Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Heavy-loaded Vehicles Detection Model Testing using Synthetic Dataset

Daniel Avian Karjadi[1)*], Bayu Yasa Wedha[2)], Handri Santoso[3)]
[1)2)3)]Universitas Pradita
[1)] daniel.avian@student.pradita.ac.id, [2)] bayu.wedha@student.pradita.ac.id, [3]handri.santoso@pradita.ac.id

Submitted : Apr 10, 2022 | **Accepted** : Apr 12, 2022 | **Published** : Apr 13, 2022

Abstract: Currently, many roads in Indonesia are damaged. This is due to the presence of large vehicles and large loads that often pass. The more omissions are carried out, the more damaged and severe the road is. The central government and local governments often carry out road repairs, but this problem is often a problem. Damaged roads are indeed many factors, one of which is the road load. The road load is caused by the number of vehicles that carry more than the specified capacity. There are many methods used to monitor roads for road damage. The weighing post is a means used by the government in conducting surveillance. This research is not a proposal to monitor the road, but this is only to create a model for the purpose of detecting heavily or lightly loaded vehicles. This research is to classify using Convolutional Neural Network (CNN) with pre-trained Resnet50. The model generated from the Convolutional Neural Network training process reaches above 90%. Generate Image deep learning algorithms such as the Generative Adversarial Network currently generate a lot of synthetic images. The testing dataset that will be used is generated from style transfer. The model is tested using a testing dataset from the generated style transfer. Style transfer is a method of generating images by combining image content with image styles. The model is pretty good at around 92% for training and 88% for testing, can it detect image style transfer? The Convolutional Neural Network model is said to be good if it is able to recognize the image correctly, considering that the accuracy of the model is very good. One of the reasons why the training model is good but still makes errors during testing, then the image dataset is overfitting.

Keywords: Convolutional Neural Network; Resnet50; Pre-Trained; Overfitting; Loaded Vehicle.

INTRODUCTION

One cause of road damage is because the vehicles that cross exceed the load. Actually, vehicles with large loads are not allowed to cross the road. There should be some kind of tool to detect passing vehicles, for example by installing a camera so that there is a kind of warning with the information system. The detection results can be stored as a dataset and become big data to be processed into a road repair decision. For example, how many times has a high-loaded vehicle crossed the highway, and predict the maintenance or repair time of the highway.

Detection of vehicle objects using the classification method. Using Convolutional Neural Network (Kazemzadeh et al., 2022) with Resnet50. Resnet is among the best in the feature extraction process to the classification process. Residual Network (ResNet) is an architecture that consists of several blocks connected by shortcuts connections (Talo, 2019). The residual network has two variants, namely 2 convolution layers measuring 3x3 (18-layer and 34-layer), and 3 convolution layers (Das et al., 2022) measuring 1x1, 3x3, and 1x1 (50-layer to 152-layers). The first 1x1 convolution layer in variant second aims to reduce the dimensions, while the size of the same after 3x3 is used to return output dimensions in the previous layer. This second variant better known as the bottleneck which aims to reduce computational costs due to the large convolution size smaller.

The skip connection introduced by the Resnet architecture is called as a residual connection to avoid information loss during deep network training. Skip connection can be used to train very deep networks and improve network performance. Residual connections in the ResNet architecture, have trained a 1001-layer CNN model. The ResNet architecture consists of residual blocks and connections between residual blocks. The advantages of residual connections in the ResNet architecture are connections between residual blocks, retaining

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the knowledge gained during training and speeding up the model training process by increasing the network capacity.

The research question in this study is whether the testing dataset can be recognized by the model from the Convolutional Neural Network? The hypothesis in this study is that the results of the model can recognize the image synthesis and style transfer. Where the purpose of this research is to test the model from Resnet50 with dataset testing the results of the style transfer. This study will discuss the Literature Review, Methods, Result, Discussion, and Conclusion.

