TABLE III: Execution time of discrimination of variables from words in methods

Methods	Execution time (seconds)
Method using orientation of gradient [11]	37.52
Method using discrete wavelet decomposition [12]	6.95
Method using Alexnet and SVM	223.33
Method using ResNet-50 and SVM	1814.65

angle of characters. The method using 2-D discrete wavelet decomposition in [12] focuses on the stroke analysis of characters. Therefore, the existing methods are not efficient for short words in different fonts. The obtained classification accuracy of ResNet-50 is higher than that of Alexnet because there are more layers in the architecture of the network.

Comparing to the fine-tuning of CNNs, the combination of CNNs and SVM shows better performance. In this case, fine-tuning the pre-trained CNNs is used for the new task of classification of variable and word. Due to the small number of variable and word images, the use of fine-tuning is not as efficient as the combination of CNNs and SVM.

In reality, the feature extraction depends on the clearness of testing datasets. In some circumstances, there are some noises generated from the segmentation of images. It can cause the misclassification (e.g some short words such as *"in"* or *"of"* can be detected with errors as variables). The variable containing a single character has caused higher error rate in the classification than the variable containing an index. Actually, there are more variances of positions in the layout of the variable containing an index than that of the variable containing a single character. The feature allows to discriminate the type of variable from textual word more accurately.

C. Time efficiency

In order to compare the execution performance of the methods, the execution time of the classification of testing variable and textual word is evaluated. The methods are implemented in Matlab R2018a environment and run on a computer with 6GB RAM and Core i3-2.67 GHz processor. The execution time of the discrimination of variables from words in testing phase is shown in the table III. Actually, CNNs consumes more time than those of traditional approaches in the feature extraction because the CNNs aim to automatically extract more visual features than traditional methods in order to improve the classification accuracy. The ResNet-50 consists of more layers than Alexnet, therefore, it takes more time for the CNN to extract features of images. The mathematical expression detection is a key step of building mathematical retrieval systems. For the systems, the prerequisite step is executed offline. It is not required to perform the step in real time and the accuracy in the detection is important to achieve.

V. Conclusion and future work

We have presented the method for detecting variables in inline expressions in document images. In the work, the employment and optimization of CNNs including Alexnet and Resnet-50 allow to obtain the highest precision of 99.5% for the classification of variable and word images. The uses of pre-trained CNNs as feature extractors and the SVM as a classifier out-perform traditional methods in the classification. Besides, pre-trained CNNs allow to extract discriminant features of image with an appropriate consumption of time and resource. In the future, context information of variable can be investigated and integrated to improve the accuracy for the detection. Moreover, different machine learning algorithms can be used in order to determine the optimized choice for the classification of variable and word.

References

[1] "mathinsight," https://mathinsight.org/definition/variable, accessed: 2018-03-14.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, 2012.

[3] "Tesseract," https://github.com/tesseract-ocr, accessed: 2018-03-14.

[4] "Ocropus," https://github.com/tmbdev/ocropy, accessed: 2018-03-14.

[5] A. M. Namboodiri and A. K. Jain, "Document structure and layout analysis," *Digital Document Processing*, pp. 29–48, 2007.

[6] X.Lin *et al.*, "A text line detection method for mathematical formula recognition," in *The 12th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2013.

[7] H.Lee and J.Wang, "Design of a mathematical expression understanding system," *Pattern Recognition Letters*, vol. 18, 1997.

[8] X. J.Jin and Q.Wang, "Mathematical formulas detection," in *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*. IEEE, 2003.

[9] U. Garain, "Identification of mathematical expressions in document images," in *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*. IEEE, 2009, pp. 1340–1344.

[10] W.Chu and F.Liu, "Mathematical formula detection in heterogeneous document images," in *Proceedings of Conference on Technologies and Applications of Artificial Intelligence*. IEEE, 2013.

[11] C. Sun and D. Si, "Skew and slant correction for document images using gradient direction," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1997.

[12] L.Zhang, Y. Lu, and C.Tan, "Italic font recognition using stroke pattern analysis on wavelet decomposed word images," in *Proceedings of the 17th International Conference on Pattern Recognition*. IEEE, 2004.

[13] W. He, Y. Luo, F. Yin, H. Hu, J. Han, E. Ding, and C.-L. Liu, "Context-aware mathematical expression recognition: An end-to-end framework and a benchmark," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, 2016, pp. 3246–3251.

[14] B. H. Phong, T. M. Hoang, and T.-L. Le, "A new method for displayed mathematical exressions detection based on fft and svm," in *Nafosted conference on Information and Computer Science*. IEEE, 2017, pp. 90–96.

[15] E. Kavallieratou *et al.*, "Handwritten character recognition based on structural characteristics," in *2002. Proceedings. 16th International Conference on Pattern Recognition*. IEEE, 2002.

[16] "ImageNet," ImageNet. http://www.image-net.org/, accessed: 2018-03-14.

[17] M. Afzal and S. Capobianco, "Deepdocclassifier: Document classification with deep convolutional neural network," in *13th International Conference on Document Analysis and Recognition (ICDAR) (2015)*. IEEE, 2015.

[18] X.Lin and L.Gao, "Performance evaluation of mathematical formula identification," in *2012 10th IAPR International Workshop on Document Analysis Systems*. IEEE, 2012.

[19] I. Jolliffe, "Principal component analysis," *2nd ed., Springer, 2002.*, 2002.