Welding Defect Classification Based on Convolution Neural Network (CNN) and Gaussian Kernel

Agus Khumaidi[1], Eko Mulyanto Yuniarno[1,2] , Mauridhi Hery Purnomo[1,2]
[1] Department of Electrical Engineering, [2] Department of Computer Engineering,
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
agus15@mhs.ee.its.ac.id[1], ekomulyanto@ee.its.ac.id[2], hery@ee.its.ac.id[3]

***Abstract*—Visual inspection process for weld defects still manually operated by human vision, so the result of the test still highly subjective. In this research, the visual inspection process will be done through image processing on the image sequence to make data accuracy more better. CNN as one of the image processing technique can determine the feature automatically which is suitable for this problem in order to classify the variation of each weld defect pattern. Classification using Convolution Neural Network (CNN) consist of two stages: extraction image using image convolution and image classification using neural network. Gaussian kernel used for blurring image, it helps the extraction of images without losing the main information from the original image, this filter also minimize the occurrence of interference or noise. Results of the classification used to get the category of weld defects with high accuracy as a variable of a weld inspection process whether the weld is pass the standard or not. The proposed system has obtained classification with validation accuracy of 95.83% for four different type of welding defect. The data input of this research is the result of images captured by a webcam.**

Keywords—Welding defect, Visual Inspection, Convolution Neural Network.

I. Introduction

Welding inspection method is done through two methods, Non Destructive Testing (NDT) and Destructive Testing (DT). NDT is an inspection of an object to detect defects, cracks, or other discontinuity without damaging the things that we test or inspect. Basically, this test is done to ensure that the materials we use are safe and have not passed the damage tolerance, while for DT is a way of testing welds by disrupting the tested weld. The goal is to determine the strength of the weld by some kind of load testing [1].

Image classification using CNN has been performed in many researches. In 1989 Yann LeCun and his team have managed to perform image classification in the zip code using the Feed forward Neural Network which became known as the Convolution Neural Network (CNN) [2]. The problem of object detection and recognition is one of the oldest problem in the field of computer vision. The object detection deals with finding the objects of interest in a scene where the multiple objects can be present at once, while the object recognition tries to classify either objects extracted by object detection tools or the whole images. The object detection problem is inherently much more difficult than object recognition since objects can be present in an image in vast amount of various locations, scales or different aspect ratios. One of the approaches is commonly used for object recognition is using convolutional neural networks.

Convolutional Neural Networks (CNN) [3] have been applied as an effective structure of models for understanding image content, giving state-of-the-art results on image detection, segmentation and recognition [4] [5]. The key to enable the several factors behind those results were techniques for scaling up the networks to tens of millions of parameters and massive labeled datasets that able to support the learning process. Under these conditions, CNN have been shown to learn interpretable and powerful image features [6] [7].

In this research the input welding image using the original image without going through the process of radiography. The welding image will be captured directly from webcam. Extraction of features using convolution process and gradient descent neural network for image classification, convolution consists of several different filer for each performing its own convolution and those are then stacked on top of each other to create a 3 - dimensional feature map where depth is determined by number of filter. Classification using CNN doesn't need to search features and define welding defect characteristics manually, but the feature will be obtained automatically through convolution process, Therefore the classification by using CNN will increase the value of data accuracy compared with previous research.

This paper structure is as follows: the proposed architecture was discussed in Part II, aspects of the implementation of image classification welding defects present in section III, further discussion of results and experiments are being considered in section IV, and the end result is summarized in section V.

II. CNN Architecture

The illustration of the proposed CNN structure for the classification of welding image sequence of weld defect is shown in Fig. 1.

