

Review on the Recent Welding Research with Application of CNN-Based Deep Learning Part II: Model Evaluation and Visualizations

합성곱 신경망기반 딥러닝의 용접연구 적용 Part II: 모델의 평가와 시각화

Kidong Lee^{*,**}, Sung Yi^{*}, Soongkeun Hyun^{**} and Cheolhee Kim^{*,***,†}

^{*}Department of Mechanical and Materials Engineering, Portland State University, OR 97229, USA

^{**}Department of Materials Science and Engineering, Inha University, Incheon, 22212, Korea

^{***}Joining R&D Group, KITECH, Incheon, 21999, Korea

[†]Corresponding author: chkim@kitech.re.kr, cheol@pdx.edu

(Received November 10, 2020; Accepted January 18, 2021)

Abstract

With the development of deep learning technology, research on classification and regression models on welding phenomena using convolution neural networks (CNNs) are gradually increasing. Part 1 of this study introduced the characteristics of deep learning models using CNNs and their application to welding studies. In this paper, we reviewed recent welding research papers to analyze how to evaluate CNN models and visualize the modeling output, and details of evaluation index, comparison models, and visualization methods were explained.

Key Words: Deep learning, Convolution Neural Network, Image, Welding, Evaluation, Visualization

1. Introduction

There has been recent increase in the application of machine learning to a wide range of industrial sectors, and artificial neural networks, one of the machine learning techniques, have drawn much interest. Research with adoption of neural networks in welding was first introduced in the 1990s¹⁻⁴⁾, and a number of studies have recently been published applying deep learning and convolutional neural networks (CNN) to welding. CNNs triggered a breakthrough in image recognition, and compared to the previous approaches of extracting features and learning the relationship between the extracted features and results, CNNs facilitate more generalized learning, and its application to welding research has drastically increased.

Part I⁵⁾ of this review introduces the basic structure and learning methods of CNN, and the previous studies⁶⁻²²⁾ that applied CNNs in welding research have

been categorized into 5 groups according to the criteria of supervised/unsupervised learning, application of transfer learning, adoption of feature extraction or data augmentation in data preprocessing, and the characteristics and applications of the models used were introduced.

In the case of CNN, due to its complexity in dimension, it is not easy to perform evaluation and visualization of the results compared to more traditional and intuitive models such as linear regression. In this review, we classify the applicable publications according to evaluation indices, evaluation method and visualization method used in the research as can be seen in Table 1, and the cases of applications are introduced

2. Evaluation Metric

For evaluation of the performance of classification and regression models developed through neural network modeling, evaluation indices such as accuracy

Table 1 Summary of research papers reviewed in this study

Classification / Regression	Evaluation index	Comparison with		Visualization	1stAuthor (Publication Year)	Ref. No.
		Conventional method	Neural Networks			
Classification	Accuracy				A. Khumaidi (2017)	6)
Regression/ Classification	Accuracy				S. Choi (2019)	7)
Classification	Accuracy, Confusion matrix				B. Zhang (2020)	8)
Regression	MAPE*	Edge detection			T. Ashida (2019)	9)
Classification	Accuracy, Recall, F1 score		SCNN* ⁸ /CNN (pretrained and tuned models)		C. V. Dung (2019)	10)
Classification	Mean error rate		FCN* ⁹		Y. Zhang (2019)	11)
Classification	Accuracy, Training time		FCN, CNN(No. of Class)		D. Bacioiu (2019)	12)
Classification	Accuracy		CNN (Classification module)		H. Zhu (2019)	13)
Classification	Accuracy		CNN (with and without transfer learning)		W. Jiao (2020)	14)
Classification	Accuracy, Training Time		CNN (reference)	Intermediate activations	Z. Zhang (2019)	15)
Classification	Accuracy, Recall, Training time		AlexNet, VGG-16, Resnet-50, Densenet-121, Mobile NetV3-Large (pretrained and tuned models)	Intermediate activations	Y. Yang (2020)	16)
Classification	Training time, Mean error rate	Visual inspection of expert	CNN (No. of layers)		Z. Guo (2017)	17)
Classification	Accuracy	SVM* ⁴	CNN (Proposed model, Transfer learning), FCN		N. Yang (2018)	18)
Regression/ Classification	TPR** ₂ , TNR* ³	Hough Circle/ HOG* ⁵ +SVM, LBP* ⁶ +SVM	CNN		J.-K. Park (2019)	19)
Classification	Accuracy	Haralick feature, HOG feature	SSAE* ¹⁰ , CNN	t-SNE* ¹²	W. Hou (2019)	20)
Classification	Accuracy, Recall	HOG, LBP, HOG+LBP, BOF* ⁷	SAE* ¹¹ , CNN(No. of layers)	t-SNE	Z. Zhang (2020)	21)
Classification (Auto-encoder)	Accuracy				A. Muniategui (2017)	22)

*Mean Absolute Percentage Error; **True Positive Rate; *³True Negative Rate; *⁴Support Vector Machine;

*⁵Histogram of Oriented Gradient; *⁶Local Binary Pattern; *⁷Back Of Features; *⁸Shallow ConvNet;

*⁹Fully Connected Net; *¹⁰Stacked Sparse Autoencoder; *¹¹Sparse autoencoder;

*¹²t-Distributed Stochastic Neighbor Embedding

and learning rate are used. In most cases, in addition to the use of evaluation indices, comparative analysis is performed through comparisons with other models.

2.1 Classification model

The field with most active utilization of image-based CNN is classification by supervised learning. In a classification model, each result can be represented using a confusion matrix according to the prediction result²³⁾, and the confusion matrix of binary classification, which has the simplest form, is shown in Fig. 1.

The prediction results of the binary classification model are classified into True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Of these, true positives and true negatives represent the cases of good prediction performance, and false positives and false negatives represent errors. A false positive is an error that incorrectly classifies a negative as positive, and a false negative is an error that classifies a positive as negative.

In some studies, a true positive rate (TPR), a false positive rate (FPR), a true negative rate (TNR), and a false negative rate, which represent the percentage of the each classification result, have been adopted.

When the classification datasets are balanced datasets, accuracy, which is the ratio of correct classification out of the total datasets, is sometimes used as an evaluation index. Conversely, in some cases, the mean error rate, which is the case of using incorrect classification ratio out of the total datasets, is employed.

$$(accuracy) = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$(Mean\ Error\ Rate) = \frac{FN + FP}{TP + FN + FP + TN} \quad (2)$$

		Prediction	
		Positive	Negative
Actual value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Fig. 1 Confusion matrix for binary classification

However, if the datasets are imbalanced, the indices may be biased, so complementary indices such as precision and recall are needed. Among the data classified as true by the model, the ratio of real true cases is called precision, and the ratio of classification as true when the real value is true is called recall.

$$(precision) = \frac{TP}{TP + FP} \quad (3)$$

$$(recall) = \frac{TP}{TP + FN} \quad (4)$$

The precision and recall are complementary indices, and the model shows good classification performance when the values of both indices are high. In addition, the F1 index (F1 score), which is the harmonic mean of precision and recall, can be used.

$$(F1\ score) = 2 \times \frac{1}{\frac{1}{(precision)} + \frac{1}{(recall)}} \quad (5)$$

$$= 2 \times \frac{(precision) \times (recall)}{(precision) + (recall)}$$

When CNN was applied to the classification model, mean accuracy was used as an evaluation index in most cases^{6-8,10,13-16,18,20,21)}. Y. Yang et al.¹⁶⁾ and Z. Zhang et al.²¹⁾ used recall as well as mean accuracy as an evaluation index. In particular, C. V. Dung et al.¹⁰⁾ used all of the mean accuracy, recall, and F1 score for evaluation. Z. Zhang et al.²¹⁾ presented a confusion matrix with mean accuracy for evaluation. Y. Zhang et al.¹¹⁾ and Z. Guo et al.¹⁷⁾ used mean error instead of mean accuracy. J.-K. Park et al.¹⁹⁾ emphasized the importance of identifying defective products in the manufacturing process of parts for which safety is a primary concern, such as engine transmissions, and used the TP ratio as well as the TN ratio, which is the accuracy of defect detection, as evaluation indices.

