly using a digital camera from 6 villages in the historic area of Lasem. Transfer learning is carried out with the aim of duplicating a small dataset and the results of the trained model are used as additional "knowledge" in the second training process.

Deep learning research with image data sources that depend on primary data which is usually limited, transfer learning methods to increase knowledge in training, and they can be applied. As a result, the trained model tested with the test and validation datasets has an accuracy rate of 68% for the visual character classification and 60% accuracy for the damage level classification. In the future, there is a need to enrich this by providing datasets with relational characters to generate intelligence in the proposed learning model.

Keywords: CNN, building identification, visual character, damage, Lasem.

Introduction

The historical buildings that exist as cultural heritage must be preserved so that the current and future generations also get rich information from them. Conventionally, manual drawings or photographic records have been employed to records them. However, today, there are digital technologies that can do the same with better outcomes.

Recently, deep learning as a shape recognition method has been used in the field of architecture for building classification and recognition. Convolutional Neural Network (CNN) has also been used as a method of classifying the styles and functions of buildings in urban areas (Taoufiq, Nagy and Benedek, 2020), in identifying the construction era and period of ancient buildings (Hasan *et al.*, 2021). The classification is based on age, price and the number of floors in the villas and the apartment types (Muhammad and Serte, 2021). Deep learning-based frameworks for understanding architectural styles and eras employ deciphering building facades based on Google Street View images (Sun *et al.*, 2022).

Convolutional Neural Network is also applied as a method of automatic detection of defects and damages to buildings. For example, CNN has been applied for automatic detection of concrete cracks (Cha, Choi and Büyüköztürk, 2017). Evaluating the application of CNNs for automatic detection and localisation of defects (mould, damage and stains) from images (Perez, Tah and Mosavi, 2019) and automatic detection and localization of building defects (Su and Wang, 2020) are common. Ma *et al.* (2020) have applied YOLOv3 object detection method based on the Convolutional Neural Network (CNN) to locate collapsed buildings from the post-earthquake remote sensing imagery. CNN backbones have been considered and exploited to produce transfer learning effects of crack detection on the masonry surfaces carried out at the patch level (Dais *et al.*, 2021). Liu *et al.* (2022) has detected real-time building damage based on terrestrial images using an enhanced YOLOv5, with damage type annotations of debris, collapse, spalling, and cracks.

The studies described above have conducted the recognition of historical buildings through 2D and 3D images of buildings from both primary and secondary data collected, with an approach through deep learning with the basic CNN algorithms. The research has resulted in the recognition of urban and historic building facades for the purpose of classification of style, function, period, age, and building elements. Detection of structural cracks, types of defects, and brick wall cracks due to age and disasters have been successfully carried out by utilizing a CNN-based framework as a new detection method.

Lasem is known as a historical area in Central Java that has a variety of cultural heritage related buildings and is one of the heritage cities in Indonesia. Ancient houses are some of the many cultural heritage buildings scattered in the Lasem area. Chinese and Javanese culture is strongly reflected in the architectural forms of the buildings. The area, which has the nickname "Little China", is known to have problems; many houses are damaged because they have been abandoned by their owners and there are buildings that have been replaced with modern architectural models (Purwanto, 2018).

It is seen in this context that deep learning studies and their use in the fields of architecture and civil engineering could be employed to identify these damages to the buildings in Lasem. The objectives of this research are to develop an algorithm to identify the character and damage of cultural heritage building facades using the Convolutional Neural Networks (CNN), and produce a classification of the condition of cultural heritage buildings in the Lasem District. The results of this research are expected to contribute to policy makers, communities and building owners, as a reference for the rehabilitation and restoration of ancient buildings, and increase the value of cultural heritage areas.

Visual Character of Buildings

Krier states that the visual character of a building can be seen in the elements of the building facades (roof, walls, arcades, doors & windows), elements of the building interior spaces (interior walls, doors and windows, floors, columns, and ceilings), and the building mass (Krier, 1983). Nelson mentions the characters that are visible on the entirety of the building: roofs, openings, intersection projections, materials, settings, material details, ornaments, individual spaces, similar spaces, surface finishes, and exposed structures. Parolek complements some of these architectural elements with elements of the building mass, the facade configuration, windows and doors, elements and details, as well as colour and material combinations (Hilmy, Sardjono and Pandelaki, 2019). Facade configuration elements that can shape the character of a building are space opening elements, facade constituent planes, application of dominant facade materials, facade types and finishes and colour processing techniques (Sastra, 2016).

The word facade comes from French. It is taken from the Italian *facciata* or *faccia*. Faccia is taken from the Latin word *facies*. In its development, it became face in English and in the field of architecture the facade means the face of the building. Thus, it becomes very clear that the facade or the front of a building is an integral part of a building. The facade is the exterior part of the building that faces the outside. The main facade is usually the side of the building with the main entrance and has a specific style (Purnomo, 2019).

Chinese courtyard residential elements can be classified into three categories, namely the gate, the roof, and the building facade. The type of gate in each Lasem Chinese village dwelling can be distinguished based on the style and the materials used. The Chinese style in the gates of the Chinatown residences in the Lasem area is quite dominant, where almost all buildings have gates both in large and small sizes and in a Chinese style (Duhita, 2019).

