AUTOMATIC ESTIMATION OF DETERIORATION LEVEL ON TRANSMISSION TOWERS VIA DEEP EXTREME LEARNING MACHINE BASED ON LOCAL RECEPTIVE FIELD

Keisuke Maeda, Sho Takahashi, Takahiro Ogawa and Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {maeda, sho, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents an automatic estimation method of deterioration levels on transmission towers via Deep Extreme Learning Machine based on Local Receptive Field (DELM-LRF). Although Convolutional Neural Network (CNN) requires a large number of training images, it is difficult to prepare a sufficient number of training images of transmission towers. Thus, we generate a novel estimation method which enables training from a small number of training images. Specifically, we automatically extract image features based on Local Receptive Field (LRF) which combines convolution and pooling without using hand-craft features and estimate deterioration levels via Deep Extreme Learning Machine (DELM), which is a part of efficient deep learning methods. The derivation of DELM-LRF is the biggest contribution of this paper, and it can be trained from less training images compared to CNN. Experimental results show the effectiveness of DELM-LRF for the estimation of deterioration levels on transmission towers. Consequently, the proposed method makes it possible to approach challenging tasks with high expertise having difficulty in preparing enough images.

***Index Terms*—** Deep extreme learning machine, local receptive field, transmission tower, deterioration level estimation.

1. INTRODUCTION

In any country, there are several critical infrastructures such as power grid, oil/gas pipelines, railway, etc. Especially, innumerable transmission towers have been constructed around the world, and maintenance inspection of these infrastructures is important [1]. In order to maintain these transmission towers, visual inspection has been mostly performed by inspectors. However, since there are an enormous number of transmission towers in various areas, more efficient inspection methods in visual inspection are required [2]. It should be noted that the visual inspection which is performed by climbing towers is dangerous. Therefore, for improvement of efficiency of maintenance inspection and reduction of the risk, there have been proposed various automated analysis methods, and they support inspectors by using surface images on transmission towers [2–5]. In particular, an automatic estimation of deterioration levels of transmission towers is an indispensable task for supporting maintenance inspection. Inspectors determine the deterioration levels on the transmission towers by visual inspection. In the actual maintenance operation, although the manual of inspection is prepared, determining the deterioration levels is difficult due to the ambiguity in inspectors' decision. This is because inspectors observe the surface and estimate the deterioration level based on their experience and individual knowledge [6, 7]. Therefore, quantitative analysis on the deterioration levels is necessary. In order to construct the quantitative analysis method, we have promoted collaborative researches with Tokyo Electric Power Company (TEPCO) Research Institute and have analyzed data provided by TEPCO Research Institute.

Recently, in image recognition fields, there have been proposed several image classification methods having high performance such as Convolutional Neural Network (CNN) [8–10] which requires a large number of training images. When it is difficult to prepare a lot of training images, CNN-fine-tuning is used instead of CNN trained from scratch. However, it is reported that the estimation performance of CNN-fine-tuning does not necessarily improve when it is applied to a task for which it was not designed, that is, in case that technical data with a property different from that of the pre-trained data is used [11, 12]. Note that we define the data requiring the professional knowledge as the technical data. In fact, similar tendencies have been confirmed in the experiments in this paper. Therefore, a new approach based on deep learning that can effectively handle a small amount of technical data is necessary.

In this paper, we propose an automatic estimation method of deterioration levels from surface images on transmission towers. We generate a novel estimation method which enables training from a small amount of technical data. Specifically, we automatically extract image features via Local Receptive Field (LRF) [13] which performs convolution and pooling like CNN without using hand-craft features. Furthermore, we estimate deterioration levels based on Deep Extreme Learning Machine (DELM) [14] which is a part of efficient deep learning methods. The number of images required for both LRF and DELM is less than that of CNN. Therefore, construction of Deep Extreme Learning Machine based on Local Receptive Field (DELM-LRF) which integrates LRF and DELM is the biggest contribution of this paper. Although our DELM-LRF enables the estimation in the same manner as CNN, it requires few training images. Consequently, experimental results show that automatic and accurate estimation of deterioration levels is realized by using our DELM-LRF.