1. LITERATURE REVIEW

Table 1, Previous research that discussed the topic of Convolutional Neural Network with Resnet50

Author	Topic	Advanced	Disadvanced
(Sunario Megawan & Wulan Sri Lestari, 2020)	Face Spoofing Detection Using Faster R-CNN with Resnet50 Architecture on Video	The use of object detection in this case is the face. Using the Fast R Convolutional Neural Network method. The results of the training accuracy are very good and reach 99%.	Testing to detect faces will decrease if you use lighting. This means that the test results are not as accurate as the training results which achieve very good results, but in testing the object images or videos are not as accurate as the 99% training results.
(Kade Bramasta Vikana Putra et al., 2021)	Meat Image Classification Using Deep Learning with Hard Voting Optimization	This research is good by comparing the training models from the VGG16, VGG19, Densetnet-121 and Resnet 50 models. The results are good around 98%.	Testing datasets are not explained more diversely, because testing datasets are very important, considering that testing must be more numerous and datasets must be more diverse. It is not possible to do dataset testing by splitting from the main dataset, for example 80% for training data and 20% for testing datasets.
(et al., 2021)	Classification of food / non-food images using Transfer Learning method with Residual Network model	The discussion about the Resnet18 algorithm is complete with many explanations and the results of the training reached 98% and training with Alexnet reached 97%. The training dataset is 3000 images and the testing dataset is 1000 images.	The testing dataset does not vary, where to test a very high model it is necessary to test with a more varied test dataset. Moreover, if the training model reaches 98%, the testing dataset must be more varied. Testing the model with the testing dataset generated from the Generative Adversarial Network.
(Sandhopi et al., 2020)	Identification of Jepara Motifs on Carvingsby Utilizing Convolutional Neural Network	Discussion of classification on Jepara engraved images using Resnet50, Inception V3, VGG16, Xception and resulted in 95% model accuracy.	Dataset testing does not use a variable testing dataset. Especially for models with an accuracy above 90%. A synthetic dataset, the result of a generative adversarial network, style transfer,

*name of corresponding author

	will provide a better variety of tests.

In previous studies, many have discussed classification using Convolutional Neural Networks using Residual Networks. The results of the training accuracy reached above 90%. This study does not look for weaknesses in previous studies, but this research provides a proposal to complete the test method with a synthetic testing dataset from Generative Neural Network or Style Transfer. **State of the Art** in this research is to test the synthetic image, the result of the style transfer. Because the training model reaches above 90%, testing is carried out with more varied datasets. **The novelty of this research** is to propose a more varied dataset testing for the training model that reaches above 90%, considering that by using a more varied testing dataset, we can determine the model's performance accuracy.

METHOD

This study proposes a detection method using the Convolutional Neural Network (CNN) Resnet50 method. Where is currently using a pre-trained model. The vehicle dataset is used for classification. After completing the training process, to do testing using the results from the Style Transfer image generate which produces a synthetic image. This method is still related to generating a Style Transfer image(Kim et al., 2022), (Lee et al., 2019).

