2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI 2022)

Yogyakarta, Indonesia
8 December 2022

IEEE Catalog Number: **CFP22AAH-POD**
ISBN: **978-1-6654-5513-8**

Copyright © 2022 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number: CFP22AAH-POD
ISBN (Print-On-Demand): 978-1-6654-5513-8
ISBN (Online): 978-1-6654-5512-1

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Facial Expression in Tourism Destinations Using A Deep Learning Approach

1st Budi Dwi Satoto
Department of information system
University of Trunojoyo Madura
Bangkalan, East Java, Indonesia
budids@trunojoyo.ac.id

2nd Achmad Yasid
Department of Information System
University of Trunojoyo Madura
Bangkalan, East Java, Indonesia
ayasid@trunojoyo.ac.id

3rd Budi Irmawati
Department of informatics engineering
University of Mataram
Mataram, NTB, Indonesia
budi-i@unram.ac.id

4th Muhammad Yusuf
Department of information system
University of Trunojoyo Madura
Bangkalan, East Java, Indonesia
muhammadyusuf@trunojoyo.ac.id

5th Mohammad Syarief
Department of information system
University of Trunojoyo Madura
Bangkalan, East Java, Indonesia
mohammad.syarief@trunojoyo.ac.id

6th Siti Oryza Khairunnisa
Department of information system
Tokyo Metropolitan University
Tokyo, Japan
siti-oryza-khairunnisa@ed.tmu.ac.jp

2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-6654-5512-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISRITI56927.2022.10052854

***Abstract*— The development of the world of tourism is currently supported by information technology. New tourist attractions are introduced through social media. Apart from being easy and cheap to reach for visitors, tourist attractions must also be responsive to improved facilities and services. The step is through regular visitor surveys. One way that can be done is through visitors' facial expressions with the help of deep learning. In this research, the proposed contribution is using a convolutional neural network with 34 layers. The goal is to get good accuracy, but the computational burden of the training process is light. The image data used comes from secondary datasets to identify angry, disgusted, scared, happy, neutral, sad, and surprised classes with a total of 5,250 images. Data Augmentation technique is used to overcome the class imbalance. The results showed that the system could recognize facial expressions with an average accuracy of 99.38%. The average computational time for the training process to get the model is 25 minutes 23 seconds, with a testing time of 1-2 seconds. Experimental data testing results above 98%: error MSE 0.0445, RMSE 0.2110, and MAE 0.0150.**

Keywords*— *facial expression, tourism destination, deep learning, data augmentation, custom layer

I. INTRODUCTION

Handling tourism after the COVID-19 pandemic is one of the country's concentrations to increase income. The emergence of new tourism objects based on local wisdom also needs to be considered and evaluated periodically regarding the infrastructure and services provided [1]. The level of availability of age-appropriate play areas or parks at tourist sites is also needed to increase the reference level of visitor comfort. Public facilities such as road access, toilets, and places of worship are a concern because some visitors also pay attention to these conditions [2].

Research Improvement of tourism service facilities using machine learning has begun to be widely developed. The method uses photos of online tourist destinations in Australia as a data source. An artificial intelligence (AI) framework was used to identify tourist photos without human interaction. The step is to place 25 categories of environmental conditions from the images recorded at each tourist location. The results show that the deep learning process with convolutional neural networks gets the best image identification performance [3].

The importance of conducting a tourism customer service review is aimed at increasing the number of tourists. The data analysis seeks to understand the level of customer satisfaction and their demand for services. Although several studies have found factors that influence customer satisfaction in the tourist area, there are shortcomings regarding a large amount of social data and analysis of online tourist behavior. In addition, machine learning techniques in tourism have not explored the research of online customer reviews based on camera-captured analysis [4] [5].

Subsequent research explores the use of Extreme Learning Machine (ELM) to introduce facial micro-expressions of tourist visitors. The step is to compare the use of machine learning and deep learning. It begins with the search for the main features on the face, and ELM has a fast learning ability and higher performance than other models. In this case, machine learning techniques in micro-automatic facial expression recognition promise more effective results, saving time and resources. [6].