In this research, we utilized the data that were provided by Tokyo Electric Power Company Research Institute. This work was partly supported by JSPS KAKENHI Grant Number JP17H01744.

2. DETERIORATION LEVEL ESTIMATION VIA DELM-LRF

In this section, we explain the automatic estimation of deterioration levels via DELM-LRF. The proposed method consists of two procedures as shown in Fig. 1. First, we automatically extract image features from surface images on transmission towers based on LRF in **2.1**. Second, construction of DELM from the extracted features is explained in **2.2**. Then we can obtain the deterioration level based on outputs from the DELM's output layer.

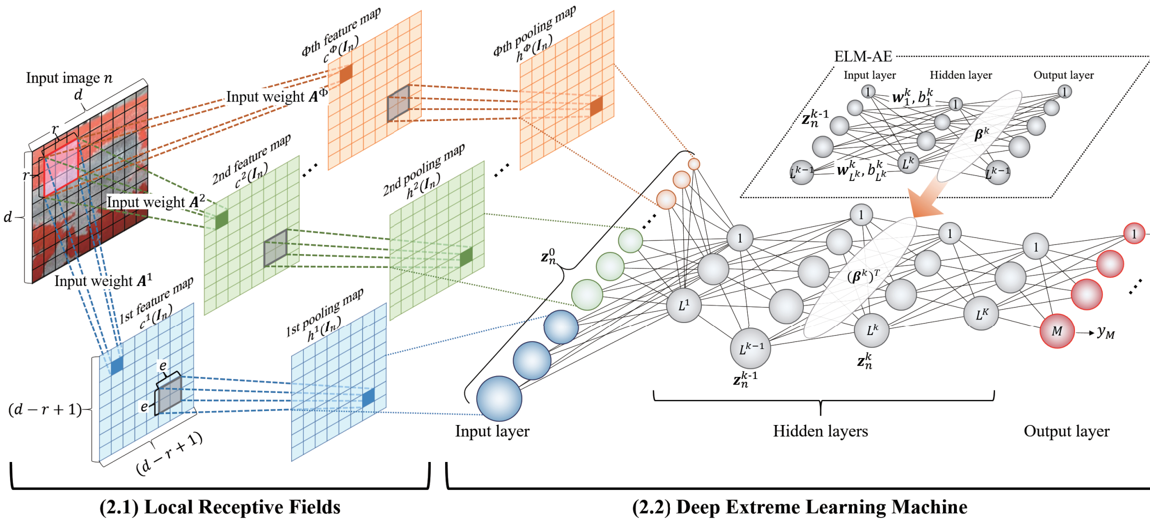


Fig. 1. An overview of DELM-LRF which consists of two procedures. The first procedure is feature extraction by using convolution and pooling based on LRF [13]. The second procedure is estimation of deterioration levels based on DELM [14] which is generated by using ELM-Auto Encoder. The main contribution of this paper is integration of both LRF and DELM.

2.1. Automatic Feature Extraction Based on LRF

Given a training image n ($n = 1, 2, ..., N$; N being the number of training images), we extract image features from an input luminance matrix $\boldsymbol{I}_n \in \mathbb{R}^{d \times d}$ corresponding to image n. Regarding the color channel, we perform the same processing as in [13]. In order to extract image features, the proposed method performs two procedures, which include generation of feature maps and pooling maps as shown in Fig. 1.