Facade configuration elements can shape the character of a building, and the physical integrity of a building facade greatly affects the appearance and character of a building. The less the forming elements, the more the character fades. However, a loss of elements or damage to a building facade can affect the integrity of a building character. Both elements are very important to be recorded for the sake of restoration and reconstruction of the historic buildings.

Obtaining the Data Set

Based on the theory of identifying the visual character of buildings and the architecture of the Chinese building facades in Lasem from previous studies, an architectural analysis in the preparation of data sets is carried out by identifying elements of building facades which include roofs, columns, exterior walls, openings and gates. The process of analyzing building image data was carried out based on building architecture studies and was validated by the Lasem researchers who are administrators of the Indonesian Architecture Association (IAI) of the Central Java Province and the Indonesian Built Environment Research Association (IPLBI). Data validation is carried out for data labeling which are used as training data.

Sources of data in this study were obtained from the research location directly by documenting buildings from the Lasem City. Data labeling is done to produce training data, where all documented images are first identified and sorted, so as to produce a data set. The data set consists of 260 collections of building facade images which were photographed directly using digital cameras and smartphones. These images were taken from the historic city of Lasem which is in Soditan, Karangturi, Babagan, Sumbergirang, Gedungmulyo and the Dorokandang Village.

Secondary data is also generated by adding images from the Google Street view to enrich the data. Data labeling was carried out by analyzing the data based on two research analysis variables, namely the visual character and the condition of the facade identified from the shape and condition of the building facade elements, such as the gates, fences, roofs, exterior walls, exterior columns, doors and the windows.

The gate type of each Chinese Lasem courtyard residence can be distinguished based on the style and the materials used. The Chinese style at the gates of the Chinese courtyard residence in the Lasem area is quite dominating, where almost all buildings have gates in the Chinese style both in large and small sizes (Duhita, 2019).

The condition of the houses that have become the data set are on the neighborhood roads and alleys with a road width of two meters to six meters. The shapes of the Lasem Chinese houses are dominated by a complex with high fences and gates. The image data of the building facades are limited to the range where the camera is located and the possibility to photograph with an angle view in front of the building. Based on these limitations, the facade elements of the historical Chinese houses from the front view of the building can be taken which are dominated by the gates and fences seen as the elements that make up the facade of the building.

The residential character of Lasem's Chinese courtyard when captured from the outside area (street corridor) is a row of white walls of two meters height with repeated Chinese roofs at the beginning of each plot that functions as the main gate.

The Chinese courtyard residential elements can be classified into three categories: gates, roofs and building facades. The following are examples of the three types of gates found in the Lasem Chinese settlement, namely a temple gate, a small gate and a large gate.

(a) (b) (c)

Fig. 1: (a) Small gate, (b) Large gate (House gate), (c)Temple gate.
Source: Author

Labeling data is done by categorizing images based on the style of a building which refers to the architectural research of the Lasem building known to have a form of Chinese, Javanese, Colonial and mixed architecture. Some of the typical building styles in the Karangturi village Lasem include the Temple, Chinese Laseman, Chinese Vernacular, Javanese and Chinese-Javanese Acculturation. The large wall of the Chinese house in Lasem has a gateway in the middle of the wall with attractive colours and shapes (Kuasa and Wuryanto, 2017).

Lasem Chinese architectural style can be seen from the back of a curved roof: the tip of the back of the hill is tapered like a wallet tail and on the wall under the roof ridge, there is a carving of Chinese characters. Javanese culture elements, reflected in a symmetrical layout and a pavilion in front of the dwellings (Sudarwani, Purwanto & Rukayah, 2022).

Fig. 2: The building forms in Lasem that provided the basis for the research data.
Source: Author

Wall is one part that is vulnerable to moisture, both from the vertical and horizontal directions (Pratiwi, Wijayanto & Putri, 2021). Bricks are susceptible to weathering when compared to other building materials, due to their high porosity values between 11-40% in the contemporary buildings and 30-38% in the historical buildings (Martinez *et al.*, 2016). A substantial increase in wetness can be seen visually through wet, decaying walls that experience cracking, flaking and micro-organism growth (U.S. Environmental Protection Agency, 2013).

(a) (b)

Fig. 3: (a) Facade Conditions in Small Gate House buildings,
(b) Facade Conditions in Large Gate House buildings.
Source: Author

Damage detection is carried out on the walls which are the constituent elements of the building facade because a wall is a part that is vulnerable to damage due to both climate and time influences. The damages to cultural heritage buildings have been caused by the shrinking of the aged building, human activities, or natural disasters. In general, the level of damages to cultural heritage buildings can be classified into three, namely: light damage, moderate damage, and heavy damage. The damage assessment to the cultural heritage building was carried out to look at two interrelated things, namely damages to the building as a whole and damages to the main physical attributes, in the context of protecting the value and physical character of cultural heritage buildings. The identification of the building damage level refers to the Regulation of The Minister of Public Works and Public Housing Republic of Indonesia Number 19 of 2021. These are the Technical Guidelines for The Implementation of a Preserved Cultural Heritage Building and building damage category based on the building damage survey guidebook. It is also based on the Regulation of the Minister of Public Works of the Republic of Indonesia accessed from the website (https://simantu.pu.go.id/content/?id=4134).