Furthermore, research with the theme of tourism uses artificial intelligence (AI) technology as a possible solution. Based on a systematic literature review, research on the antecedents and consequences of the encounter triad between customers, employees, and Artificial Intelligence is needed. This research identified four service modes: meeting, outcome, mediation, and facilitation. In addition, this research develops an integrated model to determine the factors that influence service encounters using AI technology and the customer service outcomes resulting from those encounters. It contributes to service management and AI applications both theoretically and practically [7].

Information and Communication Technology Research also has excellent potential to increase public awareness of the importance of Cultural Conservation. Providing tools to make historical site visitors more excited and enjoyable is essential. Interaction paradigms and innovative methods were developed to enable tour operators and cultural site guides to build scenarios. Domain experts help manage interactive IoT-based environments with information that stimulates emotion, understanding, and content use. It creates an Intelligent Interactive Experience by synchronizing related things on the observed object with pattern recognition and computer vision

Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.

techniques. Innovative Interactive Experience will adapt intelligent object behavior adequately [8] [9].

The above literature shows the importance of artificial intelligence in finding visitor responses through facial expressions. The gap with previous research is that deep learning does not require a geometric approach, for example, features of lip position or image matching with image refining to get a face match. Feature map layers in deep learning can recognize facial expressions through good training. In addition, higher accuracy can be obtained. In machine learning research, the selection of facial features is still the subject of an investigation by researchers. The secondary data used in this research comes from Kaggle. The size of the image stored to save repository space has a small measure of 48x48 pixels and needs to be enlarged to 224x224 pixels. In addition, additional variations are carried out using data augmentation techniques. Data sharing includes 80% training, 20% testing, and 20-80% training data is validation data. There is also experimental data taken outside of the training data. This research hopes that using the proposed 34-layer CNN design can reduce computational time during the training process.

II. LITERATURE REVIEW

A. Visitor response to tourist attraction

For local tourism owners, customers are an easily accessible and significant source of knowledge that often remains unused. This research explores the encounters of tourism business owners with customers, where the customer is used as an opportunity to learn throughout the journey. Companies can research the subjective perception of customers so that the value of content liked by users can be known. The findings show that the company needs the object of knowledge about the purpose of the scenario process designed with the whole experience [10]. It aims to get benefits such as quality meetings in promoting a product. The description concludes that in the context of learning what customers do, even though having a small scale provides opportunities for micro-tourism companies to engage in personal desires that their customers like and want to do. The method is to turn customers into participants based on data collected through participant observation, interviews, and user-generated content reviews. [11]

B. Use of Deep learning for tourism

Content created by tourism users has cognitive and emotional information, which is valuable data to build a destination image describing the tourist experience and destination assessment during the Tour. Multiple destination images can help tourism managers explore similarities and differences to investigate the elements of tourist interest and increase the competitiveness of destinations. The interactive and multi-level visual analysis enables understanding and analysis of cognitive themes and emotional experiences from various destination images [12].

C. Face recognition

It is a machine learning-assisted face recognition research method to extract features. The research implements a global averaging layer. Increasing the data in the pre-training dataset can improve the model's generalizability—the network performance test using the FER2013 emoticon dataset. Facial expression recognition accuracy is 68.4%. This method can make predictions or test for about 0.12 seconds.[13].

D. Relationship of tourist facilities services with facial expressions

Research on visitors' emotional responses to tourist facilities was conducted by exploring psychological mechanisms through facial expressions and questionnaires. Facial expressions convey more emotion because facial expressions are subordinate to desires, such as happiness, sadness, and anger. Self-report emphasizes the emotions evoked by stereotypes, especially disgust. The proposed model of emotion formation can interpret differences in emotional expression [14].

E. Convolutional Neural Network

Pattern recognition works by multiplying the input image feature matrix m x n with a kernel filter.

- The convolution layer is the core of feature maps. This layer generates a new image containing the multiplication of the features of the image inputted by the kernel.

$$h[x,y] = f[x,y] * g[x,y] \quad (1)$$

with $h[x,y]$=output convolution, $f[x,y]$=matrix image, $g[x,y]$=kernel filter, x,y=index.The filter applied is a matrix with a size of 3x3 or 5x5. This convolution process produces a feature map which will then be used during the activation layer [15]

- The batch Normalization Layer uses to reduce the covariance shift or equalize the distribution of each input value that is constantly changing due to changes in the previous layer during the training process.

$$output = \frac{W-N+2P}{S} + 1 \quad (2)$$

with W = width of Image, N = width of filter, P = Zero padding and S = Stride [16].