First, we randomly generate an initial weight matrix $\hat{\boldsymbol{A}} \in \mathbb{R}^{r^2 \times \Phi}$. Note that $r \times r$ means the receptive fields size, and Φ is the number of feature maps. We orthogonalize the initial weight matrix $\hat{\boldsymbol{A}}$ using the singular value decomposition (SVD) method, and orthogonal vector $\hat{\boldsymbol{a}}^{\phi} \in \mathbb{R}^{r^2}$ ($\phi = 1, 2, ..., \Phi$) is calculated. Since SVD cannot be performed on the column in the case $r^2 < \Phi$, we perform the following steps: 1) $(\hat{\boldsymbol{A}})^{\top}$ is orthogonalized via SVD, 2) transposed back. Thus, an input weight matrix $\boldsymbol{A}^{\phi} \in \mathbb{R}^{r \times r}$, which corresponds to $\hat{\boldsymbol{a}}^{\phi}$ column-wisely, for ϕ th feature map is obtained. By using the obtained input weight matrix $\boldsymbol{A}^{\phi}$, ϕ th feature map $c^{\phi}(\boldsymbol{I}_n)$ is calculated as:

$$\begin{aligned} c^{\phi}_{i,j}(\boldsymbol{I}_n) &= \sum_{s=1}^{r}\sum_{u=1}^{r} \boldsymbol{I}_{n(i+r-s,j+r-u)} \times \boldsymbol{A}^{\phi}_{s,u}, \\ i,\ j &= 1,\ 2,\ ...,\ (d-r+1)\,, \end{aligned} \tag{1}$$

where $c^{\phi}_{i,j}(\boldsymbol{I}_n)$, $\boldsymbol{I}_{n(i+r-s,j+r-u)}$ and $\boldsymbol{A}^{\phi}_{s,u}$ represent (i, j) element of $c^{\phi}(\boldsymbol{I}_n)$, $(i + r - s, j + r - u)$ element of $\boldsymbol{I}_n$ and (s, u) element of $\boldsymbol{A}^{\phi}$, respectively. Consequently, feature map $c^{\phi}(\boldsymbol{I}_n)$ whose size is $(d - r + 1) \times (d - r + 1)$ is calculated.

Second, we calculate pooling maps by using the obtained feature maps. Pooling size used in the proposed method is $e \times e$, and the pooling map is of the same size with the feature map $(d - r + 1) \times (d - r + 1)$. By using a square/square-root pooling, ϕ th pooling map $h^{\phi}(\boldsymbol{I}_n)$ is calculated as:

$$\begin{aligned} h^{\phi}_{p,q}(\boldsymbol{I}_n) &= \sqrt{\sum_{i=p-e}^{p+e}\sum_{j=q-e}^{q+e}(c^{\phi}_{i,j}(\boldsymbol{I}_n))^2}\,, \\ p,\ q &= 1,\ 2,\ ...,\ (d-r+1)\,, \\ &\text{if} \quad (i, j) \text{ is out of bound; } c^{\phi}_{i,j}(\boldsymbol{I}_n) = 0\,, \end{aligned} \tag{2}$$

where $h^{\phi}_{p,q}(\boldsymbol{I}_n)$ represents (p, q) element of $h^{\phi}(\boldsymbol{I}_n)$. The square/square-root pooling is also used in [15, 16], and its effectiveness has been proved. Finally, we can obtain an image feature vector $\boldsymbol{z}^0_n \in \mathbb{R}^{(d-r+1)\times(d-r+1)\times\Phi}$ by aligning each pixel's value of all pooling maps as shown in Fig. 1.

2.2. Deterioration Level Estimation Based on DELM

Given a feature vector $\boldsymbol{z}^0_n \in \mathbb{R}^{(d-r+1)\times(d-r+1)\times\Phi}$, $\boldsymbol{z}^0_n$ is input into DELM. DELM consists of one input layer, K hidden layers and one output layer, that is, the number of layers of DELM is $K + 2$ as shown in Fig. 1. The aim of training of the DELM is calculation of a weight matrix between $(k - 1)$-th layer and k-th layer. Specifically, **(I)** in case of k smaller than $K+1$, the weight matrix is calculated by using ELM-Auto Encoder (ELM-AE), which is an unsupervised learning method. **(II)** in case of $k = K + 1$, the weight matrix is calculated by the same manner as ELM [17] which is a supervised learning method.