Methods

A typical CNN architecture consists of a series of layers such as convolution, fully connected pooling, and logistic (softmax regression). Additional layers such as dropout and batch normalization can also be added to avoid overfitting and increase the generalizability of the model (Ji, Liu & Buchroithner, 2018).

CNN-based segmentation works on three-dimensional images such as the height, width and the number of channels. When compared to the area or the cluster-based image segmentation, CNN-based segmentation has the advantage of providing the most precise results, although it has the disadvantage of taking a long time to train the model (Kaushik, 2019).

Two-dimensional image datasets with grayscale and RGB colour spaces are the most widely used datasets for semantic segmentation with the deep learning techniques. The main advantage of the deep learning techniques is the possibility to learn a feature representation suitable for the problem at hand (Garcia-Garcia et al., 2017)

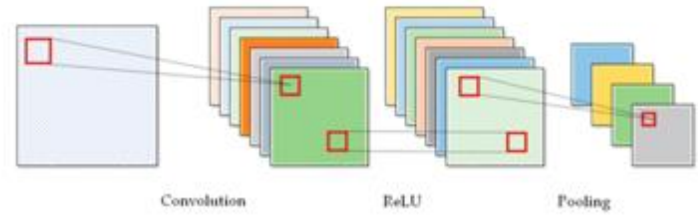


Fig. 4: CNN illustration with operation *max pool*
Source: Ji, Liu & Buchroithner, 2018

We formulate a good connection between the research that has been done and the research that will be done related to the identification of cultural heritage buildings with deep learning models. Therefore, we propose research on the identification of historical buildings through images using the Convolutional Neural Network model. We conduct the identification of cultural heritage buildings on two aspects, namely visual character and building condition, which have not been done together and become the focus of Taoufiq, Nagy and Benedek (2020), Hasan *et al.* (2021), Muhammad and Serte (2021), Sun *et al.* (2022), Cha, Choi and Büyüköztürk (2017), Perez, Tah and Mosavi (2019), Ma *et al.* (2020), Dais *et al.* (2021), and Liu *et al.* (2022).

The VGG-16 architecture is applied for training data processing based on the advantages of the VGG architecture in deep visual representations (Simonyan & Zisserman, 2014). This research proposal is applied to a case study in a historic area in Indonesia that has a variety of buildings with cultural heritage values, one of which is in the Lasem District.

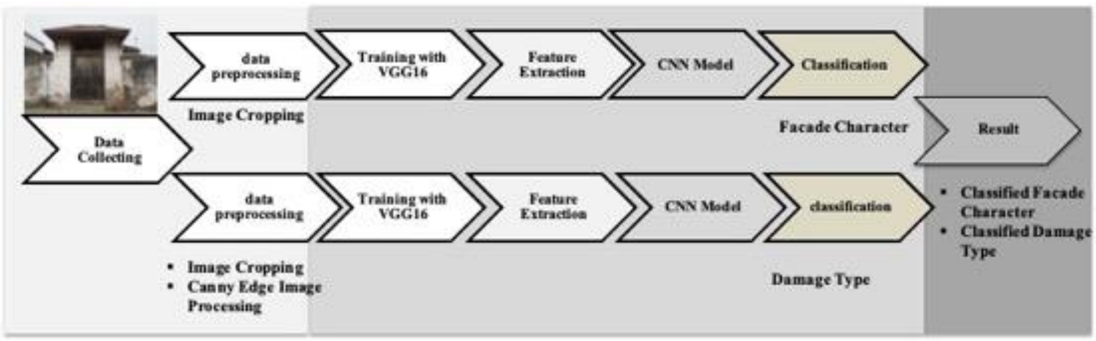


Fig. 5: CNN implementation scheme.
Source: Author

The image pre-processing consists of image cropping for the classification of facade visual characters. Image cropping along with Canny Edge for the damage classification, after pre-processing the data are used as training materials in CNN. Then, CNN will extract features from the images and look for a suitable formula in each layer so as to produce a model that can classify the images. The CNN model uses VGG-16 architecture as a digital image processing method to build facades to be applied to identify any damage to the historical building facades in the Lasem region. Images of the building facade are in RGB color space and in the PNG format. Image pre-processing consists of image cropping for the classification of façade character, and image cropping and canny edge for damage classification. Image pre-processing with a canny edge aims to get a pattern of damage to the building facade. After pre-processing, the data is used as training material in the Convolutional Neural Network (CNN). Then the model will extract features from the image and find a suitable formula in each layer so as to produce a model that can classify the image.

