15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Transfer learning of a deep convolutional neural network for localizing handwritten slab identification numbers

Sang Jun Lee
POSTECH, Korea
lsj4u0208@postech.ac.kr

Gyogwon Koo
POSTECH, Korea
ggkoo99@postech.ac.kr

Hyeyeon Choi
POSTECH, Korea
hyeyeon@postech.ac.kr

Sang Woo Kim
POSTECH, Korea
swkim@postech.ac.kr

Abstract

Most machine learning methods assume that previous and future data have same distribution in same feature space. This paper presents a real-world problem that violates the common assumption, and we propose a practical methodology to handle the problem. In the steel making industry, automated marking systems are widely used to inscribe slab identification numbers (SINs). In the previous work, a deep learning based algorithm was developed to automatically extract regions of printed SINs. However, as the marking system is outdated, few SINs are marked by hand in uncommon situations, and the existing algorithm does not work for the handwritten SINs. This paper proposes a practical method that uses very small training data (10 images) to localize handwritten SINs. The knowledge of mid-level layers or entire layers in the pre-trained deep convolutional neural network is transferred to overcome the shortage of training data in the target domain. Experiments were conducted with actual industrial data to demonstrate the effectiveness of the proposed algorithm.

1 Introduction

In the steel manufacturing industry, identification of individual products is important for the production management. Automated marking systems with a specialized paint that is endurable to the high temperature of steel products are widely used to inscribe product identification numbers. For the automatic recognition of a product identification number, accurate localization is a challenging problem due to intricate background of actual factory scenes. In our previous work [1], a deep learning based algorithm was developed for localizing slab identification numbers (SINs) that were printed by a paint marking machine. This previous algorithm used a deep convolutional neural network (DCNN), and it achieved better performance compared to a rule-based localization algorithm [2] for printed SINs. However, in an actual steelworks, SINs are marked by hand in few exceptional circumstances, and the previous localization algorithm does not work for the few uncommon situations. Furthermore, the number of handwritten SINs is not sufficient for training a new DCNN. In this paper, we propose a practical methodology that uses transfer learning to handle the actual industrial problem.

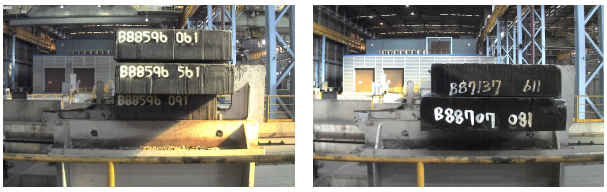
A DCNN, firstly proposed in [3], is a popular deep learning structure for image data. Recently, noticeable performance improvements have been achieved with the use of DCNNs in computer vision tasks such as image classification [4], object detection [5], and other applications [6, 7]. In spite of its outstanding performance, construction of sufficiently big training data is a difficult and time-consuming task in some industrial fields. This issue can be handled by transferring source-domain knowledge that is related to a target task.

Table 1. Outline of transfer learning for the localization of handwritten SINs.

	Source	Target
Domain	Printed SINs	Handwritten SINs
Task	Localization of SINs (classification of sub-regions)	

Transfer learning is the use of knowledge acquired from a source domain for solving a related problem. General frameworks for the transfer learning and keywords such as domain and task are well-summarized in [8]. In many machine vision applications, knowledge transfer from a source domain to a target task have been used to handle the problem of insufficient training data or different distributions of previous and future data [9, 10, 11, 12]. The architecture of a DCNN is suited for transfer learning, and source-domain knowledge can be transferred without the use of source-domain data. There are two major approaches for transferring source-domain knowledge of a DCNN. The first approach is knowledge transfer of mid-level representations [13, 14]. In this approach, low and mid-level parameters of a DCNN are reused, and new adaptation layers or simple machine learning techniques such as a support vector machine are added to solve a target task. Another approach is transferring the entire model of a DCNN. A DCNN that was previously trained with source-domain data was fine-tuned using a small number of target-domain data in [15, 16]. In this paper, both two approaches are employed to solve the actual industrial problem.

Our problem has different domains and same task for source and target as summarized in Table 1. A large number of printed SINs were used as source-domain data, and target-domain data were constructed with few number of handwritten SINs. Fig. 1 presents actual factory scenes in source and target domains. Images in both domains have similar background contents, but the shape of SINs are different. The objective of this paper is to develop a localization algorithm for handwritten SINs using very small number of target-domain data.



(a) Source-domain data. (b) Target-domain data.

Figure 1. Examples of actual factory scenes in source and target domains.

2 Transfer learning of a DCNN

2.1 Source-domain data (printed SINs)

In our previous work, 4934 actual factory scenes that contain 10007 slabs with printed SINs were utilized to generate patch images. From regions of background and SINs, 4.6 million patch images with the fixed size of 20×48 were collected to train a DCNN. Architectures of DCNN were designed to classify sub-regions in a test image for the localization of printed SINs. The DCNN model for the transfer learning has three convolutional layers and two fully-connected layers, and it contains 161,568 trainable parameters.