(I) $k = 1, 2, ..., K$

In DELM, the relationship between the k-th hidden layer's output matrix $\boldsymbol{Z}^k = [\boldsymbol{z}^k_1, \boldsymbol{z}^k_2, ..., \boldsymbol{z}^k_N]^{\top}$ and $(k-1)$-th hidden layer's output matrix $\boldsymbol{Z}^{k-1}$ can be obtained as follows:

$$\boldsymbol{Z}^k = G\Big(\boldsymbol{Z}^{k-1}(\boldsymbol{\beta}^k)^{\top}\Big), \tag{3}$$

where $(\boldsymbol{\beta}^k)^{\top}$ is the weight matrix between k-th and $(k - 1)$-th hidden layers, and G is the activation function. If $k = 1$, $\boldsymbol{Z}^0 = [\boldsymbol{z}^0_1, \boldsymbol{z}^0_2, ..., \boldsymbol{z}^0_N]^{\top}$ consists of the obtained feature vector $\boldsymbol{z}^0_n$ in the subsection **2.1**.

In order to calculate the weight matrix $(\boldsymbol{\beta}^k)^\top$, we generate ELM-AE layer by layer. Although general deep learning methods determine parameters of network by using the back propagation approach which needs high computation costs and a large number of training images, the proposed method determines the weight matrix by using ELM-AE, which is layer-by-layer unsupervised learning. Thus, DELM-LRF does not require a large number of training images. The ELM-AE network model with L^{k-1} input nodes, L^k hidden nodes and L^{k-1} output nodes is shown in the upper-right of Fig. 1. Given an input vector $\boldsymbol{z}_n^{k-1}$, the outputs $\boldsymbol{h}_n^k$ of hidden layer in ELM-AE can be obtained as

$$\boldsymbol{h}_n^k = G(\boldsymbol{W}^k \cdot \boldsymbol{z}_n^{k-1} + \boldsymbol{b}^k), \tag{4}$$

$$\boldsymbol{W}^{k\top} \cdot \boldsymbol{W}^k = \boldsymbol{I}, \tag{5}$$

$$\boldsymbol{b}^{k\top} \cdot \boldsymbol{b}^k = 1, \tag{6}$$

where the activation function G is a sigmoid function. ELM-AE has orthogonal random weight $\boldsymbol{W}^k = [\boldsymbol{w}_1^k, \boldsymbol{w}_2^k, ..., \boldsymbol{w}_{L^k}^k]^\top$ and the random bias $\boldsymbol{b}^k = [b_1^k, b_2^k, ..., b_{L^k}^k]^\top$.

By using ELM-AE's input matrix $\boldsymbol{Z}^{k-1}$ and its hidden layer output matrix $\boldsymbol{H}^k = [\boldsymbol{h}_1^k, \boldsymbol{h}_2^k, ..., \boldsymbol{h}_N^k]^\top$, the calculation of the output weight $\boldsymbol{\beta}^k$ can be divided into the following two patterns.

(I-A) $L^{k-1} \neq L^k$

This means that the number of nodes of an input layer of ELM-AE is not equal to the number of nodes of a hidden layer of ELM-AE (sparse and compressed representation). The output weight $\boldsymbol{\beta}^k$ is obtained as follows:

$$\boldsymbol{\beta}^k = \left(\frac{\boldsymbol{I}}{C}\sum_{l^k=1}^{L^k} \mathrm{KL}(\rho \| \hat{\rho_{l^k}}) + (\boldsymbol{H}^k)^\top \boldsymbol{H}^k\right)^{-1} (\boldsymbol{H}^k)^\top \boldsymbol{Z}^{k-1}, \tag{7}$$

where $\mathrm{KL}(\rho \| \hat{\rho}_{l^k}) = \rho \log \frac{\rho}{\hat{\rho}_{l^k}} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_{l^k}}$ is the KL divergence, ρ is a sparsity parameter ($\rho = 0.05$), and $\hat{\rho}_{l^k}$ is the average activation of each hidden layer's node l^k of ELM-AE.