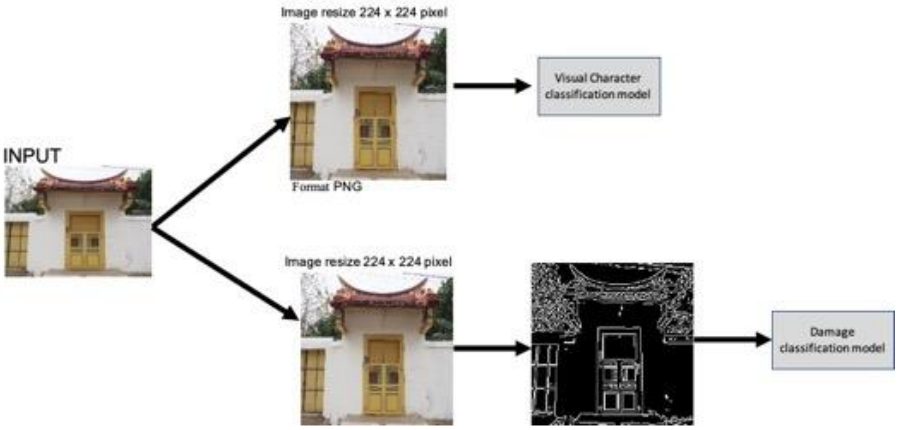


Fig. 6: Schematic of using Grayscale for classification and detection and Canny Edge to obtain the damage pattern of the building facade.
Source: Author

(a) (b)

Fig. 7: (a) Actual condition, (b) canny edge image processing
Source: Author

This program is run on Google Collaboration to run the Python Code. The first thing to do is to make a connection between Google Collaboration and the Google Drive for storage, then prepare the data using the CV2 library, NumPy, Keras and Tensor Flow. The Convolutional Neural Network (CNN) layer used for the detection and classification is a kernel with a size of 3x3, because the type of data used is a 2-dimensional image, and the VGG16 architecture uses this kernel size. The kernel is used to automatically extract useful features (automatically, without manual settings) from data points (images) to complete the object classification task. Max Pooling is used for the classification on the basis that the result of the pooling will be clearly visible lines that become the pattern of the identified image.

In the CNN architecture, Max Pooling was chosen, because the results of the pooling are clearly visible lines that will become the pattern of the identified image. Max Pooling is different from Average Pooling. When Max pooling is used, the pattern can be seen more clearly. Due to the limited data set amounting to 260 image data sets, the transfer learning method is carried out, so that the results of the trained model will be used as an additional "knowledge" in the second training process using the "Model Checkpoint" function. The processing of training data uses GPU for training data because the processed data are in the form of images. Thus, it requires more power.

The models that have been made are compiled to finalize them so that they are ready to use. For training, the Adam Optimizer is used because the results are better than the other optimization functions. It requires fewer parameters than the other functions so that the computation time is less. The loss function suitable for use is categorical cross entropy because in this case, it is used for classification. Using Adam as a training optimizer requires fewer iterations to reduce losses and save time. This test was carried out on the MNIST dataset (handwritten numbers 0-9) with a multilayer neural network and used a dropout layer (Kingma and Ba, 2015).

Early stopping in Keras is used to stop the training process before overfitting occurs between the training and validation. The checkpoint model is applied to save the best results during training and it opens the results of the training model that have been done as an additional knowledge.

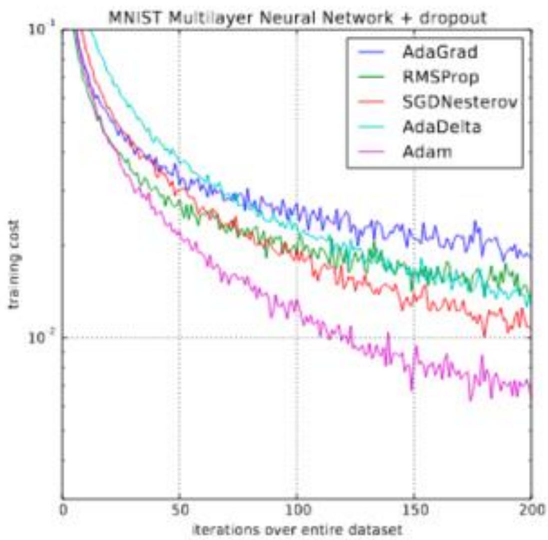


Fig. 8: Time difference between Adam optimizer function and other functions
Source: Kingma and Ba, 2015

Training, Test and Evaluation

In the training model for the classification of the visual character of building facades, the image data set that was successfully collected shows disproportionate characteristics. The dominance of the amount of image data is in the Chinese building styles with facade characters in the form of small gates and large gates, as well as Javanese buildings. Collections of images with other building styles such as colonial, modern and mixed are not in a quantity that is able to provide a balanced data set. Based on this, the transfer learning method is carried out, due to the limited dataset, so that the results of the model that has been trained will be used as additional "knowledge" in the second training process using the "Model Checkpoint" function.

The classification model consists of 2 steps: (1) Step Training: the model is built from the training data, (2) Test phase: a model is used to assign a label to unlabelled test data. The division of data groups was carried out for the training process by 70%, and training data and validation data by 30% from 260 data sets. The training data and test data used are facades of historical houses and traditional houses with a composition of 70% training data and 30% test data. In pre-processing the training data, the image resolution is changed to 224 x 224 pixels because the VGG architecture uses that resolution as the input. The class mode category is used because this process aims to classify the image. Next is to create the VGG16 architecture. The result of the VGG16 architecture shows the layers and parameters of each layer.