2.2 Evaluation indices of the regression model

In the regression model, the mean square error (MSE) based on the difference between the actual and predicted values is used as a loss function, and the mean absolute error (MAE) is most frequently used as an evaluation index.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y'(i) - y(i))^2 \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'(i) - y(i)| \quad (7)$$

Here, n represents the total number of data, i is the index number of data, $y(i)$ is the measured value of the i -th data, and $y'(i)$ is the predicted value of the i -th data by the regression model. The closer the mean absolute error is to 0, the higher the fit of the regression model. In addition, the mean absolute percentage error (MAPE) which represents the percentage of the mean absolute error is defined as in Eq. 8, and mean accuracy, the opposite concept, is defined as in Eq. 9.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y'(i) - y(i)|}{y'(i)} \quad (8)$$

$$(\text{mean accuracy}) = 1 - MAPE \quad (9)$$

The coefficient of determination (R^2) is also used to indicate how well the predicted value obtained through the regression model explains the correct value.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y'(i) - y(i))^2}{\sum_{i=1}^n (y'(i) - \bar{y}'(i))^2} \quad (10)$$

$\bar{y}'(i)$ represents the mean of the predicted values. The coefficient of determination is in the range of $0 \leq R^2 \leq 1$, and the closer the value is to 1, the higher the accuracy of the regression model. S. Choi et al.⁷⁾ and T. Ashida et al.⁹⁾, who applied the CNN model to the regression model, used mean accuracy and mean error as evaluation indices, respectively.

3. Comparative evaluation with other models

Various regression and classification models have been developed for image-based prediction of welding phenomena. CNN models proposed in a number of papers are compared with the traditional classifier or other neural network models and the performance of the model was verified.

3.1 Comparison with traditional feature extraction method and classifiers

In addition to neural networks, there are a variety of methods used for classification and regression. The conventional models presented in this section describe methods that have been employed before the introduction of neural networks, and include image processing algorithms such as Hough transform and contour extraction algorithm, traditional classifiers, and visual inspection of skilled personnel that has been traditionally used in manufacturing process.

T. Ashida et al. compared the proposed CNN re-

gression model with the traditional contour extraction image processing technique and confirmed that errors in the prediction of molten pool width and leading end width can be reduced⁹⁾. Z. Guo et al. evaluated the performance of a CNN model with comparison against the visual inspection of the skilled worker¹⁷⁾.

Traditional image classification models other than neural networks can be described by categorizing into feature extraction and classifier. Feature extraction methods include HOG (Histogram of Oriented Gradient), LBP (Local Binary Pattern), Haralick feature²⁴⁾, and for classifiers, SVM (Support Vector Machine) is often used. In order to improve the performance of traditional feature extraction methods, image processing such as binarization or contour measurement to highlight features is conducted first. HOG is an algorithm that segments an image into a certain size and normalizes local cells through a gradient direction distribution to extract features, and LBP is an algorithm that binarizes 3x3 areas around all pixels in an image to 0 or 1 with relative brightness. LBP was originally developed to classify texture, but it is mainly used for image analysis such as face recognition. Also, Haralick feature is a feature extraction algorithm based on the gray level gradient of adjacent pixels. SVM is a classification algorithm using machine learning and consists of a set of hyper-planes. It is a classifier that performs classification using a maximum-margin hyper-plane that maximizes the distance between data classes in high dimensions.

W. Hou et al.²⁰⁾, N. Yang et al.¹⁸⁾, Z. Zhang et al.²¹⁾, and J.-K. Park et al.¹⁹⁾ compared traditional classifiers with the proposed CNN model. In the comparison for performance evaluation, W. Hou et al. used HOG feature and Haralick feature extraction method²⁰⁾, and N. Yang et al. used SVM as a classifier¹⁸⁾. Z. Zhang et al. compared various traditional classifiers such as HOG, LBP, HOG+LBP, and BOF with the proposed CNN model²¹⁾. J.-K. Park et al. compared the Hough circle contour extraction algorithm with the CNN regression model to evaluate the performance of tracking the center point of the engine transmission weld, and also compared the CNN classification model with HOG+SVM and LBP+SVM to evaluate performance¹⁹⁾.

3.2 Comparison with other neural network models

For image recognition, SAE(Sparse Auto-Encoder) image feature extraction method, one of unsupervised learning neural networks, and CNNs are mainly used. SAE is a type of auto-encoder, which is unsupervised learning frequently used for image feature extraction. Through data flattening, a fully connected neural network (FCN) can also be used for image feature extraction.

N. Yang et al.¹⁸⁾, Y. Zhang et al.¹¹⁾, and D. Bacioiu et al.¹²⁾ compared the proposed CNN classification model with the FCN model for performance evaluation. In addition, W. Hou et al.²⁰⁾ and Z. Zhang et al.²¹⁾ conducted performance evaluation using SAE and stacked SAEs, respectively.

CNN models vary widely depending on their structure and the adoption of transfer learning, so the performance of the CNN model can be improved through structure optimization, and modeling techniques. Thus, modeling performance can be evaluated using various CNN model structures and modeling techniques and in more than half of the papers reviewed, performance was evaluated through comparison between CNN models^{10,12,13,15-21)}.

D. Bacioiu et al. compared the difference in the accuracy of the proposed CNN model as the classification classes became more complex from 2, 4, to 6¹²⁾. C. V. Dung et al. used mean accuracy, recall, and F1 score for as evaluation indices comparison of three CNN models, which are SCNN (Shallow CNN), BN (Bottle Neck)-CNN, and FT (Fine Tuning) CNN. The comparison between SCNN and CNN model with application of VGG-16-based transfer learning showed that superior performance was obtained in the model with application of transfer learning. In addition, by comparing the BN-CNN model and the FT-CNN model, it was confirmed that through fine-tuning, the accuracy can be improved through optimization even in the same transfer learning model¹⁰⁾.

W. Jiao et al. compared the accuracy of the proposed 9 layer CNN model and the transfer learning CNN model. Through the comparison, it was confirmed that the application of transfer learning is effective for learning with a small amount of training data¹³⁾.