(I-B) $L^{k-1} = L^k$

This means that the number of nodes of an input layer of ELM-AE is equal to the number of nodes in a hidden layer of ELM-AE. The output weight $\boldsymbol{\beta}^k$ is simply obtained as

$$\boldsymbol{\beta}^k = (\boldsymbol{H}^k)^{-1} \boldsymbol{Z}^{k-1}, \tag{8}$$

where $(\boldsymbol{\beta}^k)^\top \boldsymbol{\beta}^k = \boldsymbol{I}$.

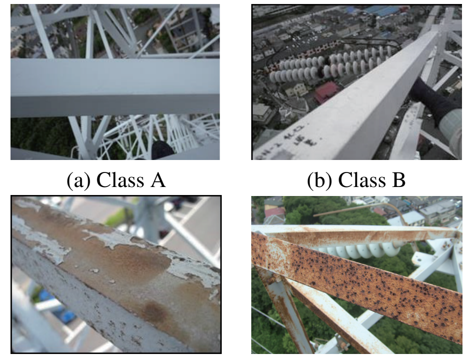
Consequently, we can obtain each layer's output weight $(\boldsymbol{\beta}^k)^\top$ ($k = 1, 2, ..., K$) of DELM by using the output weight $\boldsymbol{\beta}^k$ of the ELM-AE.

(II) $k = K + 1$

The output weight matrix $\boldsymbol{\beta}^{K+1}$ between K-th hidden layer and the output layer is calculated by supervised learning. We try to minimize the training error $\boldsymbol{\xi}_n = [\xi_{n,1}, \xi_{n,2}, ..., \xi_{n,M}]^\top$ (M being the number of output nodes in M-class problem) as well as the output weights.

$$\begin{aligned} \min_{\boldsymbol{\beta}^{K+1} \in \mathbb{R}^{L^{K+1} \times M}} R \quad &= \quad \frac{1}{2}\|\boldsymbol{\beta}^{K+1}\|_F^2 + \frac{C}{2}\sum_{n=1}^{N}\|\boldsymbol{\xi}_n\|^2, \\ \text{s.t. } \boldsymbol{z}_n^{K+1}\boldsymbol{\beta}^{K+1} \quad &= \quad \boldsymbol{t}_n^\top - \boldsymbol{\xi}_n^\top, \end{aligned} \tag{9}$$

where $\boldsymbol{t}_n = [t_{n,1}, t_{n,2}, ..., t_{n,M}]^\top$ is a vector whose m-th element is one, while the rest of elements are zeros if the original true class label is m. Furthermore, C is a regularization parameter. Based on the KKT theorem, Eq. (9) is equivalent to solving the following optimization problem:



(a) Class A (b) Class B

(c) Class B (d) Class C

Fig. 2. Examples of images [18] used in the experiment.

$$\begin{aligned} \min_{\boldsymbol{\beta}^{K+1} \in \mathbb{R}^{L^{K+1} \times M}} \tilde{R} = \frac{1}{2}\|\boldsymbol{\beta}^{K+1}\|_F^2 + \frac{C}{2}\sum_{n=1}^{N}\|\boldsymbol{\xi}_n\|^2 \\ - \sum_{n=1}^{N}\sum_{m=1}^{M} \alpha_{n,m}(\boldsymbol{z}_n^{K+1}\boldsymbol{\gamma}_m^{K+1} - t_{n,m} + \xi_{n,m}), \end{aligned} \tag{10}$$

where $\boldsymbol{\gamma}_m^{K+1} = [\boldsymbol{\beta}_{1,m}^{K+1}, \boldsymbol{\beta}_{2,m}^{K+1}, ..., \boldsymbol{\beta}_{L^{K+1},m}^{K+1}]^\top$ is a vector of the weights linking the hidden layer to the m-th output node. By taking derivatives with $\boldsymbol{\gamma}_m^{K+1}$, $\boldsymbol{\xi}_n$ and $\boldsymbol{\alpha}_n$, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_N]^\top$ and $\boldsymbol{\alpha}_n = [\alpha_{n,1}, \alpha_{n,2}, ..., \alpha_{n,M}]^\top$, the optimal solution of $\boldsymbol{\beta}^{K+1}$ can be obtained as