The batch normalization operation is performed before the activation function of each input layer. The process consists of a zero-center input process by calculating the mini-batch mean and minibatch variance and then normalizing by calculating the scale & shift [17].

- Relu Layer is used to change the negative value to 0. ReLu Activation function to replace the sigmoid activation function

$$ReLU\,(x) = f(x) = \begin{cases} x, & jika\ x \geq 0 \\ 0, & lainnya \end{cases} \quad (3)$$

With f(x)=output layer function, x=variable whose value is equal or more significant than zero [18].

- Polling Layer reduces the dimensions of the features obtained from the convolutional layer. There are two types of pooling layers: max pooling and average pooling, as shown in Fig. 1 [19].

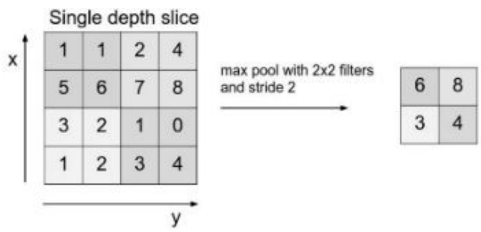


Fig. 1. Max Pooling layer

Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.

Average pooling is getting the average value of the matrix value of the part of the image covered by the convolution process. [20].

F. Architecture of CNN

Sequence layer designs are arranged so that the best combination is obtained. Several architectures are often used, including alexnet, squeezenet, xception, googlenet, VGG, resnet, densenet, and customnet [21]

G. Augmentation data

In some cases, it can be applied to unbalanced datasets, but the point is to use augmentation to make variations:

$$A = \begin{bmatrix} 1 & s_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

With A=matrix kernel shears vertical, s_v=value shears. The result is $x' = x + s_v y$ and $y' = y$ [22]

H. Confusion matrix

The system (model) generated by the training process is used to compare with the actual classification results so that the Confusion matrix provides class error information from the classification results.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

TP (true positive) = number of expression records classified as valid, TN (true negative) = number of non-expression records classified as non-facial expression class [23].

$$Sensitifity/\,Recall = \frac{TP}{TP+FN} \quad (5)$$

FP=False Positive. The population whose class results match when the population is not that class. FN=False Negative means that the results of the tests carried out do not match the type but are included in that class[24]. Error calculations are using MSE (Mean Square Error)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \quad (7)$$

RMSE (Root Mean Square) and Mean Absolute Error (MAE) [25]

$$MAE = \frac{1}{n}\sum_{j=1}^{n} \left|y_j - \hat{y}_j\right| \quad (6)$$

with N=number of sample data, i=indexing, y_i=actual label, and $\tilde{y}_i$ Label prediction [26].

III. RESEARCH METHODS

At this stage, planning for the method will be made as follows:

A. Dataset

This research collects data through direct fieldwork and secondary data sets from kaggle.com. The dataset file has a size of 48x48 pixels with a resolution of 300 dpi. It is too small, so it requires preprocessing to be 224x224x3 with 96 dpi with 24-bit depth, which means RGB. Each folder contains 750 images; there are seven classes: angry, disgusted, fearful, happy, neutral, sad, and surprised. There is a total of 5,250 images. There are black-and-white images that must then be converted to RGB. Dataset link:https://www.kaggle.com/datasets/mahmoudima/mma-facial-expression