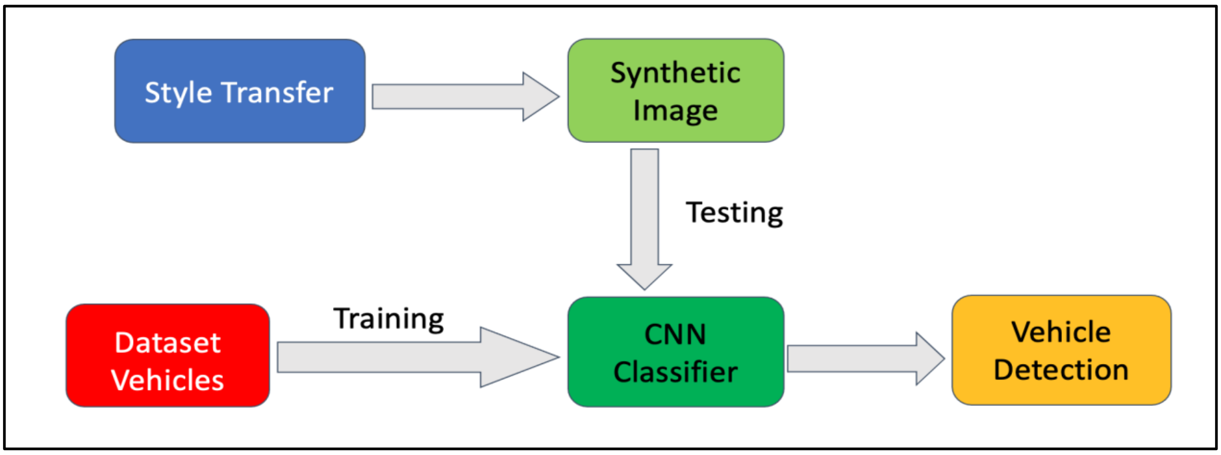

Fig. 1 Research method proposal classification
Source : researcher property

The classification as in Figure 1 shows the model that has been studied using CNN with Resnet50, will be tested with images that have never been tested. The result is whether it gets better or doesn't recognize the type of vehicle that is being classified. This is very important considering that training data and testing data are different from different models.

Style Transfer

This method is a method for combining two images into one. The image consists of two images, namely the content image and the style image. Image content and image style are performed by feature extraction with the encoder method. Then convolutional is done until the image becomes a bottleneck. The image is a feature map of the image content and image style. Adaptive Instance Normalization (Zhang et al., 2021) method , is one method by taking the average and variance of image content and image style. After that, the decoder is carried out to produce an image generator style transfer.

Convolutional Neural Network Classifier

The layers contained in CNN are generally (Arther Sandag et al., 2021): the convolutional layer, pooling layer, and fully connected layer. The layers perform different tasks on the input data. On the convolutional layer, filters are used to extract features. The pooling layer performs a maximization and an average, which extracts the maximum value in the filter region or the average value in the filter region. Meanwhile, the fully connected layer collects information from the map features and produces the final classification. One of the algorithms related to computer vision is Convolutional Neural Network (CNN). Within the CNN scientific branch, there are several

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

algorithms, such as VGG16, VGG19, Inception, Xception, AlexNet, ResNet, and many more. The focus of this research is only on the discussion of Resnet-50. ResNet-50 algorithm with two working methods, feature extraction, and fully connected or classification. Classification is discussed using two classes, namely car, and truck. The Resnet-50 training model will detect an object or image of a car or truck.