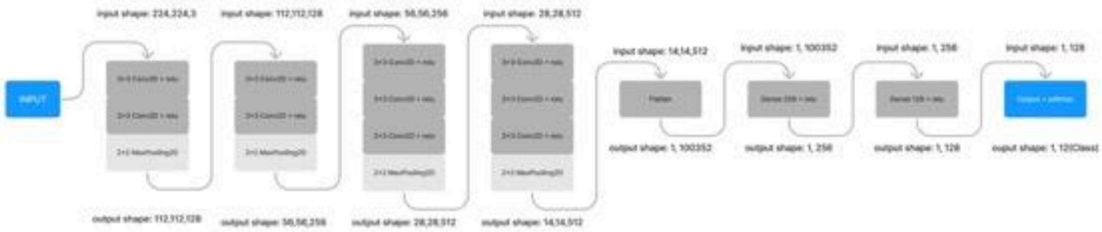


Fig. 9: VGG-16 architecture for building visual character classification.
Source: Author

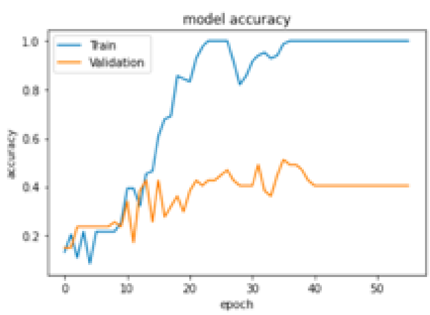
We evaluate the performance of trained classification models in machine learning based on the test patches. The performance evaluation of the classification method uses the confusion matrix parameter. The model evaluation uses accuracy metrics and the loss function calculation using the categorical cross-entropy. This is applied because the classification model is based on many classes, and the accuracy model is the easiest and fastest computation for multi-class classification needs. The multi-class classification problem is done by not calculating the overall F-1 score. The F-1 score is calculated per class in a one-to-one manner. This approach is done by assessing the success of each class separately, as if there is a different classifier for each class. The following is the formula for calculating precision and the F-1 Score.

$$Precision\ (P) = \frac{TP}{TP+FP}, \qquad (1)$$

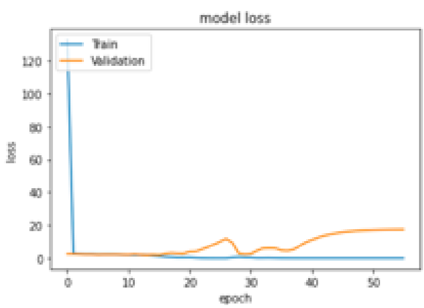
$$Recall\ (R) = \frac{TP}{TP+FP}, \qquad (2)$$

$$F1score = 2 * \frac{(precision*recall)}{precision*recall}, \qquad (3)$$

The training data model on the classification of building facade character types gets 100% accuracy for the training data in the first training process and 40% for the validation data. While in the results of the calculation of the loss function, there is a drastic decrease in the iterations (epochs) 1-3. In the next iteration process, the results of the loss function calculation are quite stable until the 50th iteration (epoch). In the next training process, the accuracy value is slightly increased but not much different from the first process. The calculation of the model accuracy uses metrics accuracy and the calculation of the loss function uses categorical cross entropy.

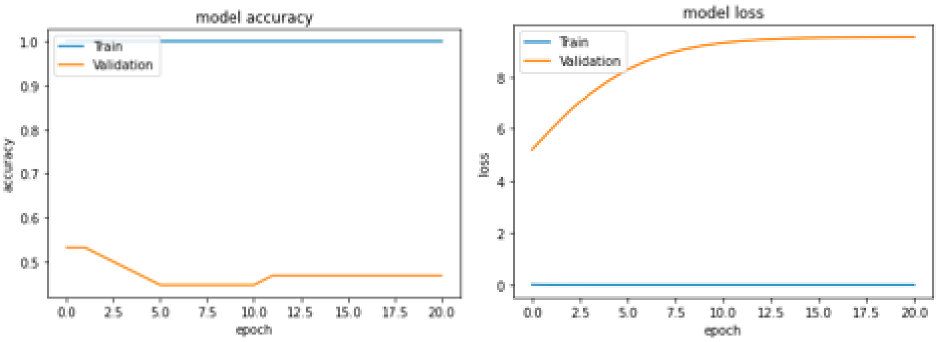


(a)



(b)

Fig. 10: (a) Accuracy model of training data for the category of building facade character types, (b) Results of the calculation of loss function training for the classification of facade character types.
Source: Author



(a) (b)

Fig. 11: (a) Training accuracy model for the visual character category of building data in the second training process with an accuracy of 68%, (b) The measurement results of the loss function for each iteration in the second training.
Source: Author

The data set that has been prepared is a collection of images of building facades from direct locations that have been analyzed based on the damage experienced referring to the Regulation of the Minister of Public Works and Public Housing of the Republic of Indonesia Number 19 of 2021, concerning Technical Guidelines for the Implementation of a Preserved Cultural Heritage Building and building damage category based on the building damage survey guidebook based on the Regulation of the Minister of Public Works of the Republic of Indonesia. This data set is used as material to train the model so that the machine can learn and gain knowledge. With this training, the machine is expected to replace people to recognize and classify the condition of the building facade according to the knowledge that has been given.