B. Research Methods

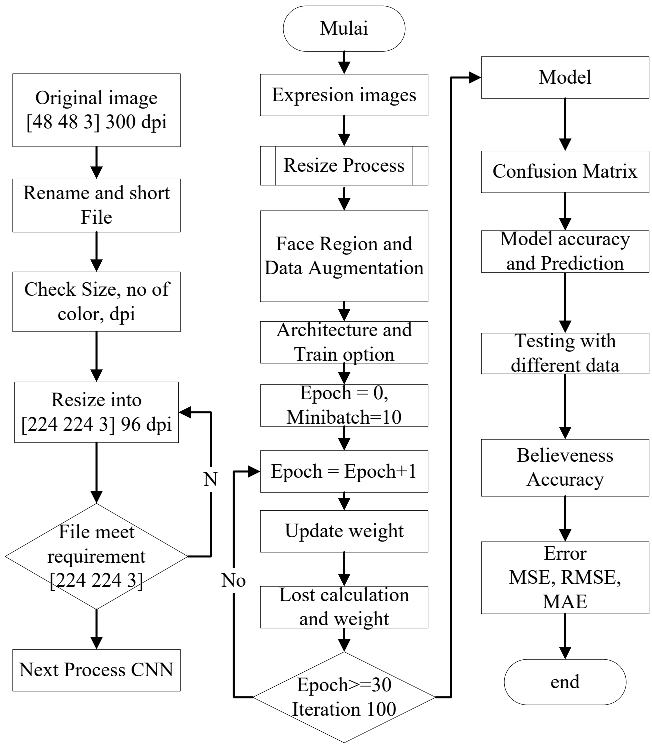


Fig. 2. Research Methods

To save image storage space, the kaggle repository stores images in the size of [48 48 3] 300 dpi, so this data needs to be restored first to [224 334 3] 96 dpi. After all the files are resized, they can be used for training, data augmentation, testing, and obtaining models.

C. Hardware dan Software

Hardware and software used in this research are Core i-7 Laptop, 12GB Memory, Nvidia GTX1050 4GB, Windows 10 software, and Matlab 2020a.

D. Design Architecture 34-Layer

The deep learning implementation uses a combination of feature maps and classification with 34 layers, where the layer combination adjusts the number of kernel filters when convoluting, as shown in Fig. 3.

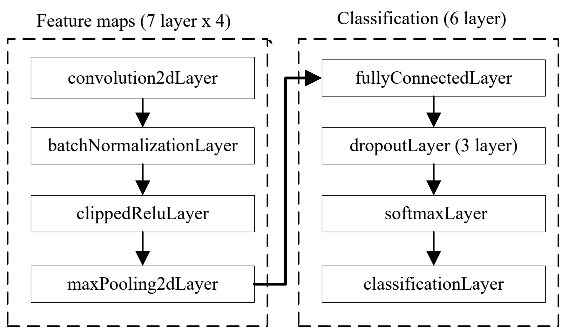


Fig. 3. Architecture 34-Layer

Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.

The dropout layer reduces the number of nodes in the fully connected layer process, making the classification process lighter. At the same time, softmax will perform the probability order of the classification results.

E. *Trial Scenario*

The training also has several CNN architectures to compare the computational load. The architectures used for comparison are alexnet, squeezenet, resnet, and googlenet.

IV. RESULT AND ANALYSIS

At this stage, the analysis is carried out from the initial process, starting with the data acquisition stage

A. *Preprocessing Data*

The secondary image shows that it is small [48 48 3] with 300 dpi, and there is also a colorless image, so it must be sorted and size modified before running CNN.

Fig. 4. Data Acquisition

Preprocessing prepares data before CNN processes it. The first step is to check that all images are the same size, resolution, and color bit depth. The photos are stored in a small size because the repository saves image storage space. The result of the resizing process for the entire image is from [48 48 3] to [224 224 3], shown in Fig. 5.

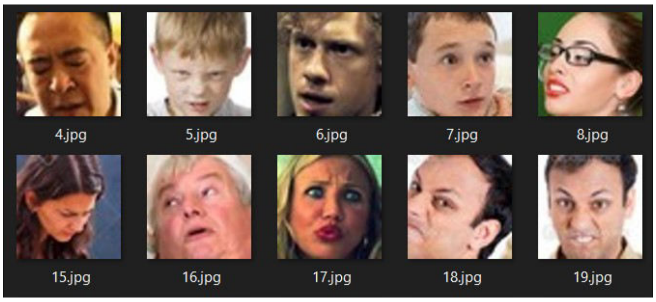


Fig. 5. Preprocessing with image resize

B. *Training Process*

Initial settings for learning parameters including Initial Learn Rate=3e-4, minibatch size = 10, Val Frequency, optimization sgdm (gradient descent) or adam (adaptive moment), Max Epochs = 30, Shuffle = every-epoch, valFrequency = Verbose, false = Plots and the training-progress shows in Fig. 6.