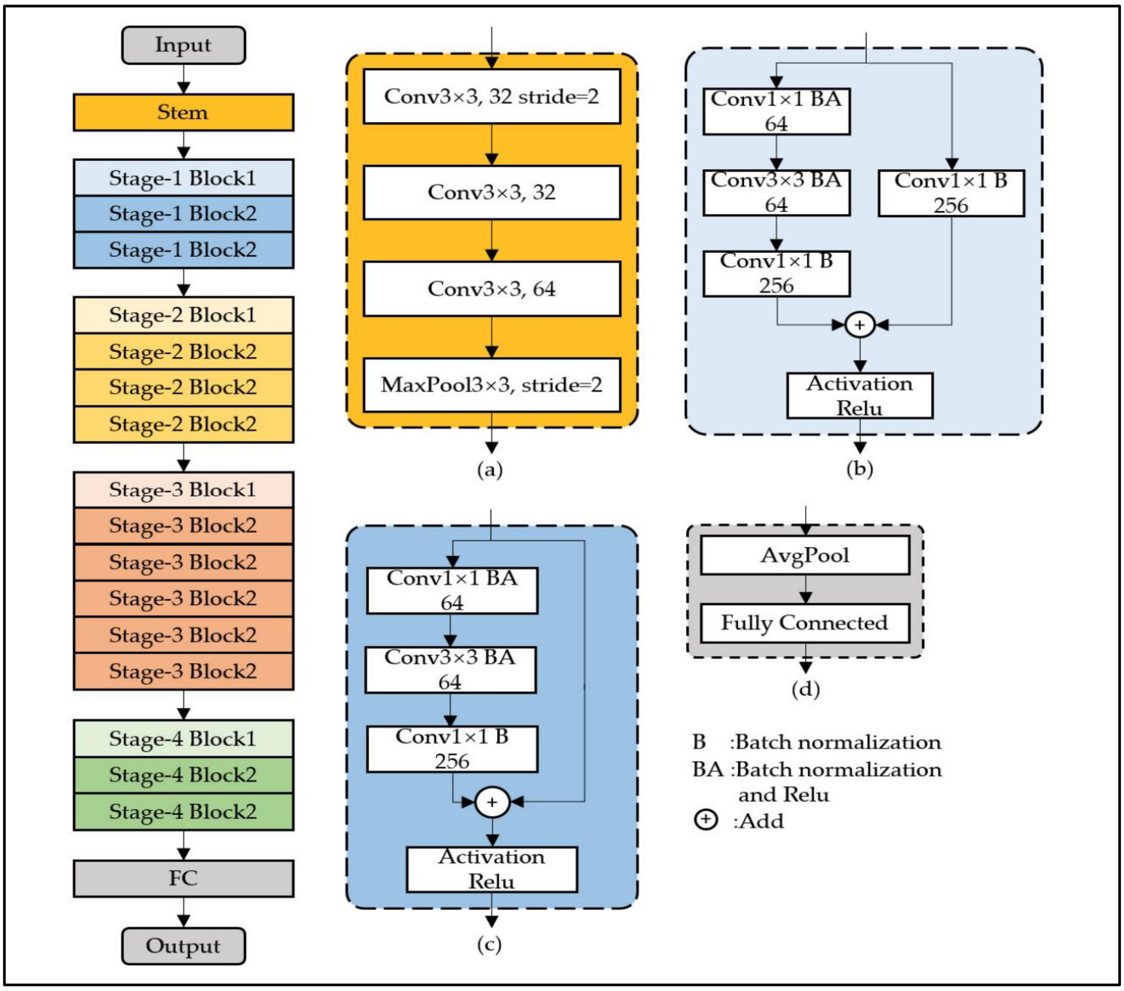

Fig. 2 Resnet50 Architecture
Source : (Wang et al., 2021)

Resnet50 (Residual Network-50 Layer)

ResNet (Residual Network) is a residual network that has a deep network (Zhao et al., 2021). The deepest network of ResNet consists of 152 layers. This network is 8 times deeper than the VGG network but the complexity is still lower than the VGG network. In 2015, the network won first place in the ILSVRC (ImageNet Large Scale Visual Recognition Challenge) and COCO competitions in terms of image classification, detection and segmentation on COCO and ImageNet datasets (Niswati et al., 2021)

Figure 2 shows the detailed architecture of the ResNet50-vd network (ABDULFATTAH et al., 2021), (Qiu et al., 2021). The entire network in figure 2 on the left consists of a parent module, four residual modules, and a fully connected neural network layer. The numbers 32, 64, 256 in figure 2 represent the number of convolution channels. The parent module consists of three 3×3 convolutions and max pooling, and uses stride = 2 in the first convolution and Max pooling to achieve downsampling. Therefore, the output feature map of the master module is half the size of the input, and the number of channels is 64. The modules from stage 1 to stage 4 contain one block1 and several blocks2. It can be seen from Figure 2b that block1 consists of two paths to form a downsampling block. On the left side is a bottleneck structure consisting of three convolutions, which are used to learn new features. On the right side is a structure consisting of convolution and AvgPooling, which is used to process the input into the same size and scale as the output of the bottleneck structure. However for stage2 to Stage4, block1 with parameter stride = 2 to scale the feature map size in half. Block2, like block 1, consists of

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

two paths. The difference is that the structure on the right side of block 2 is a shortcut connection that forms the residual module with the lower structure. The output feature map sizes from trunk to Stage4 are [7, 14, 28, 56, 112]. The number of convolution channels from stem to Stage 4 is [64, 256, 512, 1024, 2048]. Therefore, the final output feature map of Stage4 is 7×7×2048. The last FC layer of the model uses the mean pooling layer and the fully connected layer.

Convolution traditionally cannot make adaptive changes when the image is enlarged or rotated. This is because the calculation rules are predefined, whereas the deformable convolution can make adaptive changes by changing the position and sampling the input. Therefore, dcn-v2 (Zhu et al., 2019) was added to ResNet-vd to improve object deformation adaptability. Deformable convolutions use an additional convolution layer to study offsets. The input and offset feature maps are then used together as input to the deformable convolution layer. The sample points are shifted first by offset, then deflected by the input feature map. In this article, deformable convolution is applied to all 3×3 convolution layers from stage 2 to stage 4 of ResNet50-vd. Therefore, there are 13 deformable convolution layers in the network.

