TABLE I: Number of word and variable images in the testing datasets

Number of word images	Number of variable images	
	Variable containing a single character	Variable containing a single character and an index
420	308	72

TABLE II: Classification accuracy results (Bold value indicates the highest scores of the methods)

Methods	Precision	Recall	F-score
Method using orientation of gradient [11]	86.38%	76.81%	81.31%
Method using DWT [12]	92.63%	83.45%	87.80%
Method using the fine-tuning of Alexnet	93.38%	86.58%	89.85%
Method using the fine-tuning of ResNet-50	94.13%	88.25%	91.09%
Method using Alexnet and SVM	98.13%	96.11%	97.11%
Method using ResNet-50 and SVM	**99.5%**	**98.95%**	**99.23%**

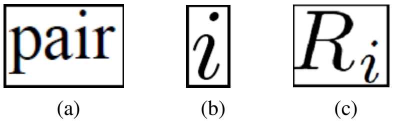

Fig. 4: Examples of a word (a); a variable contains a single character (b) and a variable contains an index (c)

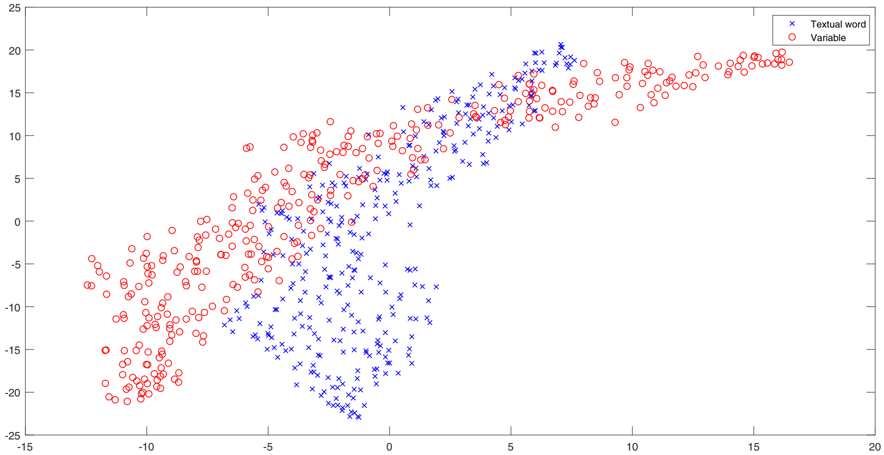

Fig. 5: Representation of feature of variable and word extracted by Resnet-50

B. *Performance evaluation*

1) Baseline and performance measures: As a baseline, the existing methods in [11] and [12] are used for the classification of variable and word. Moreover, fine-tuning CNNs is used for evaluating the performance of the classification. Fine-tuning a network is the reuse of a tuned network for a new task of classification. The technique is adopted because the number of variable and word images is not enough to train the CNNs from scratch. By using the technique, the variable and word images are classified by the *Softmax* classifier that is the last layer of the CNNs.

In our work, the Precision (P), Recall (R) and F1 score are used for the performance evaluation. Precision is the proportion of the true positives against all the positive results; Recall is the proportion of the true positives against all the true results and F1 score is the harmonic mean of precision and recall.

2) Performance: The performance comparisons of the proposed method and existing ones are shown in the table II. Comparing to traditional methods, the uses of CNNs and SVM in our work show higher accuracy. The accuracy in the classification of variable and textual word much improves. The out-performance comes from the fact that the CNNs allow to extract features of images better than existing methods. Fig. 5 illustrates the features of 340 images of each type of variable and word that are extracted by Resnet-50. The feature distribution is represented by using the dimensional reduction technique that is Principal Component Analysis (PCA) [19]. Actually, 1000 features of variable and word images are extracted by using Resnet-50. The PCA technique allows to represent the features in 2-D space efficiently. The feature representation illustrates the classification possibility of variable and textual word. The method using orientation of gradient in [11] heavily relies on the calculation of skew