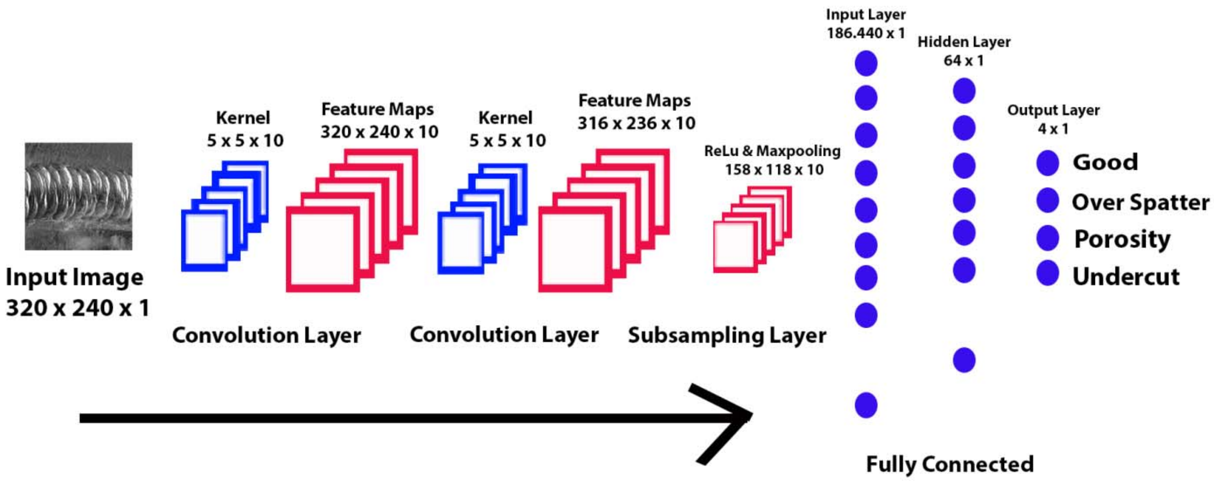


Fig. 1. CNN Architecture.

the structure generally consists of two convolution processes, one subsampling process and the last result of the subsampling process multiplied by the artificial neural network weights through the training process. The image size of 320 x 240 will be multiplied by the value of a 5 x 5 Gaussian filter, the number of filters for each convolution is 10, after the calculation of the input image with the filter value will produce 10 new images after two convolution processes. The next process is the process of subsampling where this process have two functions, first the function of the RELU (rectified linear unit) or a kind of thresholding process that will change the value of negative number to zero and max pooling function that takes the pixel value every 2 x 2 matrix and finds the maximum value of it to produce extraction image with the size of 158 x 118 x 10. Each pixel value will be the input vector of the neural network, the number of input vector is 186440. The number of neurons in the hidden layer is 64 and the output vector is 4, there are good, porosity, undercut or over spatter.

A. Convolution Layer

Feature extraction performed on convolution layer, this layer performs convolution operation on the output of the previous layer. Convolution is a mathematical term, applying a function on the output of other functions repeatedly. In the image processing is applied to the image in all the offset, that allows as shown in Fig. 2. The black box as a whole is an image that will be in the convolution, the kernel moves from the upper left corner to the bottom right. The results of convolution layer can be seen in the convolution features [8].

The purpose of the convolution on image data is to extract features from the input image. Convolution will produce a linear transformation of the input data (I) corresponding to the spatial information data from filter (H). The weights on that layer specifies the convolution kernel is used, so that the convolution kernel can be trained based on input on CNN.

$$I' = H \otimes I$$

$$I'(x,y) = \sum_{i=-n}^{n} \sum_{j=-n}^{m} h(i,j) I(x+i, y+j) \tag{1}$$

m, n is the size of the filter function in the matrix. The functions x and y are the size of the x and y pixel values of an image.

B. Gaussian Kernel

Every convolutional process has a kernel or a feature, in this research Gaussian kernel has been chosen to filter the grayscale image in order to 'blur' images and remove noises. Here is the equation of two-dimensional Gaussian kernel [8]:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

the values of x and y are the size of the Gaussian matrix value with the value x = 0 and y = 0 were located in the diagonal intersection, where σ is the width of Gaussian kernel, in statistic it is usually called standard deviation and the square of it called the variance. In this research σ is used as inner scale or shortly scale.