Residual Learning

H(x) is a base mapping that corresponds to multiple stacked layers with x as the input to the first layer. There is a hypothesis that the nonlinear multiple layer asymptotically approaches complex function 2 and is equivalent to the hypothesis that it asymptotically approaches the residual function, that is, $H(x)-x$ (input and output have the same dimensions). The nested layer approximates H(x), explicitly allowing this layer to approximate the residual function $F(x) := H(x) - x$. The original function becomes $F(x) + x$. Although the two forms should be able to asymptotically approach the desired function (as hypothesized), the ease of learning may differ. This reformulation is motivated by the counter-intuitive phenomenon of the degradation problem. If the added layer can be constructed as an identity mapping, the deeper model must have training errors no larger than its shallower counterpart. The degradation problem suggests that the solver may have difficulty approaching the identity mapping by multiple non-linear layers. With the residual learning reformulation, if the identity mapping is optimal, the solver can easily direct the weights of several non-linear layers towards zero to approximate the identity mapping (Kaiming, 2016).

Skip Connection

The skip connection in Figure 2c shows input x to the resulting two-layer output F(x), so that the overall output is $F(x) + x$. The formula gives the intuition that this network has an unbroken gradient flow from the first layer to the last layer and is assigned a Rectified Linear Unit (ReLU) (Sandhopi et al., 2020). The purpose of skip connection is to handle missing gradient information and retain knowledge (Kaiming, 2016), (Talo, 2019). This architecture has several connecting blocks with skip connections. This method is the mainstay of ResNet50 to overcome vanishing or exploding gradients.

Skip Connection (or Shortcut Connection) (Peng et al., 2019), is to skip multiple layers in a neural network and feed the output of one layer as input to the next layer. Skip Connection was introduced to solve different problems in different architectures. In the case of ResNets, the skip connection solved the degradation problem.

Skip connections (Hua et al., 2019) were introduced in the literature before residual networks (ResNet). The highway network is passed by connections with gates that control and flow of information to the deeper layers. This method is similar to LSTM. While ResNets is a special case of the Highway network, its performance does not match that of ResNets. This shows that it is better to keep the gradient highway clean than to go to any gate. Neural networks can study any function of a problem whose complexity can be both high-dimensional and non-convex. Visualization has the potential to answer some of the questions about why neural networks work. There is some good work being done to visualize complex loss surfaces. The result in the figure is that the network with the skip connection shows performance improvements

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

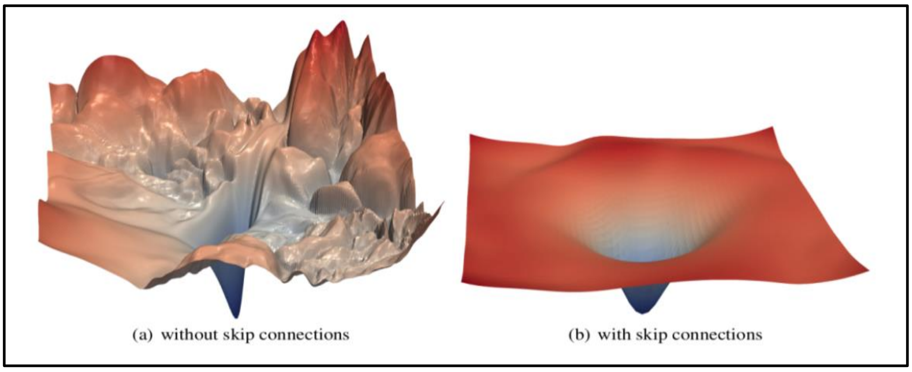

Fig. 3 Gradien with skip connection and without skip connection
Source : https://www.analyticsvidhya.com/blog/2021/08/all-you-need-to-know-about-skip-connections/

RESULT

The results of the detection of light vehicles and heavy vehicles. Experiment of the Restnet50 method using datasets from the car and truck groups. The testing dataset uses the image generator dataset of the image transfer style. This indicates that the testing dataset is not leaking into the training dataset. If the dataset is training and dataset is testing, there will be overfitting.