H. Zhu et al. compared the accuracy of each classifier in the CNN model. The feature map was extracted through the CNN model, and when compared with softmax and SVM techniques, which are commonly used classification methods, the proposed random forest method showed superior performance¹³⁾.

Y. Zhang et al. conducted performance evaluation by comparing the accuracy of the FCN and the proposed CNN model. In addition, the cause of the error was investigated through analysis of misclassified samples¹¹⁾.

4. Visualization

In the reviewed papers, evaluation of the model through visualization methods is largely divided into two types: a method of feature extraction of the classification model through the intermediate activations and a method of visualizing the classification through a 2D

graph by using t-SNE.

4.1 Intermediate activations

A feature map that has been filtered by a convolutional filter in the CNN learning process can be represented as an image, and the highlighted features can be identified to some extent in lower layers. Intermediate activations are used as an evaluation method by 1) extracting feature map of lower layers and determining if the highlighted features are suitable for the purpose of the model, or 2) identifying the number of images with significant features at higher layers.

Y. Yang et al. showed that by visualizing the image of the first layer of the optimized CNN model through intermediate activations, the features of the images suitable for the purpose of the model were properly highlighted¹⁶⁾.

Z. Zhang et al. visualized and output a feature map at each CNN layer. Using the number of images with significant pixels (images in which values of all pixels are non-zero) in the feature map of each layer, it was confirmed that the proposed model shows effective performance in feature extraction¹⁵⁾.

4.2 t-SNE

t-SNE (t-Distributed Stochastic Neighbor Embedding) is a nonlinear dimensionality reduction technique suitable for visualization of high dimensional datasets. It represents the similarity of data as a distance in two or three dimensions. This method calculates neighboring probability in high-dimensional space and low-dimensional space assuming normal distribution by Kullback-Leibler divergence method and minimizes the probabilities and this process of calculation and minimization is learned and visualized.

W. Hou et al. extracted features using the Haralick feature, HOG and SSAE, and two CNN methods with different depths, and visualized these extracted features with a 2D map through t-SNE. In addition, the performance of the proposed model was evaluated by comparing the mean accuracy of classification results obtained from each method. It was confirmed that the proposed CNN model showed clear classification in the t-SNE 2D map compared to other methods, and the clearer the classification in the visualized 2D map, the higher the mean accuracy²⁰⁾.

Z. Zhang et al. extracted features from the optimized CNN model and visualized the features with a 2D map through t-SNE. The classification between each group was visually confirmed on the map, and the cause of the error was presented by analyzing the section where

the classification was not clear²¹⁾.

5. Visualization

There are various types of auto encoders, which are unsupervised neural networks described above, and among them, a convolutional auto-encoder is advantageous for image compression. The autoencoder compares the original images and the restored images which encoded (compressed) and decoded (decompressed) through the autoencoder, and performs unsupervised learning. Whether the classification model retains the features required for classification after image compression can be evaluated through the performance of the final classification model.

Muniategui et al. used a deep convolutional auto-encoder for image compression and used the compressed images as inputs to the fuzzy model²²⁾. 225 vectors were extracted from the compressed 15x15-pixel image and grouped by prediction class of Very Good, Good, Regular, Bad, and Very Bad. For evaluation of the compressed image, the mean of the grouped vectors was compared with each pattern. The patterns of Very Good, Good, and Regular groups of the proposed four fuzzy classification models were similar, and in the case of Bad and Very Bad groups, the patterns showed specificity depending on the classification model. In this case, in the compressed images, the similarity of the compressed image vectors from the similarity of the production parts was confirmed, and it was also identified that there were specific patterns related to the difference between each classification model. Finally, the performance of the fuzzy classification model when images were compressed using a deep convolutional auto-encoder was evaluated using accuracy.

6. Summary and Outlook

With the rapid development of deep learning technology, it is actively applied in welding research. Developing a model for application of deep learning technology is an important part of research, but evaluation of the model by assessing the performance of the proposed deep learning technology also plays a key role in research. The excellence of the model can be verified through visualization and performance indices in classification and regression models. If an error occurs in prediction through the developed model, it may be due to various causes such as lack of reproducibility and nonlinearity of physical phenomena, errors in the measurement process, and the performance of the model itself. It is necessary to identify the cause of the error through the development of a visualization method.

Although the deep learning model is oriented for end-to-end learning, it is necessary to identify the causes of the error through the development of adequate evaluation indices and visualization methods.

Acknowledgement

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008750) supervised by the Korea Institute for Advancement of Technology (KIAT)

ORCID: Kidong Lee: <http://orcid.org/0000-0001-7565-1965>

ORCID: Soong Keun Hyun: <http://orcid.org/0000-0002-3434-8465>

ORCID: Cheolhee Kim: <http://orcid.org/0000-0003-4127-3171>

References

1. K. Andersen, G. E. Cook, G. Karsai and K. Ramaswamy, Artificial Neural Networks Applied to Arc Welding Process Modeling and Control, *IEEE Trans. Ind. Appl.* 26(5) (1990) 824-830.
<https://doi.org/10.1109/28.60056>
2. K. Bae and S.-J. Na, A Study of Vision-Based Measurement of Weld Joint Shape Incorporating the Neural Network, *Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf.* 208(1) (1994) 61-69.
https://doi.org/10.1243/PIME_PROC_1994_208_060_02
3. G. E. Cook, R. J. Barnett, K. Andersen and A. M. Strauss, Weld Modeling and Control Using Artificial Neural Networks, *IEEE Trans. Ind. Appl.* 31(6) (1995) 1484-1491.
<https://doi.org/10.1109/28.475745>
4. H. S. Moon and S. J. Na, A Neuro-Fuzzy Approach to Select Welding Conditions for Welding Quality Improvement in Horizontal Fillet Welding, *J. Manuf. Syst.* 15(6) (1996) 392-403.
[https://doi.org/10.1016/S0278-6125\(97\)83053-1](https://doi.org/10.1016/S0278-6125(97)83053-1)
5. K. Lee, S. Yi, S. Hyun and C. Kim, Review on the Recent Welding Research With Application of CNN-Based Deep Learning - Part I: Models and Applications, *J. Weld. Join.* 39(1) (2021) 10-19.
<https://doi.org/10.5781/JWJ.2021.39.1.1>
6. A. Khumaidi, E. M. Yuniarno and M. H. Purnomo, Welding Defect Classification Based on Convolution Neural Network (CNN) and Gaussian Kernel, *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia (2017) 2 61-265.
<https://doi.org/10.1109/ISITIA.2017.8124091>
7. S. Choi, I. Hwang, Y. Kim, B. Kang and M. Kang, Prediction of the Weld Qualities Using Surface Appearance Image in Resistance Spot Welding, *Met.* 9(8) (2019) 831.
<https://doi.org/10.3390/met9080831>