The factor 2 in the exponent is a kind of convention. The semicolon between spatial and scale parameters is conventionally put to make difference between these parameters explicit. Increasing σ will reduces the amplitude substantially.

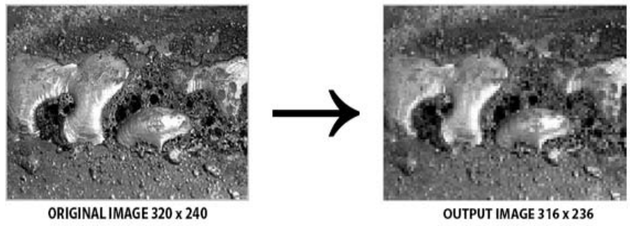


Fig. 2. Image result of convolution process calculation using 5 x 5 Gaussian kernel with σ = 30.

C. Max Pooling

After passing through the convolution process, the next stage is RELU layer will implement the activation function elements, such as max (0, x) thresholding zero [9]. So, if the pixel value is less than zero then it will be switched equal to zero. Subsampling laye is the process of reducing the size of an image data. In image processing, subsampling also aims to improve the position invariance of features. In most of CNN, subsampling method used is the max pooling. Max pooling output of the convolution layer divides into several smaller grids and taking the maximum value of each grid to compose the image matrix that has been reduced as shown in Fig. 3. The results of this process can be seen on the set of the grid on the right. The process will ensure the same features found even though the object image translated (shift).

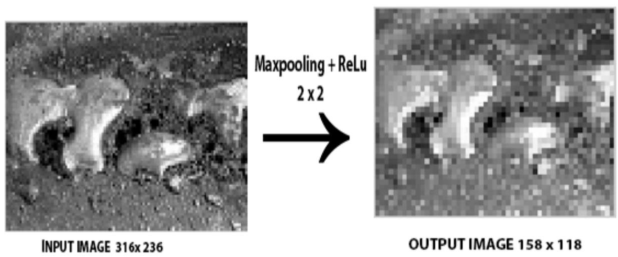


Fig. 3. Image result of maxpooling dan ReLu.

D. Classification using NN

A neural network is made of simple processing units called neuron. In the proposed scheme, the classification method is a feature matching method for which NN are typically used. One of the most universal and frequently used NN learning algorithms is the back-propagation algorithm which is based on the error energy minimalization. Each data output neuron will be activated, activation function is non-linear function that allows a neural network to be able to transform input data into a higher dimension. In the CNN are used activation function is sigmoid function. Sigmoid function to transform the range of values of the input x being between 0 and 1 in the form of the distribution function. So that the sigmoid function has the following form [10]:

$$f(x) = \frac{1}{(1 + e^{-f(x)})} \quad (3)$$

x in equation 3 or the equation is the output value of the calculated neuron by weight. A general training algorithm for any neural network is usually done by evaluating the network on a training example and comparing the result of network with expected result by called loss function. Loss function or error function is usually designed in a way and measured by how much the result differs from the expected output and generally to minimize the output of such loss function over all training examples in a training set. This minimization can be seen as optimization problem and to solve this problem weights are adjusted in the direction of maximal gradient. This method is called as a gradient descent algorithm.

III. Implementation Aspects

A. Welding Image Dataset and Data Divisions

Data set in this study have the same proportions for each category show in Fig. 4: 30 for good (GO) image, 30 for overs patter (OS), 30 for porosity (PO) and 30 for undercut (UC). The total of contents in this dataset are 120 data, then those are divided into two separated data. First is training data and second is testing data. The proportion of each data are 80 % (96 images) for training data and 20 % (24 images) for testing (validation) data.