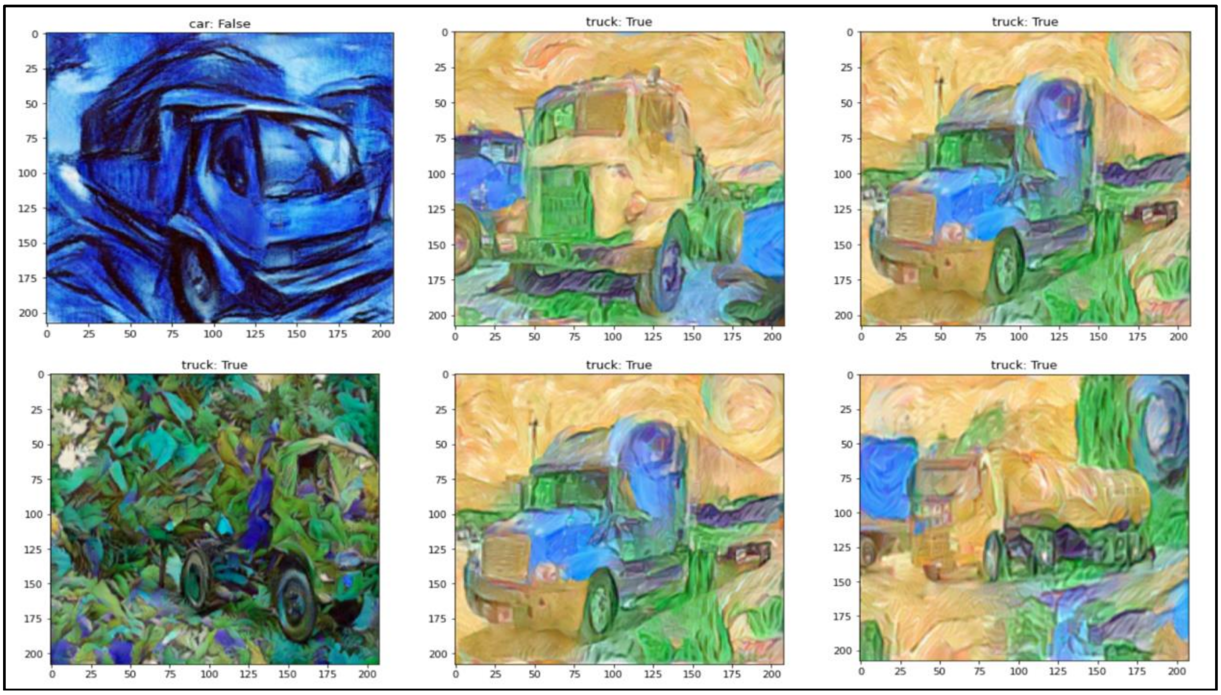

Fig. 3 Hasil deteksi dari dataset testing dengan model Resnet50
Source : researcher property

In Figure 3, there are six images of the detection results from the Resnet50 model. Figure 3a, the detection is still wrong, the detected image should be a truck, but a car is detected, then the result will be False. Figure 3b. and so on is a truck image and the detection result is a truck, then the detection value is True.

Even though the average accuracy is around 93%, there are still errors in detecting the image shown in Figure 3. This result requires a training process with large datasets and changes in the Resnet50 architecture. This results in a model with higher accuracy and more varied testing.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