8. B. Zhang, K.-M. Hong and Y. C. Shin, Deep-Learning-Based Porosity Monitoring of Laser Welding Process, *Manuf. Lett.* 23 (2020) 62-66.
<https://doi.org/10.1016/j.mfglet.2020.01.001>
9. T. Ashida, A. Okamoto, K. Ozaki, M. Hida and T. Yamashita, Development of Image Sensing Technology for Automatic Welding (Image Recognition by Deep Learning), *Kobelco Technol. Rev.* 37(April) (2019) 77-81.
10. C. V. Dung, H. Sekiya, S. Hirano, T. Okatani and C. Miki, A Vision-Based Method for Crack Detection in Gusset Plate Welded Joints of Steel Bridges Using Deep Convolutional Neural Networks, *Autom. Constr.* 102 (2019) 217-229.
<https://doi.org/10.1016/j.autcon.2019.02.013>
11. Y. Zhang, D. You, X. Gao, N. Zhang and P. P. Gao, Welding Defects Detection Based on Deep Learning with Multiple Optical Sensors During Disk Laser Welding of Thick Plates, *J. Manuf. Syst.* 51 (2019) 87-94.
<https://doi.org/10.1016/j.jmsy.2019.02.004>
12. D. Bacioiu, G. Melton, M. Papaalias and R. Shaw, Automated Defect Classification of Aluminium 5083 TIG Welding Using HDR Camera and Neural Networks, *J. Manuf. Process.* 45 (2019) 603-613.
<https://doi.org/10.1016/j.jmapro.2019.07.020>
13. H. Zhu, W. Ge and Z. Liu, Deep Learning-Based Classification of Weld Surface Defects, *Appl. Sci.* 9(16) (2019) 3312.
<https://doi.org/10.3390/app9163312>
14. W. Jiao, Q. Wang, Y. Cheng and Y. Zhang, End-To-End Prediction of Weld Penetration: A Deep Learning and Transfer Learning Based Method, *J. Manuf. Process.* (2020) *In-Press, Available online 4 February 2020.*
<https://doi.org/10.1016/j.jmapro.2020.01.044>
15. Z. Zhang, G. Wen and S. Chen, Weld Image Deep Learning-Based On-Line Defects Detection Using Convolutional Neural Networks for Al Alloy in Robotic Arc Welding, *J. Manuf. Process.* 45 (2019) 208-216.
<https://doi.org/10.1016/j.jmapro.2019.06.023>
16. Y. Yang, L. Pan, J. Ma, R. Yang, Y. Zhu, Y. Yang and L. Zhang, A High-Performance Deep Learning Algorithm for the Automated Optical Inspection of Laser Welding, *Appl. Sci.* 10(3) (2020) 933.
<https://doi.org/10.3390/app10030933>
17. Z. Guo, S. Ye, Y. Wang and C. Lin, Resistance Welding Spot Defect Detection with Convolutional Neural Networks, *Proceeding of International Conference on Computer Vision Systems*, Thessaloniki, Greece (2017) 169-174.
https://doi.org/10.1007/978-3-319-68345-4_15
18. N. Yang, H. Niu, L. Chen and G. Mi, X-ray Weld Image Classification Using Improved Convolutional Neural Network, *AIP Conference Proceedings*, 1995 (2018) 020035.
<https://doi.org/10.1063/1.5048766>
19. J. K. Park, W. H. An and D. J. Kang, Convolutional Neural Network Based Surface Inspection System for Non-patterned Welding Defects, *Int. J. Precis. Eng. Manuf.* 20(3) (2019) 363-374.
<https://doi.org/10.1007/s12541-019-00074-4>
20. W. Hou, Y. Wei, Y. Jin and C. Zhu, Deep Features Based on a DCNN Model for Classifying Imbalanced Weld Flaw Types, *Meas.* 131 (2019) 482-489.
<https://doi.org/10.1016/j.measurement.2018.09.011>
21. Z. Zhang, B. Li, W. Zhang, R. Lu, S. Wada and Y. Zhang, Real-Time Penetration State Monitoring Using Convolutional Neural Network for Laser Welding of Tailor Rolled Blanks, *J. Manuf. Syst.* 54 (2020) 348-360.
<https://doi.org/10.1016/j.jmsy.2020.01.006>
22. A. Muniategui, B. Hériz, L. Eciolaza, M. Ayuso, A. Iturrioz, I. Quintana and P. Ivarez, Spot Welding Monitoring System Based on Fuzzy Classification and Deep Learning, *Proceeding of 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* Naples, Italy (2017) 1-6.
<https://doi.org/10.1109/fuzz-ieee.2017.8015618>
23. A. C. Müller and S. Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, *O'Reilly Media, Inc.*, California, USA (2016).
24. R. M. Haralick, K. Shanmugam and I. H. Dinstein, Textural Features for Image Classification, *IEEE Trans. Syst. Man. Cybern.* SMC-3 (6) (1973) 610-621.
<https://doi.org/10.1109/TSMC.1973.4309314>

합성곱 신경망기반 딥러닝의 용접연구 적용

Part II: 모델의 평가와 시각화

Review on the Recent Welding Research with Application of CNN-Based Deep Learning Part II: Model Evaluation and Visualizations

이기동^{*,**} · 이 성^{*} · 현승균^{**} · 김철희^{*,***}

^{*}포틀랜드주립대학 기계재료공학과

^{**}인하대 신소재공학과

^{***}한국생산기술연구원 용접접합그룹

1. 서 론

최근 다양한 산업분야에 기계학습의 도입이 활발하며, 그 중 하나인 신경망에 대한 관심이 높아지고 있다. 용접분야에서 신경망을 활용한 연구는 1990년대부터 소개되었으며¹⁻⁴⁾, 최근 딥러닝(deep learning)과 합성곱 신경망(Convolutional Neural Network, CNN)을 용접에 적용하려는 연구가 다수 출판되고 있다. 합성곱 신경망의 경우 이미지 인식에 강력한 성능을 보여주고 있으며, 기존에 사용하던 특징(feature) 추출 후 특징과 결과를 학습하는 방식에 비해서 일반화된 학습이 가능하여 용접적용이 급증하고 있다.

본 연구의 I편⁵⁾에서는 CNN의 기본 구조 및 학습법을 소개하고, 용접분야에 CNN을 적용한 논문들⁶⁻²²⁾을 지도학습, 전이학습(transfer learning)의 적용여부 및 데이터 전처리에서 특징 추출이나 데이터증식 여부를 기준으로 5개 그룹으로 구분하고, 사용된 모델의 특성 및 응용분야를 소개하였다.

합성곱 신경망의 경우 차원이 복잡하여 기존의 선형 회귀 등 전통적이고 직관적인 모델에 비해 결과의 평가와 시각화가 용이하지 않다. 본 논문에서는 발표 연구에서 사용한 평가 지표, 평가 방법 및 시각화 방법으로 해당논문들을 Table 1과 같이 분류하여 적용사례를 소개하고자 한다.

2. 평가지표

신경회로망 모델링을 통해 개발된 분류 및 회귀 모델의 적합성을 평가하기 위해서는 정확도와 학습 속도와 같은 평가 지표를 이용한다. 모델의 성능을 평가 지표와 다른 모델과의 비교 등을 통해 상대 비교를 수행을

병행하는 경우가 대부분이지만 일부 논문에서는 평가 지표만으로 제안된 모델의 성능을 평가하기도 한다.