C. *Data Augmentation*

It adds variation by doing rotation, magnification, shears, and reflection, and the results show that changes can be used to overcome unbalanced data. In some uses, the number of results from the augmentation process cannot always be counted, but the fulfillment of class needs can be seen. Rotation is shown in formula 8.

$$A = \begin{bmatrix} \cos(q) & \sin(q) & 0 \\ -\sin(q) & \cos(q) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

With A = kernel rotation, q = degrees of rotation. The augmentation offers in Fig. 7.

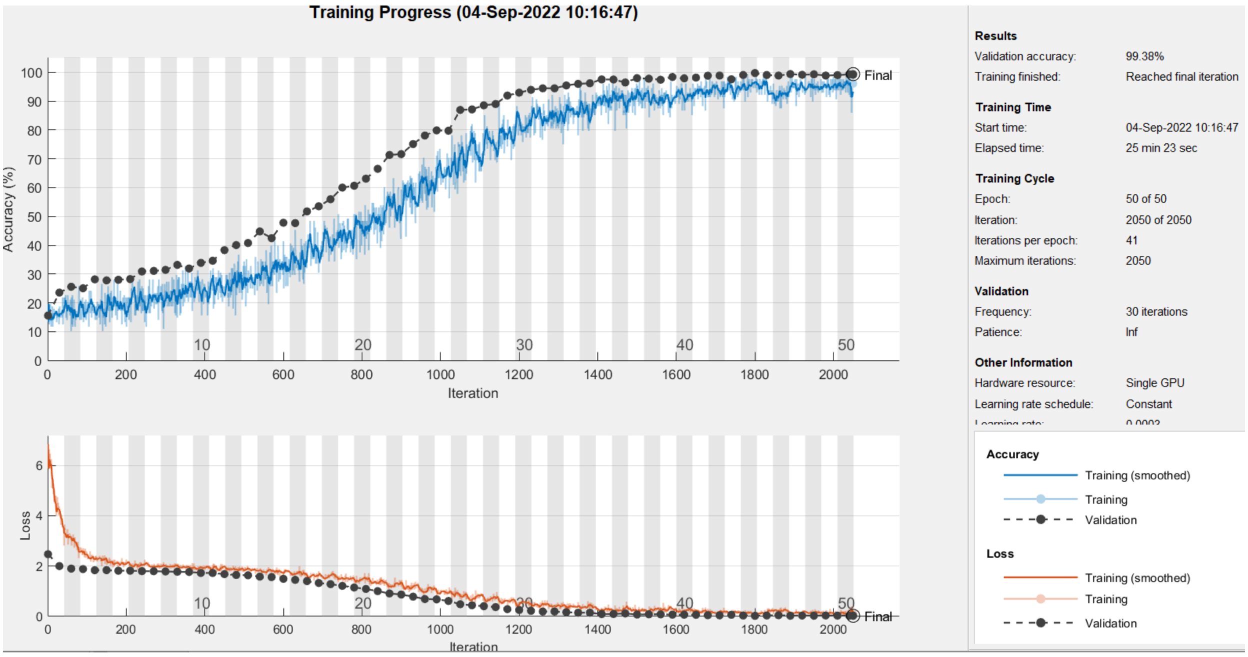


Fig 6. Training Process

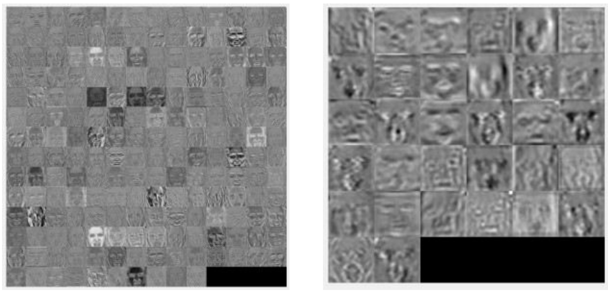
Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.

Fig. 7. Data Augmentation

The advantages of using augmentation techniques are that it does not require additional disk space for the training dataset, fills in the missing data for class unbalance, and adds variation to the data.

D. Visualization layer

The features generated during the convolution process can be visualized by the feature maps, as shown in Fig. 8.



(a) (b)

Fig. 8. layer Visualisation (a) batchnorm_2 (b) batchnorm_4

The picture shows that the farther back, the smaller and smoother the features obtained.