$$\boldsymbol{\beta}^{K+1} = \left(\frac{\boldsymbol{I}}{C} + (\boldsymbol{Z}^{K+1})^\top \boldsymbol{Z}^{K+1}\right)^{-1} (\boldsymbol{Z}^{K+1})^\top \boldsymbol{T}, \tag{11}$$

where $\boldsymbol{T} = [\boldsymbol{t}_1^\top, \boldsymbol{t}_2^\top, ..., \boldsymbol{t}_N^\top]^\top$. Consequently, we can obtain the weight matrix $\boldsymbol{\beta}^{K+1}$.

Given an input luminance matrix $\boldsymbol{I}$ of test image, by using the obtained $\boldsymbol{\beta}^{K+1}$, the output value $\boldsymbol{y} = [y_1,\ y_2,\ ...,\ y_M]$ is obtained as

$$\boldsymbol{y} = \boldsymbol{z}^{K+1}\left(\frac{\boldsymbol{I}}{C} + (\boldsymbol{Z}^{K+1})^\top \boldsymbol{Z}^{K+1}\right)^{-1} (\boldsymbol{Z}^{K+1})^\top \boldsymbol{T}. \tag{12}$$

Furthermore, the final result *level* is obtained as

$$level = \arg \max_{m \in \{1,...,M\}} y_m. \tag{13}$$

Thus, the deterioration level estimation based on DELM-LRF can be realized. DELM-LRF which integrates LRF and DELM can realize the automatic and accurate estimation of deterioration levels on transmission towers from few training images.

3. EXPERIMENTAL RESULTS

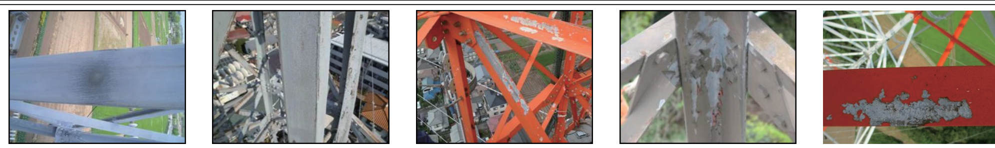
In this section, the effectiveness of our method is verified by using surface images on transmission towers. In **3.1**, the details of experimental settings are explained. The performance of the proposed method is shown in **3.2**.

3.1. Experimental Settings

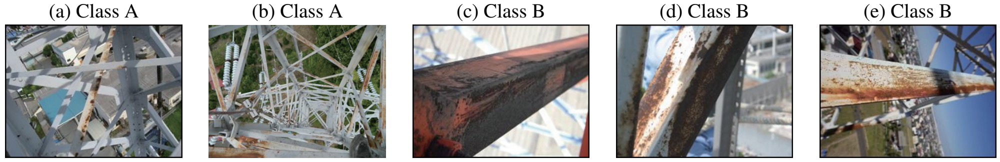
In this experiment, we prepared the dataset which has three classes. Specifically, these classes are corresponding to deterioration levels, and these classes are defined as classes A, B and C. The class C is the most dangerous. We used 2797 images which include 940 images with class A, 1184 images with class B and 673 images with class C. Examples of these images are shown in Fig. 2. The verification method was five-fold cross-validation, and Recall, Precision and F-measure were used for the evaluation.

Table 1. Recall, Precision and F-measure of the proposed method and comparative methods.