2.1 분류 모델 (classification model)

이미지 기반 CNN을 가장 많이 활용하는 분야는 지도학습을 통해 분류 작업을 수행하는 것이다. 분류 모델에서는 예측 결과에 따라 각 결과를 오차 행렬(confusion matrix)로 나타낼 수 있으며²³⁾, 가장 간단한 형태인 이진 분류에서의 오차 행렬을 Fig. 1에 나타내었다.

이진 분류모델의 예측 결과는 참 양성(True Positive, TP), 거짓 양성(False Positive, FP), 참 음성(True Negative, TN), 거짓 음성(False Negative, FN)로 구분된다. 이 중 참 양성과 참 음성은 예측을 잘 해낸 경우이며, 거짓 양성과 거짓 음성이 오차(error)에 해당된다. 거짓 양성은 음성을 양성이라 잘못 분류한 오차이고, 거짓 음성은 양성을 음성으로 분류한 오차이다.

일부 연구에서는 각 구분 결과의 비율을 나타낸 참 양성 비율(True Positive Rate, TPR), 거짓 양성 비율(False Positive Rate, FPR), 참 음성 비율(True Negative Rate, TNR), 거짓 음성 비율(False Negative Rate, FNR)을 채택하기도 한다.

분류 데이터셋이 균형적으로 이루어졌을 때는 전체 데이터셋 중 옳게 분류한 비율인 정확도(accuracy)가 평가 지표로 사용하기도 하며, 반대로 전체 데이터셋 중 잘못 분류된 경우를 이용한 평균 오차율(mean Error Rate)을 사용하기도 한다.

$$(accuracy) = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

$$(mean\ Error\ Rate) = \frac{FN + FP}{TP + FN + FP + TN} \quad (2)$$

Table 1 Summary of research papers reviewed in this study

Classification / Regression	Evaluation index	Comparison with		Visualization	1stAuthor (Publication Year)	Ref. No.
		Conventional method	Neural Networks			
Classification	Accuracy				A. Khumaidi (2017)	6)
Regression/ Classification	Accuracy				S. Choi (2019)	7)
Classification	Accuracy, Confusion matrix				B. Zhang (2020)	8)
Regression	MAPE*	Edge detection			T. Ashida (2019)	9)
Classification	Accuracy, Recall, F1 score		SCNN* ⁸ /CNN (pretrained and tunned models)		C. V. Dung (2019)	10)
Classification	Mean error rate		FCN* ⁹		Y. Zhang (2019)	11)
Classification	Accuracy, Training time		FCN, CNN(No. of Class)		D. Bacioiu (2019)	12)
Classification	Accuracy		CNN (Classification module)		H. Zhu (2019)	13)
Classification	Accuracy		CNN (with and without transfer learning)		W. Jiao (2020)	14)
Classification	Accuracy, Training Time		CNN (reference)	Intermediate activations	Z. Zhang (2019)	15)
Classification	Accuracy, Recall, Training time		AlexNet, VGG-16, Resnet-50, Densenet-121, Mobile NetV3-Large (pretrained and tunned models)	Intermediate activations	Y. Yang (2020)	16)
Classification	Training time, Mean error rate	Visual inspection of expert	CNN (No. of layers)		Z. Guo (2017)	17)
Classification	Accuracy	SVM* ⁴	CNN (Proposed model, Transfer learning), FCN		N. Yang (2018)	18)
Regression/ Classification	TPR** ² , TNR* ³	Hough Circle/ HOG* ⁵ +SVM, LBP* ⁶ +SVM	CNN		J.-K. Park (2019)	19)
Classification	Accuracy	Haralick feature, HOG feature	SSAE* ¹⁰ , CNN	t-SNE* ¹²	W. Hou (2019)	20)
Classification	Accuracy, Recall	HOG, LBP, HOG+LBP, BOF* ⁷	SAE* ¹¹ , CNN(No. of layers)	t-SNE	Z. Zhang (2020)	21)
Classification (Auto-encoder)	Accuracy				A. Muniategui (2017)	22)

*Mean Absolute Percentage Error; **True Positive Rate; *³True Negative Rate; *⁴Support Vector Machine;

*⁵Histogram of Oriented Gradient; *⁶Local Binary Pattern; *⁷Back Of Features; *⁸Shallow ConvNet;

*⁹Fully Connected Net; *¹⁰Stacked Sparse Autoencoder; *¹¹Sparse autoencoder;

*¹²t-Distributed Stochastic Neighbor Embedding

본 논문은 독자의 이해를 돕기위하여 영문논문을 국문으로 번역하여 게재한 논문입니다. 저자는 본 논문으로 연구업적과 같은 실적에 중복으로 지원받거나 인정받을 수 없음을 알려드립니다.

		Prediction	
		Positive	Negative
Actual value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Fig. 1 Confusion matrix for binary classification

하지만 데이터셋이 불균형할 경우, 지표가 편중될 수 있기 때문에 정밀도(precision), 재현율(recall)과 같이 보완하는 지표가 필요하다. 모델이 참이라고 분류한 데이터 중 실제 참인 경우의 비율을 정밀도, 실제 참인 경우 중 참으로 분류한 경우의 비율을 재현율이라 한다.

$$(precision) = \frac{TP}{TP + FP} \quad (3)$$

$$(recall) = \frac{TP}{TP + FN} \quad (4)$$

정밀도와 재현율은 상호 보완적인 지표이며 두 값이 모두 높을수록 좋은 분류 모델이다. 또한, 정밀도와 재현율의 조화 평균인 F1 지표(F1 score)를 사용할 수 있다.

$$(F1\ score) = 2 \times \frac{1}{\frac{1}{(precision)} + \frac{1}{(recall)}} \quad (5)$$

$$= 2 \times \frac{(precision) \times (recall)}{(precision) + (recall)}$$

분류 모델에 CNN을 적용한 경우 대부분의 경우 평가 지표로 평균 정확도를 사용하였다^{6-8,10,13-16,18,20,21}. Y. Yang et al.¹⁶과 Z. Zhang et al.²¹은 평균 정확도와 더불어 재현율도 평가 지표로 사용하였다. 특히 C. V. Dung et al.¹⁰은 평균 정확도, 재현율, F1 지표 또한 사용하였다. Z. Zhang et al.²¹은 평균 정확도와 함께 오차 행렬을 제시하였다. Y. Zhang et al.¹¹와 Z. Guo et al.¹⁷은 평균 정확도 대신 평균 오차율을 사용하였다. J.-K. Park et al.¹⁹은 엔진 트랜스미션과 같이 안전을 중요시하는 부품 제작 공정에서 불량품 판별의 중요성을 강조하며, 최종적으로 참

양성 비율과 함께 불량품 판별 정확도인 참 음성 비율을 평가 지표로 사용하였다.