E. Confusion Matrix

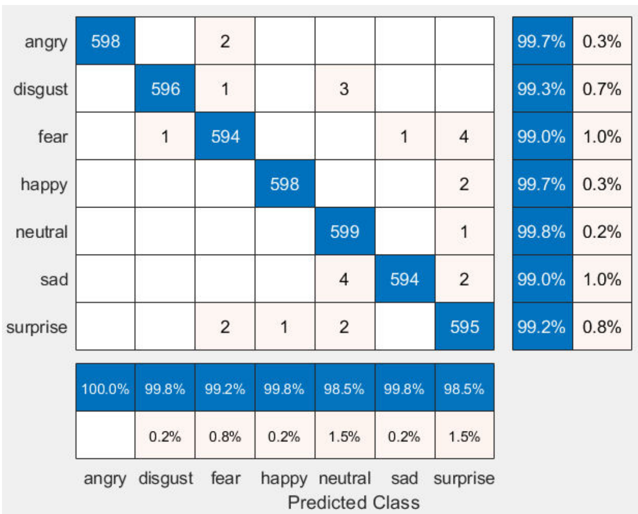


Fig. 9. Confusion Matrix

The stage of mapping image testing data against the model formed during the training process is shown in Fig 9. There are several class errors, so the accuracy obtained is around 99.2%.

F. Prediction

This stage predicts the testing data on the model generated from the training process.

Fig. 10. Prediction 01

The first prediction in Fig. 10 shows that the system can recognize fear reactions with a confidence level of 99.7%, fear at 100%, disgust at 99.6% and 99.8%, surprise at 99.5%

G. Comparison between architecture

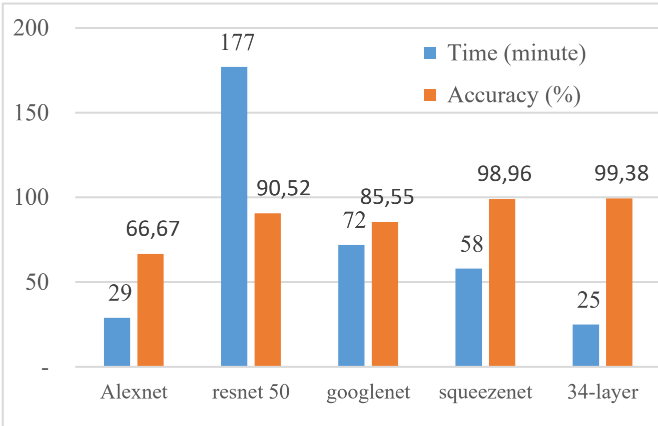


Fig. 11. Accuracy comparison

The graphic data in Fig. 11 shows that the accuracy results are obtained using a custom 34 Layer architecture. The accuracy got 99.38%.

H. Comparison with other research

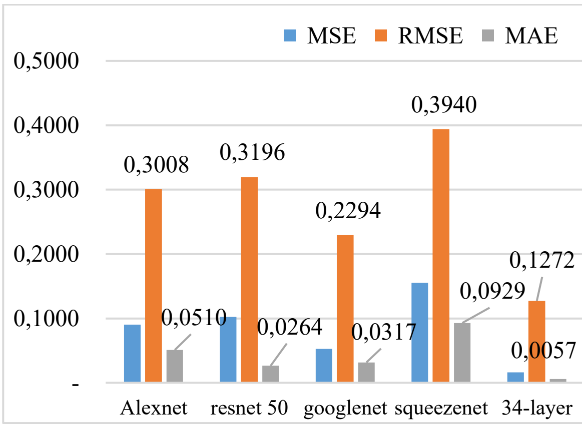


Fig. 12. Comparison of errors classification

Fig. 12 shows that the best error is a minor error obtained using resnet architecture and custom 34 layers. In the training process, the method is said to fail if the desired model is not formed for some reason; for example, the computation is too

Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.

heavy, or the computer turns off during training. In addition, it is stable or overfitting if the training and validation data rate is not in the same line. Failed predictions can also occur when the predicted value has a confidence level of less than 50%, which means that the image enters the wrong class. However, from the scenario results, the confidence level in the prediction results is good because the model produced from the training process has reasonably high accuracy. So that when testing using experimental data, the predicted results are between more than 98%. The future research that can be done is to improve the architecture using a dual graph and skip connection.