2.2 Training data (handwritten SINs)

A SIN is marked by hand in uncommon situations, and collecting sufficient number of handwritten SINs is difficult in our industrial application. In this paper, very small number of actual factory scenes (10 images) that contain handwritten SINs are used to construct training data. From the 10 training images, a sliding window method with a vertical 5-pixel/10-pixel or horizontal 12-pixel/24-pixel displacement was used for regions of SINs/background to collect patch images with the size of 20×48. In this procedure, 2223 and 23010 patches were collected from the regions of handwritten SINs and background, respectively.

2.3 Two approaches for transfer learning

The DCNN that was pre-trained in the source domain and two approaches for transferring the knowledge of printed SINs are depicted in Fig. 2. The first approach uses the knowledge of mid-level representations in the pre-trained DCNN, and new adaptation layers are added as shown in Fig. 2(a). The weights of the convolutional layers in the pre-trained DCNN are reused, the adaptation layers are trained with the training data using its activation values of the last convolutional layer. Fig. 2(b) presents the second approach of transfer learning, and the entire model of the pre-trained DCNN model is reused. The transfered DCNN model is fine-tuned using the training data in the target domain.

2.4 Training of handwritten SINs

A DCNN structure is trained in three ways using the training data. The first training method uses only training data, and the second and third methods use the training data with the source-domain knowledge. The performance for the training patches in the target-domain data are presented in Fig. 3. The accuracies

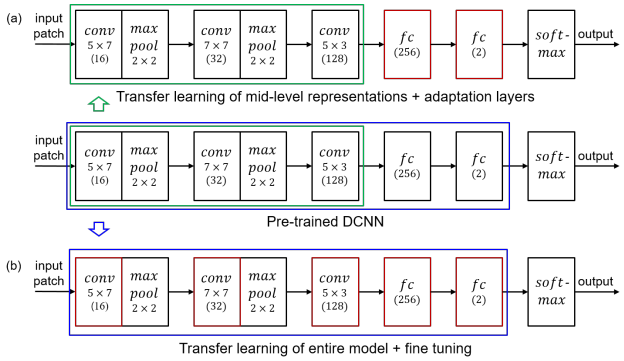


Figure 2. Two knowledge-transferring approaches for localizing handwritten SINs.

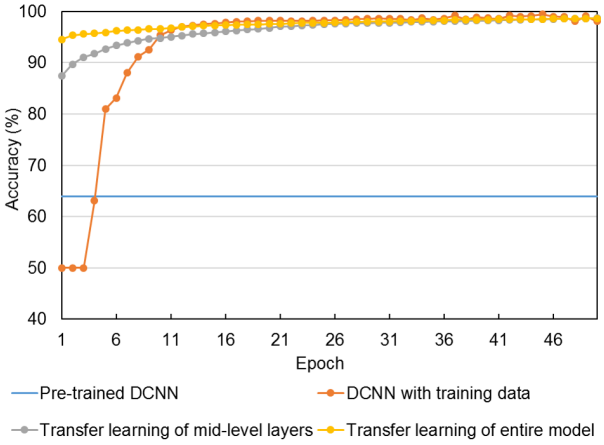


Figure 3. The accuracy for the classification of the training data.

of the pre-trained network, a DCNN trained with only target-domain data, and DCNNs with the two types of transfer learning are compared. The pre-trained DCNN that was trained in the source domain does not work for the localization of handwritten SINs. Although most of background patches were correctly classified by the pre-trained network, many patches that belong to a region of a handwritten SIN were incorrectly classified into a background patch. The performances using the other three training methods were almost saturated after 20 epochs.

3 Localization of SINs

For a test image, a sliding window method is utilized to extract patch images with the size of 20×48. The patch images are classified into a background or SIN region by using a DCNN. A character confidence map is calculated using the probabilities that individual patches in the test image belong to a region of a SIN. Fig. 4 shows a test image and its character confidence map. Bounding boxes for handwritten SINs are obtained using the character confidence map.

4 Experimental results

The proposed algorithm for the localization of handwritten SINs was tested on 387 actual factory scenes

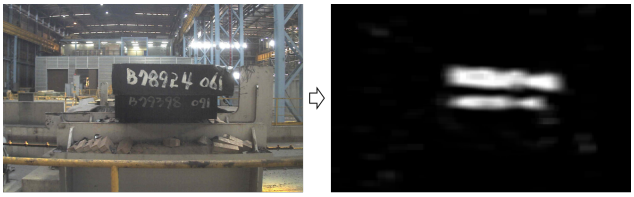


Figure 4. Localization procedure for a test image.