2.2 회귀 모델의 평가 지표

회귀 모델에서는 실제값과 예측값의 오차에 기반한 평균 제곱 오차(Mean Square Error, MSE)를 손실 함수로 사용하며 평균 절대 오차(Mean Absolute Error, MAE)를 평가 지표로 가장 많이 사용한다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y'(i) - y(i))^2 \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'(i) - y(i)| \quad (7)$$

여기에서 n은 총 데이터의 개수, i는 데이터의 순서이며, y(i)는 i번째 데이터의 측정값, y'(i)은 회귀모델을 통한 i번째 데이터의 예측값을 나타낸다. 평균 절대 오차가 0에 가까울수록 회귀 모델의 적합도가 높다. 또한 평균 절대 오차를 비율로 나타낸 평균 절대 오차율(Mean Absolute Percentage Error, MAPE)은 식 8과 같이 정의되며, 반대 개념인 평균 정확도 (mean accuracy)는 식 9와 같이 정의된다.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y'(i) - y(i)|}{y'(i)} \quad (8)$$

$$(mean\ accuracy) = 1 - MAPE \quad (9)$$

회귀 모델을 통해 얻은 예측값이 측정값을 얼마나 잘 설명하는지를 나타내는 결정 계수(Coefficient of Determination, R²)을 사용하기도 한다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y'(i) - y(i))^2}{\sum_{i=1}^n (y'(i) - \bar{y}'(i))^2} \quad (10)$$

$\bar{y}'(i)$ 는 예측값의 평균을 나타낸다. 결정 계수는 0 ≤ R² ≤ 1 범위이며 1에 가까울수록 회귀 모델의 정확도가 높다. 회귀 모델에 CNN 모델을 적용한 S. Choi et al.⁷ 와 T. Ashida et al.⁹는 각각 평가 지표로 평균 정확도와 평균 오차를 사용하였다.

3. 다른 모델과 비교평가

이미지에 기반하여 용접 현상을 예측하기 위한 다양한 회귀 및 분류 모델이 존재한다. 여러 논문에서는 제

안된 CNN 모델을 전통적 분류기 또는 다른 신경망 모델과 비교하고 모델의 성능을 검증하였다.

3.1 전통적 특징 추출 방법 및 분류기와 비교

분류 및 회귀를 위해 신경망 외에도 다양한 방법들이 전통적으로 사용되어 왔다. 이 문단에서 설명하는 기존 모델은 신경망 이전에 사용하던 방법들을 구분하여 설명하며, 이미지의 직선 추출을 위한 윤곽 측정 알고리즘, 허프 변환(Hough transform)과 같은 이미지 처리 방법, 기존에 공정에 적용해왔던 숙련자의 육안 검사와 전통적 분류기 방법들을 포함한다.

T. Ashida et al.은 제안된 CNN 회귀 모델을 전통적 방법인 윤곽 측정 이미지 처리 방법과 비교하여 응용폴 폭과 선단 길이의 예측 시 오차를 줄일 수 있음을 확인했다⁹⁾. Z. Guo et al.은 숙련자의 육안 검사와 비교하여 CNN 모델의 성능을 평가하였다¹⁷⁾.

신경망이 아닌 전통적 이미지 분류모델은 특징 추출과 분류기로 나누어 설명할 수 있다. 특징 추출 방법으로는 HOG (Histogram of Oriented Gradient), LBP (Local Binary Pattern), 하랄릭 특성(Haralick feature)²⁴⁾ 등이 있고 분류기로는 SVM (Support Vector Machine)이 자주 사용된다. 전통적 특징 추출 방법들의 성능을 높이기 위해서는 특징을 강조시키는 이진화 또는 윤곽 측정과 같은 이미지 처리가 선행된다. HOG는 이미지를 일정 크기로 분할하고, 기울기 방향 분포를 통해 국부적인 셀을 정규화하여 특징을 추출하는 알고리즘이며, LBP는 이미지 모든 픽셀 주변 3x3 영역을 상대적인 밝기로 0 또는 1로 이진화하는 알고리즘이다. LBP는 원래 텍스처(texture)를 분류하기 위해 개발되었으나 얼굴인식과 같은 이미지 분석에 사용된다. 또한, Haralick feature는 인접 픽셀의 그레이레벨 변화도에 기반한 특징 추출 알고리즘이다. SVM은 기계학습을 이용한 분류 알고리즘으로 초평면(hyper-plane)의 집합으로 이루어져 있다. 높은 차원에서 데이터 클래스 간의 거리를 최대화 하는 초평면(최대-마진 초평면, Maximum-margin hyper-plane)을 이용하여 분류하는 분류기이다.

W. Hou et al.²⁰⁾, N. Yang et al.¹⁸⁾, Z. Zhang et al.²¹⁾, J.-K. Park et al.¹⁹⁾은 전통적 분류기와 제안된 CNN 모델을 비교하였다. 성능 평가를 위한 비교에 W. Hou et al.는 HOG 특성, 하랄릭 특성 추출 방법을 사용하였고²⁰⁾, N. Yang et al.은 SVM을 분류기로 사용하였다¹⁸⁾. Z. Zhang et al.은 HOG, LBP, HOG+LBP, BOF 와 같이 다양한 전통적 분류기와

제안된 CNN 모델을 비교하였다²¹⁾. J.-K. Park et al.은 엔진 트랜스미션 용접부의 중심점 위치 추적의 성능을 평가를 위해 허프 원(Hough circle) 윤곽 추출 알고리즘과 CNN 회귀모델을 비교하였으며, CNN 분류모델과 HOG+SVM, LBP+SVM을 비교하여 성능을 평가하였다¹⁹⁾.

3.2 타 신경회로망 모델과 비교

이미지의 인식에는 신경망 중 비지도 학습 신경망인 SAE(Sparse Auto-Encoder) 이미지 특징 추출 방법과 합성곱 신경망(이 주로 사용된다. SAE는 오토엔코더의 한 종류로 이미지 특징 추출을 위해 자주 사용되는 비지도 학습이다. 데이터 평탄화를 통해 완전 연결 신경망(Fully Connected Neural network, FCN) 또한 이미지 특징 추출에 사용될 수 있다.

N. Yang et al.¹⁸⁾, Y. Zhang et al.¹¹⁾, D. Bacioiu et al.¹²⁾은 제안된 CNN 분류 모델과 FCN 모델을 비교하여 성능 평가를 하였다. 또한, W. Hou et al.²⁰⁾과 Z. Zhang et al.²¹⁾은 각각 SAE 두 개 층을 사용하는 SSAE와 SAE를 사용하여 성능 평가를 하였다.

CNN 모델은 목적에 따른 구조 설계나 전이 학습 적용 등 다양한 방법으로 최적화가 가능하여 다른 구조 최적화를 통해 CNN 모델의 성능을 향상시킬 수 있으므로 다양한 CNN 모델구조와 모델링 기법을 활용하여 모델링 성능을 평가할 수 있어 리부한 논문의 과반 이상의 논문에서 CNN 모델 간의 비교를 통해 성능을 평가하였다^{10,12,13,15-21)}.

D. Bacioiu et al.은 분류 클래스가 2개, 4개, 6개로 복잡해짐에 따라 제안한 CNN 모델 정확도의 차이를 비교하였다¹²⁾. C. V. Dung et al.은 SCNN (Shallow CNN)과 BN(Bottle Neck)-CNN, 미세조정(FT, Fine Tuning) CNN 등 세 가지 CNN 모델을 평균 정확도, 재현율, F1 지수를 평가 지표로 비교하였다. SCNN과 VGG-16 기반 전이 학습을 적용한 CNN 모델을 비교하여 전이 학습에서 성능이 우수함을 보여 주었다. 또한 BN-CNN 모델과 FT-CNN 모델을 비교하여 미세조정을 통해 같은 전이 학습 모델에서도 최적화를 통해 정확도를 높일 수 있음을 보여주었다¹⁰⁾.