V. Conclusion

From the scenarios, the system can adequately recognize the facial expressions of tourist visitors using deep learning. The results can be used to reference visitor satisfaction with the services provided. Factors supporting this system for good results are data distribution and quality, architecture selection, proper training optimization, generated model quality, confusion matrix, minimum errors, and stability. The system can recognize facial expressions with an average accuracy of 99.38%. The average computing time for the training process for constructing a model is 25 minutes 23 seconds with an experiment data testing time of 1-2 seconds: error MSE 0.0445, RMSE 0.2110, and MAE 0.0150. Tests to identify objects have a confidence level of more than 98%.

Acknowledgment

The authors would like to thank the Institute for Research and Community Service, the University of Trunojoyo Madura, which has provided a national collaboration scheme and funded this research.

References

[1] X. Wang, A. M. Ali, and P. Angelov, "Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behaviour," in *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*, 2017, pp. 1–6, doi: 10.1109/CYBConf.2017.7985780.

[2] T.-J. Loiseau, S. Djebali, T. Raimbault, B. Branchet, and G. Chareyron, "Characterization of daily tourism behaviors based on place sequence analysis from photo sharing websites," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2760–2765, doi: 10.1109/BigData.2017.8258241.

[3] A. Essien and G. Chukwukelu, "Deep learning in hospitality and tourism: a research framework agenda for future research," *Int. J. Contemp. Hosp. Manag.*, vol. ahead-of-p, no. Ahead-of-print, Jan. 2022, doi: 10.1108/IJCHM-09-2021-1176.

[4] A. Ahani *et al.*, "Evaluating medical travelers' satisfaction through online review analysis," *J. Hosp. Tour. Manag.*, vol. 48, pp. 519–537, 2021, doi: https://doi.org/10.1016/j.jhtm.2021.08.005.

[5] N. Zhou, R. Liang, and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," *IEEE Access*, vol. 9, pp. 5573–5584, 2021, doi: 10.1109/ACCESS.2020.3046715.

[6] I. P. Adegun and H. B. Vadapalli, "Facial micro-expression recognition: A machine learning approach," *Sci. African*, vol. 8, p. e00465, 2020, doi: https://doi.org/10.1016/j.sciaf.2020.e00465.

[7] M. Li, D. Yin, H. Qiu, and B. Bai, "A systematic review of AI technology-based service encounters: Implications for hospitality and tourism operations," *Int. J. Hosp. Manag.*, vol. 95, p. 102930, 2021, doi: https://doi.org/10.1016/j.ijhm.2021.102930.

[8] F. Balducci, P. Buono, G. Desolda, D. Impedovo, and A. Piccinno, "Improving smart interactive experiences in cultural heritage through pattern recognition techniques," *Pattern Recognit. Lett.*, vol. 131, pp. 142–149, 2020, doi: https://doi.org/10.1016/j.patrec.2019.12.011.

[9] M.-E. R. an emergent approach to the measurement of tourist satisfaction through emotions González-Rodríguez, M. C. Díaz-Fernández, and C. Gómez, "Facial-Expression Recognition: an emergent approach to the measurement of tourist satisfaction through emotions," *Telemat. Informatics*, vol. 51, p. 101404, Mar. 2020, doi: 10.1016/j.tele.2020.101404.

[10] A. Zatori, "Exploring the value co-creation process on guided tours (the 'AIM-model') and the experience-centric management approach," *Int. J. Cult. Tour. Hosp. Res.*, vol. 10, pp. 377–395, Oct. 2016, doi: 10.1108/IJCTHR-09-2015-0098.

[11] J. M. Yachin, "The 'customer journey': Learning from customers in tourism experience encounters," *Tour. Manag. Perspect.*, vol. 28, pp. 201–210, 2018, doi: https://doi.org/10.1016/j.tmp.2018.09.002.

[12] H. Laaroussi, F. Guerouate, and M. sbihi, "Deep Learning Framework for Forecasting Tourism Demand," in *2020 IEEE International Conference on Technology Management, Operations, and Decisions (ICTMOD)*, 2020, pp. 1–4, doi: 10.1109/ICTMOD49425.2020.9380612.