The classification is applied in a straightforward manner by processing between the training data and the test data. Classification is closely related to prediction because this method creates a model called a prediction model that can map each set of variables to the target class and then use the model to assign the target values to the newly obtained model to assign target values to the newly obtained model data set.

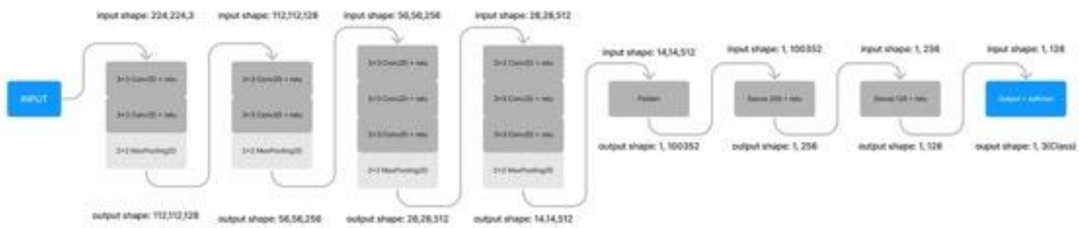
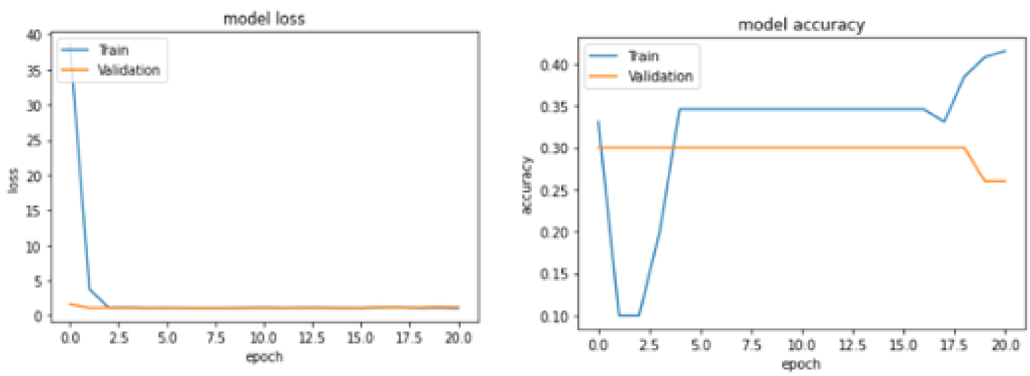


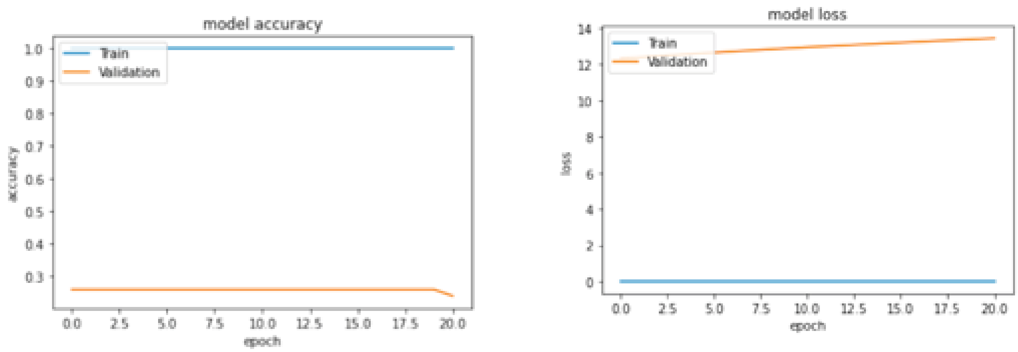
Fig. 12: VGG-16 architecture for building condition classification.
Source: Author

The training data model on the classification of types of building damages in the first training process gets a value of 40% accuracy on the training data and 30% using test data. In the results of the calculation of the loss function, the value decreases significantly until it is almost stable from the third iteration (epoch). In the seventh training process with the transfer learning method, the model already has the knowledge for the training process to continue the results of the previous training. The value of accuracy using the training data has touched 100% while the test data gets a figure of 25%-30%. In the loss function results, there is an increase to number 14 for validation data. To avoid an overfitting to the model, the iteration process is stopped automatically. The calculation of the model accuracy uses metrics accuracy and the calculation of the loss function uses categorical cross entropy.



(a) (b)

Fig. 13: (a) Accuracy model of training data for the category of building damage in the first training process, (b) measurement results of the loss function for each iteration of the first training.
Source: Author



(a) (b)

Fig. 14: (a) Accuracy model of training data for the category of building damages in the seventh training process with an accuracy of 60% when calculated using the F1-Score, (b)Measurement results of the loss function for each iteration of the seventh training.
Source: Author

Results

The first trial result of building facade character classification based on processing from training using VGG-16 obtained 66% accuracy with F1-score calculation. The results with the second training show the visualization of the test results, which is 68% when calculated with the F1-Score. These results were obtained after carrying out the training process twice. In the Heatmap table (which is colorful), it can be seen that if the composition of the numbers is greater the number on the diagonal line segment rather than vertical or horizontal, this indicates that the model is working better and the prediction is close to accurate.