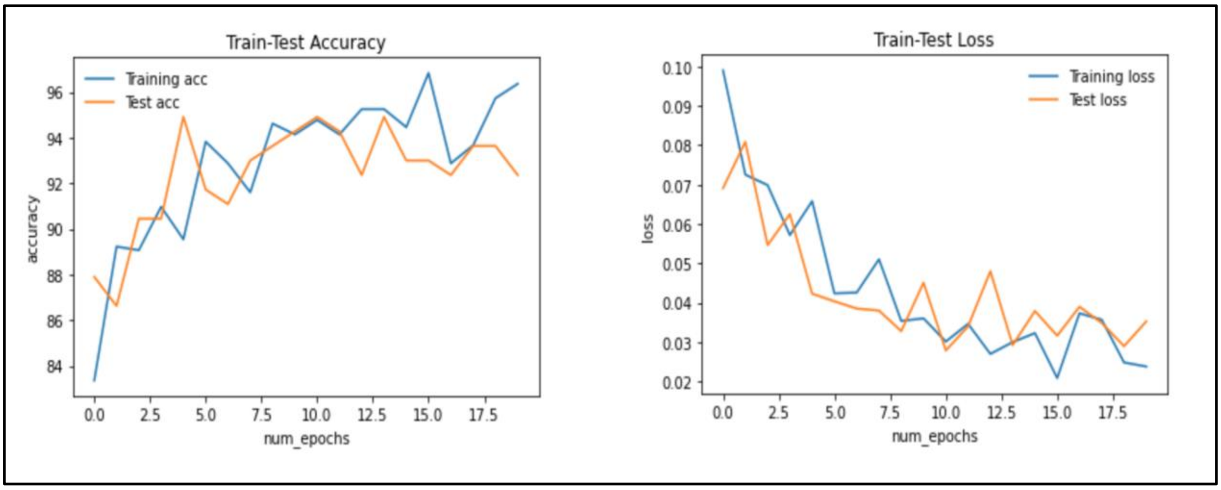

Fig. 4 Accuracy for Train and Testing, Loss for Train and Testing
Source : (Wang et al., 2021)

Performance for accuracy in training and testing is quite high, where the average training accuracy reaches 92% and testing accuracy reaches 88%. These results need to be proven by conducting various testing datasets, to obtain detection results or to recognize images or objects contained in the testing dataset. The results will actually have a good model accuracy. Loss testing and loss training average around 0.2232, which means it's better.

DISCUSSIONS

In Figure 3, from several tests using six testing images, there are five images detected correctly (True) and one image detected incorrectly (False). According to the model that has been trained, the training accuracy has reached 92%, which is a high enough value. But it's still possible there is an error. The results of the model carried out by training means that it has met the good and correct method. In addition to calculating accuracy, it also takes into account loss training and loss testing.

The limitation of this research only uses a classification algorithm using Residual Network 50 layer or better known as ResNet50. The discussion about Resnet50 is not discussed in more detail.

CONCLUSION

The conclusion from the research is that the use of the Residual Network method with 50 layers or ResNet50, is very good in terms of training performance. This is evidenced by testing with a synthetic image dataset from the generated image style transfer. The test results are quite good, but there are still errors in detecting the image or object. The training accuracy is comparable, namely 92%. In testing, errors can still occur.

REFERENCES

ABDULFATTAH, M. E., NOVAMIZANTI, L., & RIZAL, S. (2021). Super Resolution pada Citra Udara menggunakan Convolutional Neural Network. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *9*(1), 71. https://doi.org/10.26760/elkomika.v9i1.71

Arther Sandag, G., Waworundeng Universitas Klabat, J., Arnold Mononutu, J., & -Minahasa Utara, A. (2021). Identifikasi Foto Fashion Dengan Menggunakan Convolutional Neural Network (CNN). *202.62.11.57*, *7*(2), 2021. http://202.62.11.57/index.php/cogito/article/view/340

Das, S. K., Roy, P., & Mishra, A. K. (2022). DFU_SPNet: A stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification. *ICT Express*, *xxxx*, 4–8. https://doi.org/10.1016/j.icte.2021.08.022

Hua, C., Huynh-the, T., Kim, K., Yu, S., & Le-tien, T. (2019). Bimodal learning via trilogy of skip-connection deep networks for diabetic retinopathy risk progression identification. *International Journal of Medical Informatics*, *132*(July), 103926. https://doi.org/10.1016/j.ijmedinf.2019.07.005

Kade Bramasta Vikana Putra, I Putu Agung Bayupati, & Dewa Made Sri Arsa. (2021). Klasifikasi Citra Daging Menggunakan Deep Learning dengan Optimisasi Hard Voting. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(4), 656–662. https://doi.org/10.29207/resti.v5i4.3247