W. Jiao et al.은 일반적인 CNN 모델 구조와 전이 학습 CNN 모델의 정확도를 비교하였다. 이를 통해 전이 학습의 적용이 적은 양의 훈련 데이터를 이용한 학습에 효과적임을 확인하였다¹³⁾.

H. Zhu et al.은 CNN 모델에서 분류기별 정확도를 비교하였다. CNN 모델을 통해 특성 맵을 추출하고, 주로 사용되는 분류 방법인 소프트맥스, SVM과 비교

하여 제안한 랜덤 포레스트(random forest) 방법이 우수함을 주장하였다¹³⁾.

Y. Zhang et al.은 FCN과 제안된 CNN 모델의 정확도 비교를 통해 성능 평가를 하였다. 추가로 오분류 샘플에 대한 분석을 통해 오차의 원인을 파악하였다¹¹⁾.

4. 시각화

리뷰한 논문에서 시각화 방법을 통해 모델을 평가하는 것은 크게 두 가지로 나누어진다. 중간층 출력을 통해 분류 모델의 특징 추출을 확인하는 방법과 t-SNE를 통해 분류를 2D 그래프를 통해 보여주는 방법이 제시되었다.

4.1 중간층 출력

CNN 학습 과정 합성곱 필터를 거친 특성 맵(feature map)은 이미지로 표현할 수 있으며 낮은 층에서는 강조된 특성을 어느 정도 확인할 수 있다. 중간층 출력은 이러한 낮은 층의 특성 맵을 추출하여 강조된 특성이 모델의 목적을 위한 특성인지 확인하거나 높은 층에서 유의미한 특성의 개수를 확인하여 평가 방법으로 사용한다.

Y. Yang et al.은 최적화된 CNN 모델의 첫 번째 층 이미지를 중간층(intermediate activations)으로 가시화하여 보여주면서 목표로 하는 이미지의 특징이 강조되고 있음을 확인하였다¹⁶⁾.

Z. Zhang et al.은 각 CNN 레이어에서 특성맵을 이미지화하여 출력하였다. 각 레이어의 특성맵에서 유의미한 픽셀을 가진 이미지(모든 픽셀의 값이 0이 아닌 이미지)의 개수를 이용하여 제안된 모델이 효과적으로 특성을 추출하고 있음을 확인하였다¹⁵⁾.

4.2 t-SNE

t-SNE(t-Distributed Stochastic Neighbor Embedding)는 고차원 데이터 세트를 시각화에 적합한 비선형 차원 축소 기술이다. 데이터의 유사성을 2차원 또는 3차원에서의 거리로 표현하는 기술이다. 정규 분포를 전제로 한 고차원 공간의 이웃할 확률과 저차원 공간의 이웃할 확률을 쿨백-라이블러 발산(Kullback-Leibler divergence)으로 계산하여 최소화하는 방법을 학습하고 시각화하는 방법이다.

W. Hou et al.은 하랄릭 특성, HOG와 SSAE, 그리고 깊이가 다른 두 가지 CNN 각각의 방법으로 특성을 추출하고 이 특성들을 t-SNE를 통해 2D 맵으로

시각화하였다. 또한 각 방법을 이용한 분류 결과의 평균 정확도를 비교하여 제안한 모델의 성능을 평가하였다. 논문에서 제안된 CNN 모델이 다른 방법에 비해 t-SNE 2D 맵에서 더 명확하게 구분되었으며, 시각화된 2D 맵에서 명확히 분류될수록 평균 정확도가 높은 것을 확인하였다²⁰⁾.

Z. Zhang et al.은 최적화된 CNN 모델에서 특성을 추출하고 이를 t-SNE를 통해 2D 맵으로 시각화하였다. 각 그룹 간의 분리는 맵에서 확인되었으며, 분리가 명확하지 않은 구간을 분석하여 오차의 원인을 제시하였다²¹⁾.

5. 시각화

앞서 설명한 비지도 신경망인 오토 엔코더는 다양한 종류가 있는데, 그 중 합성곱 연산을 사용한 합성곱 오토 엔코더(Convolutional Auto-Encoder)는 이미지 압축에 유리하다는 장점이 있다. 오토 엔코더는 인코딩(encoding)으로 압축된 이미지와 디코딩(decoding)으로 복원된 이미지와 MSE를 비교하여 자동으로 학습된다. 분류 모델에서 이미지가 압축되면서 분류에 필요한 특징들을 유지하고 있는가는 최종적인 분류 모델의 성능의 정확도를 통해 평가할 수 있다.

A. Muniategui et al.은 심층 합성곱 오토 엔코더(deep convolutional auto-encoder)를 사용하여 이미지를 압축하여 퍼지 모델의 입력으로 사용하고자 하였다²²⁾. 압축된 15×15픽셀의 이미지에서 225개 벡터를 추출하였고 매우 양호(Very Good), 양호(Good), 보통(Regular), 불량(Bad), 매우 불량(Very Bad)의 예측 클래스별로 그룹화하였다. 압축된 이미지를 평가하기 위하여 그룹화시킨 벡터의 평균을 각 패턴과 비교하였다. 제안한 네 가지 퍼지 분류 모델의 매우 양호, 양호, 보통의 패턴은 유사하였으며, 나쁨 및 매우 나쁨의 경우에는 패턴은 분류 모델에 따라 특이성을 갖는 것을 확인하였다. 여기서 압축된 이미지에서도 생산 부품의 유사성으로 인한 압축 이미지 벡터의 유사성을 확인하였고, 각 분류 모델의 차이점에 관련된 특정 패턴도 존재함을 확인하였다. 최종적으로 정확도를 확인하여 심층 합성곱 오토엔코더를 사용하여 이미지 압축을 하였을 때의 퍼지 분류 모델의 성능을 평가하였다.

6. 시사점과 전망

딥러닝 기술의 발달로 융접 분야에서도 딥러닝 기술이 활발히 적용되고 있다. 딥러닝 기술의 적용을 위해서 모델을 구성하는 것도 중요한 부분이지만 제안된 딥

러닝 기술의 성능을 평가하여 모델을 평가하는 것 또한 중요한 부분이다. 분류 및 회귀모델에서 성능을 보여주는 각 지표와 가시화를 통해 모델의 우수성을 검증할 수 있다. 모델을 통한 예측에서 오차가 발생할 경우 물리적 현상의 재현성 부족과 비선형성, 측정단계에서의 오차, 모델 자체의 성능 등 다양한 원인에 기인할 수 있으므로 딥러닝 모델이 종단간 학습을 지향하지만 충분한 평가지표 및 시각화 방안의 개발을 통해 오차의 원인을 파악할 필요가 있다.