The test results show the number of actual data and the prediction. This can be seen with the test results that show data including: (1) The actual number of Javanese character building facade images is 29, the model successfully classifies 25 Javanese character building facade images, (2) The actual number of Small Gate Chinese character building images is 20, the model successfully classifies 14 Small Gate Chinese character building facade images, (3) The actual number of images of building facades with Chinese character Big Gate is 18 images, the model successfully classifies 11 images of building facades with Chinese character Big Gate, (4) The actual number of images of buildings with Chinese-Colonial character is 18 images, the model successfully classifies 12 images of building facades with Chinese-Colonial character.

Fig. 15: The results of the prediction accuracy of 68% success in the 2nd process
Source: Author

Results of the identification of building damages based on the processing of the damage classification produced an accuracy value of 60% on the data as a whole dataset. This calculation used the F1-score formula. These results were obtained after conducting the training process seven times. The classification of building conditions from 260 building images were based on 4 classifications: namely good, lightly damaged, moderately damaged, and severely damaged. The test results show the following data: (1) The actual number of good condition building facade data is 2, but the prediction result is 0, (2) the actual number of lightly damaged building facades is 51, the prediction success is 23 images, (3) the actual number of moderately damaged building facade conditions is 137, the prediction success is 76, and (4) the actual number of heavily damaged building facade conditions 70, prediction success of 56.

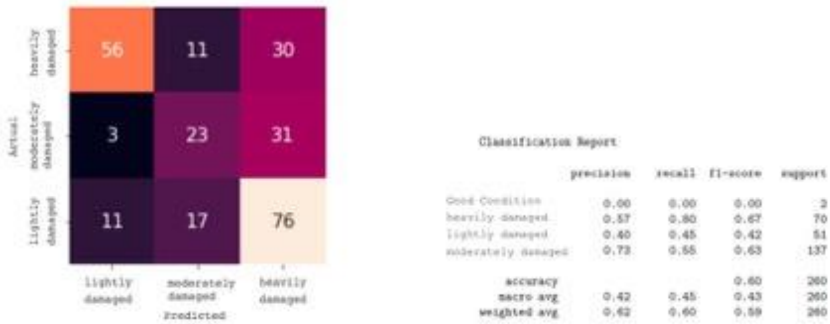


Fig. 16: Results of 60% success prediction accuracy
Source: Author

Conclusion

This paper presents a digital approach to the identification of the visual character and condition of historical buildings, using deep learning. We used a Convolutional Neural Network with VGG-16. Convolutional Neural Network (CNN) algorithm for the identification of visual character and condition of historic building facades use kernel layers in charge of automatic feature extraction. Data training techniques are added with transfer learning models aimed at enriching data knowledge, due to the limitations and disproportionate amount of data in providing character representation from the compiled data set. The application of canny edge-based segmentation is added to the VGG-16 algorithm for image processing with the aim of obtaining damage patterns in the building facade images.

The results of this trial have not received maximum accuracy even though in this scope, it has succeeded in getting an accuracy level above 50%. The author recognizes that the limited amount of data has significantly influenced training and testing processes. The identification of historical buildings using the Convolutional Neural Network model based on images in the historical area in Lasem obtained a classification with an accuracy of 68% for the visual character of the building and 60% for the level of building damage. We recognize that the limited amount of data and unbalanced data sets have significantly influenced training and testing processes. In the future, enrichment needs to be done by providing a collection of data with proportional characters, to produce intelligence in the proposal using a deep learning model.

Acknowledgements

The authors wish to thank Universitas Gunadarma who supported this paper under a PPS-PDD (Penelitian Pascasarjana-Penelitian Disertasi Doktor) 2022 grant, contract number 417/LL3/AK.04/2022, 11.3/LP/UG/VI/2022, on June, 20 2022. We would like to express our sincere gratitude to the Cultural Heritage Expert Team of the Rembang Regency Culture and Tourism Office for having provided the information and data regarding the historical buildings in Lasem, and to Ms. Margareta Maria Sudarwani from the Central Java Indonesian Architects Association and the Indonesian Association of Environmental Researchers who were willing to provide their helpful assistance for the validation of the research data. Any opinions, findings, conclusions, and recommendations are the authors' and do not necessarily reflect the sponsors

Funding

This research was supported by the National Competitive Grant Program, PPS-PD 417/LL3/AK.04/2022, 11.3/LP/UG/VI/202 and was funded by the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia.

References

Cha, Y. J., Choi, W. and Büyüköztürk, O. (2017) 'Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks', *Computer-Aided Civil and Infrastructure Engineering*, 32 (5), pp. 361–378. doi: 10.1111/mice.12263.

Dais, D. *et al.* (2021) 'Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning', *Automation in Construction*, 125(January). pp. 1-18. doi: 10.1016/j.autcon.2021.103606.

Duhita, D. (2019) 'Tipologi Courtyard pada Permukiman Tionghoa Lasem', *Jurnal Rekayasa Hijau*, 3 (1), pp. 47–60. doi: 10.26760/jrh.v3i1.2820.

Garcia-Garcia, A. *et al.* (2017) 'A Review on Deep Learning Techniques Applied to Semantic Segmentation', pp. 1–23. Available at: http://arxiv.org/abs/1704.06857.