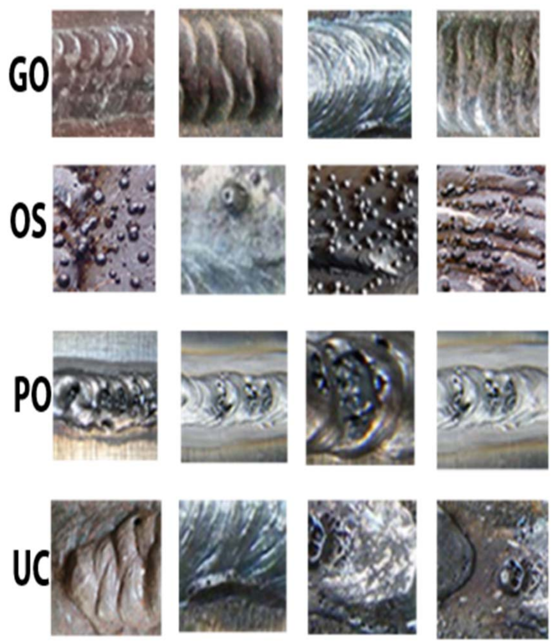


Fig. 4. Dataset of welding defects

B. Image Processing and Convolutional Parameters

This paper using 320x240x3 RGB images, which are converted into 320x240x1 Grayscale images. After all of those images has been converted into grayscale, then convolution process is started. Using 10 numbers of convolutional filters, size of pooling area is 2 x 2 and the kernel size is 5 x 5.

C. Neural Network Implementation

On the artificial neural network implementation of the proportion of data for training and testing purpose is between 70:30 or 80:20, in this study using the ratio of 80:20 because statically says the more number of data learning then the system is closer to the goal of the problem if the selected method is appropriate.

The input of NN architecture is 320x240x1 grayscale image, which is already passed the convolutional process until the pixel size is changed into 158x118x10, each pixel value on each frame is directly connected with hidden layer after passing the flattening process. In this research, the proposed NN structure has single hidden layer with total number of neurons of 64. The output is divided into four classes (binary classification), there are:

[1 0 0 0] = good
[0 1 0 0] = over spatter
[0 0 0 1] = undercut
[0 0 1 0] = porosity

Training method for fully connected layer is using gradient descent algorithm. Mean squared error (MSE) is used to evaluate the training result. It was trained for 60 times (epoch). At the end of the section on this paper, the accuracy of the testing data result will be presented.

IV. Result and Discussion

In this section, there are two parameters to be displayed. First parameter is accuracy response versus number of iteration and second parameter is MSE response versus number of epoch. Both of them are performed in two-dimension (2D) graph which consist of the training data and testing data result.

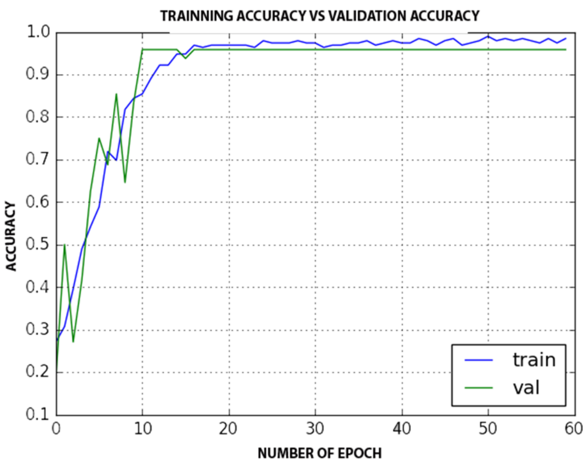
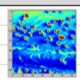
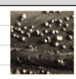
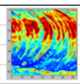
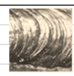
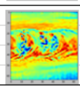
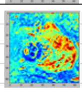
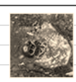
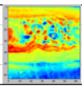
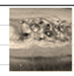


Fig. 5 Accuracy performance

Fig. 5 is graphical result between number of iteration versus accuracy. It shows blue line graphic (training response) and green line graphic (testing or validation response). The training performance is stable and accurate when the iterations are over 10. Gradient descent is able to train the data with the accuracy value of almost 100 percent for training data and testing (validation) accuracy is between 95 till 100 percent. The average of testing accuracy is 0.9583 or equal to 95.83 %. It means the gradient descent can classify the unknown welding image data until 95.83 percent respectively. CNN structure able to learn data pattern with good performance, from 24 data testing images or unknown welding images on this research can classify more than a half of them correctly. It means 23 unknown welding images has successfully classified. Table I is representing the prediction of several testing data.