Acknowledgement

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008750) supervised by the Korea Institute for Advancement of Technology (KIAT)

ORCID: Kidong Lee: <http://orcid.org/0000-0001-7565-1965>

ORCID: Soong Keun Hyun: <http://orcid.org/0000-0002-3434-8465>

ORCID: Cheolhee Kim: <http://orcid.org/0000-0003-4127-3171>

References

1. K. Andersen, G. E. Cook, G. Karsai and K. Ramaswamy, Artificial Neural Networks Applied to Arc Welding Process Modeling and Control, *IEEE Trans. Ind. Appl.* 26(5) (1990) 824-830.
<https://doi.org/10.1109/28.60056>
2. K. Bae and S.-J. Na, A Study of Vision-Based Measurement of Weld Joint Shape Incorporating the Neural Network, *Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf.* 208(1) (1994) 61-69.
https://doi.org/10.1243/PIME_PROC_1994_208_060_02
3. G. E. Cook, R. J. Barnett, K. Andersen and A. M. Strauss, Weld Modeling and Control Using Artificial Neural Networks, *IEEE Trans. Ind. Appl.* 31(6) (1995) 1484-1491.
<https://doi.org/10.1109/28.475745>
4. H. S. Moon and S. J. Na, A Neuro-Fuzzy Approach to Select Welding Conditions for Welding Quality Improvement in Horizontal Fillet Welding, *J. Manuf. Syst.* 15(6) (1996) 392-403.
[https://doi.org/10.1016/S0278-6125\(97\)83053-1](https://doi.org/10.1016/S0278-6125(97)83053-1)
5. K. Lee, S. Yi, S. Hyun and C. Kim, Review on the Recent Welding Research With Application of CNN-Based Deep Learning - Part I: Models and Applications, *J. Weld. Join.* 39(1) (2021) 10-19.
6. A. Khumaidi, E. M. Yuniarno and M. H. Purnomo, Welding Defect Classification Based on Convolution Neural Network (CNN) and Gaussian Kernel, *International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia (2017) 261-265.
<https://doi.org/10.1109/ISITIA.2017.8124091>
7. S. Choi, I. Hwang, Y. Kim, B. Kang and M. Kang, Prediction of the Weld Qualities Using Surface Appearance Image in Resistance Spot Welding, *Met.* 9(8) (2019) 831.
<https://doi.org/10.3390/met9080831>
8. B. Zhang, K.-M. Hong and Y. C. Shin, Deep-Learning-Based Porosity Monitoring of Laser Welding Process, *Manuf. Lett.* 23 (2020) 62-66.
<https://doi.org/10.1016/j.mfglet.2020.01.001>
9. T. Ashida, A. Okamoto, K. Ozaki, M. Hida and T. Yamashita, Development of Image Sensing Technology for Automatic Welding (Image Recognition by Deep Learning), *Kobelco Technol. Rev.* 37(April) (2019) 77-81.
10. C. V. Dung, H. Sekiya, S. Hirano, T. Okatani and C. Miki, A Vision-Based Method for Crack Detection in Gusset Plate Welded Joints of Steel Bridges Using Deep Convolutional Neural Networks, *Autom. Constr.* 102 (2019) 217-229.
<https://doi.org/10.1016/j.autcon.2019.02.013>
11. Y. Zhang, D. You, X. Gao, N. Zhang and P. P. Gao, Welding Defects Detection Based on Deep Learning with Multiple Optical Sensors During Disk Laser Welding of Thick Plates, *J. Manuf. Syst.* 51 (2019) 87-94.
<https://doi.org/10.1016/j.jmsy.2019.02.004>
12. D. Bacioiu, G. Melton, M. Papaelias and R. Shaw, Automated Defect Classification of Aluminium 5083 TIG Welding Using HDR Camera and Neural Networks, *J. Manuf. Process.* 45 (2019) 603-613.
<https://doi.org/10.1016/j.jmapro.2019.07.020>
13. H. Zhu, W. Ge and Z. Liu, Deep Learning-Based Classification of Weld Surface Defects, *Appl. Sci.* 9(16) (2019) 3312.
<https://doi.org/10.3390/app9163312>
14. W. Jiao, Q. Wang, Y. Cheng and Y. Zhang, End-To-End Prediction of Weld Penetration: A Deep Learning and Transfer Learning Based Method, *J. Manuf. Process.* (2020) *In-Press, Available online 4 February 2020.*
<https://doi.org/10.1016/j.jmapro.2020.01.044>
15. Z. Zhang, G. Wen and S. Chen, Weld Image Deep Learning-Based On-Line Defects Detection Using Convolutional Neural Networks for Al Alloy in Robotic Arc Welding, *J. Manuf. Process.* 45 (2019) 208-216.
<https://doi.org/10.1016/j.jmapro.2019.06.023>
16. Y. Yang, L. Pan, J. Ma, R. Yang, Y. Zhu, Y. Yang and L. Zhang, A High-Performance Deep Learning Algorithm for the Automated Optical Inspection of Laser Welding, *Appl. Sci.* 10(3) (2020) 933.
<https://doi.org/10.3390/app10030933>
17. Z. Guo, S. Ye, Y. Wang and C. Lin, Resistance Welding Spot Defect Detection with Convolutional Neural Networks, *Proceeding of International Conference on*

- Computer Vision Systems*, Thessaloniki, Greece (2017) 169-174.
https://doi.org/10.1007/978-3-319-68345-4_15
18. N. Yang, H. Niu, L. Chen and G. Mi, X-ray Weld Image Classification Using Improved Convolutional Neural Network, *AIP Conference Proceedings*, 1995 (2018) 020035.
<https://doi.org/10.1063/1.5048766>
19. J. K. Park, W. H. An and D. J. Kang, Convolutional Neural Network Based Surface Inspection System for Non-patterned Welding Defects, *Int. J. Precis. Eng. Manuf.* 20(3) (2019) 363-374.
<https://doi.org/10.1007/s12541-019-00074-4>
20. W. Hou, Y. Wei, Y. Jin and C. Zhu, Deep Features Based on a DCNN Model for Classifying Imbalanced Weld Flaw Types, *Meas.* 131 (2019) 482-489.
<https://doi.org/10.1016/j.measurement.2018.09.011>
21. Z. Zhang, B. Li, W. Zhang, R. Lu, S. Wada and Y. Zhang, Real-Time Penetration State Monitoring Using Convolutional Neural Network for Laser Welding of Tailor Rolled Blanks, *J. Manuf. Syst.* 54 (2020) 348-360.
<https://doi.org/10.1016/j.jmsy.2020.01.006>
22. A. Muniategui, B. Hériz, L. Eciolaza, M. Ayuso, A. Iturrioz, I. Quintana and P. Ivarez, Spot Welding Monitoring System Based on Fuzzy Classification and Deep Learning, *Proceeding of 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* Naples, Italy (2017) 1-6.
<https://doi.org/10.1109/fuzz-ieee.2017.8015618>
23. A. C. Müller and S. Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, *O'Reilly Media, Inc.*, California, USA (2016).
24. R. M. Haralick, K. Shanmugam and I. H. Dinstein, Textural Features for Image Classification, *IEEE Trans. Syst. Man. Cybern.* SMC-3 (6) (1973) 610-621.
<https://doi.org/10.1109/TSMC.1973.4309314>