Method	Class A			Class B			Class C			Average		
	R	P	F	R	P	F	R	P	F	R	P	F
DELM-LRF (Ours)	**0.987**	0.890	0.933	**0.593**	**0.952**	**0.721**	**0.912**	**0.621**	**0.733**	**0.831**	**0.821**	**0.796**
ELM-LRF [13]	0.986	**0.903**	**0.940**	0.399	0.939	0.471	0.869	0.585	0.620	0.751	0.809	0.677
DELM [14]	0.668	0.624	0.644	0.464	0.611	0.526	0.708	0.536	0.610	0.613	0.591	0.594
KELM [19]	0.669	0.615	0.640	0.399	0.586	0.473	0.722	0.505	0.594	0.596	0.568	0.569
ELM [17]	0.662	0.595	0.627	0.420	0.588	0.490	0.680	0.507	0.580	0.587	0.563	0.566
SVM [20]	0.713	0.559	0.626	0.266	0.564	0.361	0.739	0.480	0.582	0.572	0.534	0.523
CaffeNet-CNN [21]	0.517	0.362	0.426	0.363	0.464	0.407	0.163	0.207	0.182	0.347	0.344	0.338



(a) Class A (b) Class A (c) Class B (d) Class B (e) Class B



(f) Class C (g) Class C (h) Class C (i) Class C (j) Class C

Fig. 3. Examples of images [18] which were correctly estimated by the proposed method.

In order to verify the effectiveness of the proposed method, we compared its estimation performance with six comparative methods shown in Table 1. First, we used ELM-LRF which is a estimation method based on LRF [13]. Next, we used DELM [14], ELM with kernel (KELM) [19], ELM [17] and SVM [20] constructed by using hand-craft features. Specifically, we extracted SIFT-BoF features [22] and HS histogram features [23] from images, and performed a feature selection approach based on the mRMR algorithm [24]. Then we constructed DELM, KELM, ELM and SVM from the obtained features. Furthermore, we compared our method with a CNN-based method which has high performance in various recognition fields in order to verify the effectiveness of the proposed method. Specifically, since it is difficult to prepare a sufficient number of training images on transmission towers, we used the CNN based on fine-tuning as a comparative method. We call the CNN which is fine-tuned by using the Caffe reference network as CaffeNet-CNN [21] in this paper. We determined parameters of each method in such a way that the estimation performance of each method is the best using validation dataset. Note that in the proposed method, $r = 5$, $\Phi = 30$ and $e = 3$ were used in LRF, and the number of hidden layers K was 7 in DELM.

3.2. Performance Evaluation

The experimental results are shown in Table 1. Table 1 shows that the performance of DELM-LRF was higher than that of all comparative methods. Since the performance of KELM, ELM and SVM which are benchmarking methods in image recognition fields were low, these results indicate that deterioration level estimation is difficult and a challenging task. Furthermore, the performance of DELM which is one of the effective deep learning methods was also low. From this point, hand-craft features are not effective for the deterioration level estimation. On the other hand, in DELM-LRF, since we trained features which were calculated by the convolution and pooling, we can extract enough information for the following DELM. Thus, DELM-LRF is superior to other methods, that is, the effectiveness of the integration of LRF and DELM is verified. Moreover, the performance of CaffeNet-CNN is lower. This result indicates that CaffeNet-CNN is not suitable for a small amount of technical data such as images of transmission towers. Furthermore, the computation time in the training step of our method is 3.16×10^2 sec, and that of CaffeNet-CNN is 1.44×10^4 sec [1]. Our method can be trained much faster than CaffeNet-CNN. This is the reason why we focus on the ELM series. In this way, the proposed method makes it possible to approach the challenging tasks with high expertise having difficulty in preparing enough images such as deterioration level estimation.

Figure 3 shows the examples of images which were correctly estimated by the proposed method. From Fig. 3, images of transmission towers have many variations. Specifically, Figs. 3 (c), (f) and (g) are distant-view images, but others are near-view images. The colors of subjects, i.e., transmission towers in Figs. 3 (c), (e) and (h) are not gray or white but red. In addition, the angles of subjects are different such as Figs. 3 (b) and (j). Thus, the proposed method can estimate various kinds of images correctly. In general estimation methods such as [3], in order to cope with such many variations, the target in images are clipped manually. Thus, these are semi-automatic methods. On the other hand, since we can input the original images into our method directly, our method is a full-automatic estimation method. Therefore, the effectiveness of the proposed method is verified in view of the practical application.