Kaiming, H. Z. X. (2016). Deep Residual Learning for Image Recognition. *EEE Computer Society Conference*

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 2, April 2022
DOI : https://doi.org/10.33395/sinkron.v7i2.11378
e-ISSN : 2541-2019
p-ISSN : 2541-044X

on Computer Vision and Pattern Recognition (2016) 2016-December 770-778. https://doi.org/https://doi.org/10.1109/CVPR.2016.90

Kazemzadeh, M., Hisey, C. L., Zargar-shoshtari, K., & Xu, W. (2022). Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications. *Optics Communications*, *510*(October 2021), 127977. https://doi.org/10.1016/j.optcom.2022.127977

Kim, Y. H., Nam, S. H., Hong, S. B., & Park, K. R. (2022). GRA-GAN: Generative adversarial network for image style transfer of Gender, Race, and age. *Expert Systems with Applications*, *198*(March), 116792. https://doi.org/10.1016/j.eswa.2022.116792

Lee, S., Park, B., & Kim, A. (2019). Deep Learning based Object Detection via Style-transferred Underwater Sonar Images. *IFAC-PapersOnLine*, *52*(21), 152–155. https://doi.org/10.1016/j.ifacol.2019.12.299

Niswati, Z., Hardatin, R., Muslimah, M. N., & Hasanah, S. N. (2021). Perbandingan Arsitektur ResNet50 dan ResNet101 dalam Klasifikasi Kanker Serviks pada Citra Pap Smear. *Faktor Exacta*, *14*(3), 160. https://doi.org/10.30998/faktorexacta.v14i3.10010

Peng, Y., Zhang, L., Liu, S., Wu, X., Zhang, Y., & Wang, X. (2019). Dilated Residual Networks with Symmetric Skip Connection for image denoising. *Neurocomputing*, *345*, 67–76. https://doi.org/10.1016/j.neucom.2018.12.075

Qiu, D., Cheng, Y., & Wang, X. (2021). Progressive U-Net residual network for computed tomography images super-resolution in the screening of COVID-19. *Journal of Radiation Research and Applied Sciences*, *14*(1), 369–379. https://doi.org/10.1080/16878507.2021.1973760

Sandhopi, Lukman Zaman P.C.S.W, & Yosi Kristian. (2020). Identifikasi Motif Jepara pada Ukiran dengan Memanfaatkan Convolutional Neural Network. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(4), 403–413. https://doi.org/10.22146/jnteti.v9i4.541

Sunario Megawan, & Wulan Sri Lestari. (2020). Deteksi Spoofing Wajah Menggunakan Faster R-CNN dengan Arsitektur Resnet50 pada Video. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(3), 261–267. https://doi.org/10.22146/.v9i3.231

Talo, M. (2019). Automated classification of histopathology images using transfer learning. *Artificial Intelligence in Medicine*, *101*, 101743. https://doi.org/10.1016/j.artmed.2019.101743

Thiodorus, G., Prasetia, A., Ardhani, L. A., & Yudistira, N. (2021). Klasifikasi citra makanan/non makanan menggunakan metode Transfer Learning dengan model Residual Network. *Teknologi*, *11*(2), 74–83. https://doi.org/10.26594/teknologi.v11i2.2402

Wang, S., Xia, X., Ye, L., & Yang, B. (2021). Automatic detection and classification of steel surface defect using deep convolutional neural networks. *Metals*, *11*(3), 1–23. https://doi.org/10.3390/met11030388

Zhang, H., Pu, Y., Nie, R., Xu, D., Zhao, Z., & Qian, W. (2021). Multi-modal image synthesis combining content-style adaptive normalization and attentive normalization. *Computers and Graphics (Pergamon)*, *98*, 48–57. https://doi.org/10.1016/j.cag.2021.04.030

Zhao, J., Hu, L., Dong, Y., Huang, L., Weng, S., & Zhang, D. (2021). A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102459. https://doi.org/10.1016/j.jag.2021.102459

Zhu, X., Hu, H., Lin, S., & Dai, J. (2019). Deformable convnets V2: More deformable, better results. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 9300–9308. https://doi.org/10.1109/CVPR.2019.00953

*name of corresponding author