[13] V.-T. Dang, H.-Q. Do, V.-V. Vu, and B. Yoon, "Facial Expression Recognition: A Survey and its Applications," in *2021 23rd International Conference on Advanced Communication Technology (ICACT)*, 2021, pp. 359–367, doi: 10.23919/ICACT51234.2021.9370369.

[14] S. Zhang, N. Chen, and C. H. C. Hsu, "Facial expressions versus words: Unlocking complex emotional responses of residents toward tourists," *Tour. Manag.*, vol. 83, p. 104226, 2021, doi: https://doi.org/10.1016/j.tourman.2020.104226.

[15] K. Yan, S. Huang, Y. Song, W. Liu, and N. Fan, "Face recognition based on convolution neural network," in *2017 36th Chinese Control Conference (CCC)*, 2017, pp. 4077–4081, doi: 10.23919/ChiCC.2017.8027997.

[16] J. Cai, O. Chang, X.-L. Tang, C. Xue, and C. Wei, "Facial Expression Recognition Method Based on Sparse Batch Normalization CNN," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 9608–9613, doi: 10.23919/ChiCC.2018.8483567.

[17] R. Sadiyah, A. Fariza, and E. M. Kusumaningtyas, "Emotion Recognition Based on Facial Expression by Exploring Batch Normalization Convolutional Neural Network," in *2022 International Electronics Symposium (IES)*, 2022, pp. 511–516, doi: 10.1109/IES55876.2022.9888512.

[18] A. M. Javid, S. Das, M. Skoglund, and S. Chatterjee, "A ReLU Dense Layer to Improve the Performance of Neural Networks," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2810–2814, doi: 10.1109/ICASSP39728.2021.9414269.

[19] X. Liu, J. Centeno, J. Alvarado, and L. Tan, "One Dimensional Convolutional Neural Networks Using Sparse Wavelet Decomposition for Bearing Fault Diagnosis," *IEEE Access*, vol. 10, pp. 86998–87007, 2022, doi: 10.1109/ACCESS.2022.3199381.

[20] Y. Liang, F. He, X. Zeng, and B. Yu, "Feature-preserved convolutional neural network for 3D mesh recognition," *Appl. Soft Comput.*, vol. 128, p. 109500, 2022, doi: https://doi.org/10.1016/j.asoc.2022.109500.

[21] A. Soni, R. Koner, and V. G. K. Villuri, "Fusion of Dual-Scale Convolution Neural Network for Urban Building Footprints," *Ain Shams Eng. J.*, vol. 13, no. 3, p. 101622, 2022, doi: https://doi.org/10.1016/j.asej.2021.10.017.

[22] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in an image classification problem," in *2018 International Interdisciplinary Ph.D. Workshop (IIPhDW)*, 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[23] F. Utaminingrum, S. J. A. Sarosa, C. Karim, F. Gapsari, and R. C. Wihandika, "The combination of gray level co-occurrence matrix and back propagation neural network for classifying stairs descent and floor," *ICT Express*, vol. 8, no. 1, pp. 151–160, 2022, doi: https://doi.org/10.1016/j.icte.2021.05.010.

[24] J. Yang, X. Qu, and M. Chang, "An intelligent singular value diagnostic method for concrete dam deformation monitoring," *Water Sci. Eng.*, vol. 12, no. 3, pp. 205–212, 2019, doi: https://doi.org/10.1016/j.wse.2019.09.006.

[25] J. Guo, W. Wang, Y. Tang, Y. Zhang, and H. Zhuge, "A CNN-Bi_LSTM parallel network approach for train travel time prediction," *Knowledge-Based Syst.*, p. 109796, 2022, doi: https://doi.org/10.1016/j.knosys.2022.109796.

[26] F. Bougourzi, F. Dornaika, and A. Taleb-Ahmed, "Deep learning based face beauty prediction via robust dynamic losses and ensemble regression," *Knowledge-Based Syst.*, vol. 242, p. 108246, 2022, doi: https://doi.org/10.1016/j.knosys.2022.108246.

Authorized licensed use limited to: UNIVERSITAS GADJAH MADA. Downloaded on April 08,2023 at 02:50:10 UTC from IEEE Xplore. Restrictions apply.