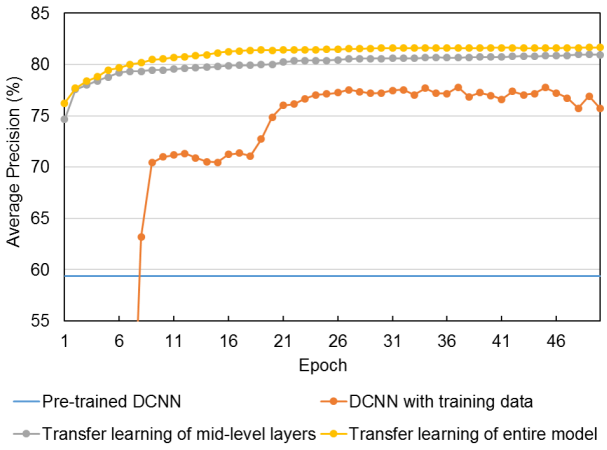


Figure 5. AP for the localiation of SINs.

that contain 683 slabs. In Fig. 5, the localization performances of the two knowledge-transferring approaches were compared to the performances of the pre-trained DCNN and a DCNN that uses only training data. The performance was measured using average precision (AP), and AP is defined as below:

$$AP = \frac{1}{N} \sum_{i} \frac{|P_i \cap T_i|}{|P_i \cup T_i|}, \tag{1}$$

where P_i and T_i are the predicted and true regions of SINs in the i-th image, N is the total number of test images, and $|P_i|$ is the area of the region P_i. Fig. 5 presents APs for the 4 cases of a DCNN during 50 training epochs. The APs of the DCNNs that use the source-domain knowledge were higher than the AP of the DCNN that uses only training data.

AP is a widely used measure in the field of object detection, and it measures the ratio of overlapped area of true and predicted regions. Generally, localization is performed for a following recognition task, and a small cropped region is not critical in most object recognition problems. However, a small cropped region is a critical problem for recognizing a SIN, because one missing or incorrectly recognized character results in failure for the recognition of a SIN. Therefore, sensitivity and precision were measured for comparing actual performances of 4 cases of the DCNNs, and it is summarized in Table 2. Sensitivity is the ratio of the number of correctly localized SINs to the number of true SINs, and precision is the ratio of the number of correct SINs to the number of predictions. A visually recognizable case is regarded as a correct SIN. The error rates in the sense of sensitivity and precision were noticeably reduced by using the source-domain knowledge.

Fig. 6 presents result images for the localization of handwritten SINs, and it compares the results with pre-trained DCNN and the results that use transfer learning of the entire DCNN model with fine-tuning. Incorrectly localized SINs in Fig. 6(a) are due to failure in classification of patches that contain a handwritten SIN. The localization algorithm that uses source-domain knowledge succeeded to the localization of SINs in Fig. 6(a). In Fig. 6(b), the localization results using the pre-trained DCNN are not recognizable due to small cropped regions of the SINs. This results show the significance of accurate localization of SINs.

5 Conclusion

This paper proposes a practical methodology for handling the problem of insufficient training data. In spite of the high performance of deep learning based algorithms, its application on an industrial problem is difficult due to challenges for collecting unusual data, tedious labeling processes, or different distributions of previous and future data. In an actual steelworks, SINs were unusually marked by hand as an automated marking system was outdated. The previous deep learning based algorithm did not work for the localization of handwritten SINs, and the number of handwritten SINs is insufficient for training a new DCNN. To solve the real-world problem, the knowledge of printed SINs was regarded as source domain, and it was transferred in two ways for the localization of handwritten SINs. Experiments show that our methodology that is based on transfer learning is practical and effective for using very small number of training data.

References

[1] S. J. Lee and S. W. Kim, "Localization of the slab information in factory scenes using deep convolutional neural networks," *Expert Syst. Appl.*, vol. 77, pp. 34–43, Jul. 2017.

[2] S. Choi, J. P. Yun, K. Koo, and S. W. Kim, "Localizing slab identification numbers in factory scene images," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7621–7636, Jul. 2012.

[3] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[5] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2553–2561.

[6] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with convolutional neural networks," *Int. J. Comput. Vis.*, vol. 116, no. 1, pp. 1–20, Jan. 2016.

[7] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, E. I. Chang *et al.*, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2014, pp. 1626–1630.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10,

Table 2. Localization performance.

	Pre-trained DCNN	DCNN with training data	Transfer learning of mid-level layers	Transfer learning of entire model
Predicted SINs	657	675	683	685
Correct SINs	192	506	603	633
Sensitivity (%)	28.11	74.08	88.29	**92.68**
Precision (%)	29.22	74.96	88.29	**92.41**

Figure 6. Result images.

pp. 1345–1359, Oct. 2010.

[9] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. Comput. vis. ECCV*. Springer, 2010, pp. 213–226.

[10] 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 *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.

[11] Y. Sawada and K. Kozuka, "Transfer learning method using multi-prediction deep boltzmann machines for a small scale dataset," in *Proc. Int. Conf. Machine Vision Applications (MVA)*. IEEE, 2015, pp. 110–113.

[12] D. Wang and T. F. Zheng, "Transfer learning for speech and language processing," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2015, pp. 1225–1237.

[13] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 1717–1724.

[14] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 806–813.

[15] S. Branson, G. Van Horn, S. Belongie, and P. Perona, "Bird species categorization using pose normalized deep convolutional nets," *arXiv preprint arXiv:1406.2952*, 2014.

[16] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, 2016.