Hasan, M. S. *et al.* (2021) 'Identification of Construction Era for Indian Subcontinent Ancient and Heritage Buildings by Using Deep Learning', in. Springer, Singapore, pp. 631–640. doi: 10.1007/978-981-15-5856-6_64.

Hilmy, M. F., Sardjono, A. B. and Pandelaki, E. E. (2019) 'Karakter Visual Arsitektur Bandara Di Jawa Tengah', *Nature: National Academic Journal of Architecture*, 6 (2), pp. 136. doi: 10.24252/nature.v6i2a4.

Ji, M., Liu, L. and Buchroithner, M. (2018) 'Identifying Collapsed Buildings Using Post-Earthquake Satellite Imagery and Convolutional Neural Networks: A Case Study of the 2010 Haiti Earthquake', *Remote Sensing*, 10 (11). pp. 1-20. doi: 10.3390/rs10111689.

Kingma, D. P. and Ba, J. L. (2015) 'Adam: A method for stochastic optimization', *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15.

Krier, R. (1983) *Element of architecture*. London: Architectural Design AD Publication Ltd

London.
Kuasa, T. A. and Wuryanto, G. S. (2017) ‘Gaya Rumah Tradisional Tionghoa Laseman sebagai Warisan Sejarah Arsitektur di Desa Karangturi, Lasem’, *Prosiding Seminar Nasional Energi Efficient For Sustainable Living*, pp. 145–156.
Liu, C. *et al.* (2022) ‘Real-Time Ground-Level Building Damage Detection Based on Lightweight and Accurate YOLOv5 Using Terrestrial Images’, *Remote Sensing*, 14 (12). pp. 1-21. doi: 10.3390/rs14122763.
Ma, H. *et al.* (2020) ‘Detection of collapsed buildings in post-earthquake remote sensing images based on the improved YOLOv3’, *Remote Sensing*, 12 (1). doi: 10.3390/RS12010044.
Martinez, P. *et al.* (2016) ‘Effectiveness of tetra-ethyl-ortho-silicate (TEOS) consolidation of fired-clay bricks manufactured with different calcination temperatures’, *Construction and Building Materials*. Elsevier Ltd, 106, pp. 209–217. doi: 10.1016/j.conbuildmat.2015.12.116.
Muhammad, M. and Serte, S. (2021) ‘Deep Learning Classification of Building Types in Northern Cyprus’, *International Journal of Computer and Information Technology(2279-0764)*, 10 (3), pp. 116–123. doi: 10.24203/ijcit.v10i3.98.
Perez, H., Tah, J. H. M. and Mosavi, A. (2019) ‘Deep learning for Detecting Building Defects Using Convolutional Neural Networks’, *Sensors (Switzerland)*, 19 (16), pp. 1–23. doi: 10.3390/s19163556.
Pratiwi, S. N., Wijayanto, P. and Putri, C. A. (2021) ‘Diagnosis of Capillary Rise in Heritage Building’, *IOP Conference Series: Earth and Environmental Science*, 780 (1). doi: 10.1088/1755-1315/780/1/012076.
Purnomo, D. (2019) ‘Tolis Ilmiah : Jurnal Penelitian Tolis Ilmiah : Jurnal Penelitian’, *Tolis Ilmiah; Jurnal Penelitian*, 1(2), pp. 124–129.
Purwanto, L. M. . (2018) ‘Kajian Penyelamatan Lasem dari Kepunahan Arsitektur Tionghoa’, pp. B023–B028. doi: 10.32315/ti.7.b023.
Ravi Kaushik, S. K. (2019) ‘Image Segmentation Using Convolutional Neural Network’, *International Journal Of Scientific & Technology Research*, 8(11), pp. 667–675. doi: 10.1109/ICCES45898.2019.9002121.
Sastra, S. (2016) ‘Kajian Estetika Bentuk Pada Fasade Perumahan Real Estate Di Yogyakarta’, *INformasi dan Ekspose hasil Riset Teknik SIpil dan Arsitektur*, 12(1), pp. 78–84. doi: 10.21831/inersia.v12i1.10355.
Simonyan, K. and Zisserman, A. (2014) ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. ICLR-2015, pp. 1–14. Available at: http://arxiv.org/abs/1409.1556.
Su, C. and Wang, W. (2020) ‘Concrete Cracks Detection Using Convolutional NeuralNetwork Based on Transfer Learning’, *Mathematical Problems in Engineering*, 2020. doi: 10.1155/2020/7240129.
Sudarwani, M.M. and Rukayah, R.S., (2022) 'The Survival of Chinatown Architecture: Lasem Chinatown, Indonesia.' ISVS e-journal (Journal of the International Society for the Study of Vernacular Settlements), 9 (2), pp.1-17.
Sun, M. *et al.* (2022) ‘Understanding architecture age and style through deep learning’, *Cities*. Elsevier Ltd, 128 (June), p. 103787. doi: 10.1016/j.cities.2022.103787.
Taoufiq, S., Nagy, B. and Benedek, C. (2020) ‘Hierarchynet: Hierarchical CNN-based urban building classification’, *Remote Sensing*, 12(22), pp. 1–20. doi: 10.3390/rs12223794.