4. CONCLUSIONS

In this paper, we have presented a novel estimation method which handles a small amount of technical data. Specifically, we automatically extract image features based on LRF which combines convolution and pooling. Furthermore, we estimate deterioration levels via DELM which is a part of efficient deep learning methods. Thus, the construction of DELM-LRF is the main contribution of this paper. The experimental results have verified the effectiveness of the proposed method. Consequently, the proposed method makes it possible to approach challenging tasks with high expertise having difficulty in preparing enough images.

[1]Note that this experiment was performed on a personal computer using Intel (R) Xeon (R) CPU E5-2699 v3 @ 2.30 GHz with 512 Gbytes RAM and GPU Tesla K80.

5. REFERENCES

[1] T. Yin, H. F. Lam, H. M. Chow, and H. Zhu, "Dynamic reduction-based structural damage detection of transmission tower utilizing ambient vibration data," *Engineering Structures*, vol. 31, no. 9, 2009.

[2] F. Zhang, W. Wang, Y. Zhao, P. Li, Q. Lin, and L. Jiang, "Automatic diagnosis system of transmission line abnormalities and defects based on uav," in *2016 IEEE 4th International Conference on Applied Robotics for the Power Industry (CARPI)*, 2016, pp. 1–5.

[3] F. Tsutsumi, H. Murata, T. Onoda, O. Oguri, and H. Tanaka, "Automatic corrosion estimation using galvanized steel images on power transmission towers," in *2009 IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific*, 2009, pp. 1–4.

[4] O. Steiger, E. Lucas, and Y. Maret, "Automatic detection of transmission towers," in *IEEE SENSORS 2014 Proceedings*, 2014, pp. 1034–1037.

[5] T. Dutta, H. Sharma, A. Vellaiappan, and P. Balamuralidhar, "Image analysis-based automatic detection of transmission towers using aerial imagery," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2015, pp. 641–651.

[6] S. Woo, I. Chu, B. Youn, and K. Kim, "Development of the corrosion deterioration inspection tool for transmission tower members," *KEPCO Journal on Electric Power and Energy*, vol. 2, no. 2, pp. 293–298, 2016.

[7] TEPCO Press Release PC Watch [Smart O & M], http://pc.watch.impress.co.jp/docs/news/1024585.html *(in Japanese)*.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[9] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3708–3712.

[10] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 IEEE 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2014, pp. 844–848.

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition." in *Icml*, vol. 32, 2014, pp. 647–655.

[12] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in neural information processing systems*, 2014, pp. 487–495.

[13] G. B. Huang, Z. Bai, L. L. C. Kasun, and C. M. Vong, "Local receptive fields based extreme learning machine," *IEEE Computational Intelligence Magazine*, vol. 10, no. 2, pp. 18–29, 2015.

[14] E. Cambria, G. B. Huang, L. L. C. Kasun, H. Zhou, C. M. Vong, J. Lin, J. Yin, Z. Cai, Q. Liu *et al.*, "Extreme learning machines [trends & controversies]," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 30–59, 2013.

[15] S. Andrew, K. P. W, C. Zhenghao, B. Maneesh, S. Bipin, and N. A. Y, "On random weights and unsupervised feature learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 1089–1096.

[16] A. Coates, H. Lee, and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," *Ann Arbor*, vol. 15, pp. 215–223, 2010.

[17] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *2004 IEEE International Joint Conference on Neural Networks*, vol. 2, 2004, pp. 985–990.

[18] H. Hamada *et al.*, Tokyo Electronic Power Company Holdings, Inc. (*private communication*).

[19] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[22] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[23] A. R. Smith, "Color gamut transform pairs," *ACM Siggraph Computer Graphics*, vol. 12, no. 3, pp. 12–19, 1978.

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.