TABLE I. Sample of Prediction Result

number	input image	greyscale image	predicted value (binary output)	real value	real condition	result correct / incorrect
1			[0 0 0 1]	[0 1 0 0]	*over spatter*	incorrect
2			[1 0 0 0]	[1 0 0 0]	*good*	correct
3			[0 0 1 0]	[0 0 1 0]	*porosity*	correct
4			[0 0 0 1]	[0 0 0 1]	*under cutting*	correct
5			[0 0 1 0]	[0 0 1 0]	*porosity*	correct

Mean squared error or the loss value in Fig. 6 also have the same result with the accuracy performance, Validation result is stable enough with the average MSE value of 0,02. Gradient descent algorithm has successfully classified the welding defect with average MSE training of 0,01.

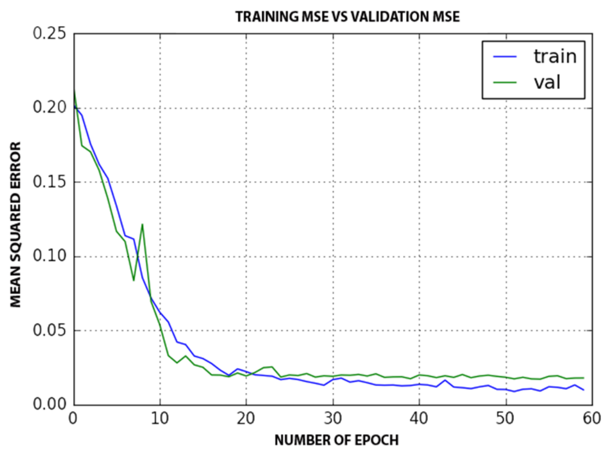


Fig. 6. MSE performance

V. CONCLUSION

Convolution Neural Network has successfully predicted welding defect with validation accuracy until 95.83 %. Using 10 layers of Gaussian kernel, 60 times iteration and gradient descent algorithm this research can classify 23 unknown welding images from 24 of total data testing correctly.

REFERENCES

[1] J. Primo, Welding Inspection Qualifications and Testing Procedures, PDHonline Course M415 (8PDH), 2012.

[2] Y. B. Y LeCun, Convolutional Networks For Images, Speech and Time Series, MIT Press, 1995.

[3] L. B. Y. B. a. P. H. Y. LeCun, "Gradientbased learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278 - 2324, 1998.

[4] I. a. H. G. E. Alex Krizhevsky and Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks," in *Curran Associates, Inc.*, 2012.

[5] H. A. S. Ali Sharif Razavian, "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, Columbus, OH, USA, 2014.

[6] M. D. Z. Fergus, "Visualizing and Understanding Convolutional Networks," in *European Conference on Computer Vision*, 2014.

[7] Z. S. X. J. J. P. Manli Sun, "Learning Pooling for Convolutional Neural Network," *Neurocomputing,* vol. 224, pp. 96-104, 2017.

[8] O. Jasek, Detecting Objects for Autonomous System Verification, Prague: Czech Technical University, 2016.

[9] R. L. Yanyi Zhang, "Selection of the parameter in Gaussian kernels in support vector machine," in *Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, China, China, 2017.

[10] F. D. G. A. D. C. Jawad Nagi, "Max-Pooling Convolutional Neural Networks for Vision-based Hand Gesture Recognition," in *International Conference on Signal and Image Processing Applications (ICSIPA2011)*, Kuala Lumpur, Malaysia, 2011.

[11] B. M. K. Thamer M.Jamel, "Implementation of a Sigmoid Activation Function for Neural Network," in *The 13th Scientific Conference of Al-Ma'moon University College*, Baghdad , Iraq, 2012.