International Journal of Web-Based Learning and Teaching Technologies
Volume 17 • Issue 5

Handwritten Documents Validation Using Pattern Recognition and Transfer Learning

Jadli Aissam, University of Hassan II, Casablanca, Morocco

Mustapha Hain, ENSAM, University of Hassan II, Casablanca, Morocco

Adil Chergui, ENSAM, University of Hassan II, Casablanca, Morocco

ABSTRACT

Handwritten documents in an enterprise resource planning (ERP) system can come from different sources and usually have different designs, sizes, and subjects (i.e., bills, checks, invoices, etc.). Given these documents were filled manually, they have to be inspected to detect various kinds of issues (missing signature or stamp, missing name, etc.) before being saved in the ERP system or processed by an OCR engine. In this paper, the authors present a transfer learning approach to detect issues in scanned handwritten documents, using an award-winning deep convolutional neural network (InceptionV3) and different machine learning algorithms such as logistic regression (LR), support vector machine (SVM), and naive Bayes (NB). The experiment shows that the combination of InceptionV3 and LR got an accuracy of 91.8% for missing stamp detection. This can allow using this approach in an ERP system as an automatic verification procedure in a document processing flow.

KEYWORDS

Deep Feature Selection, Document Classification, ERP, InceptionV3, Knowledge Transfer, Logistic Regression, Pattern Recognition, Transfer Learning

INTRODUCTION

Large Companies usually receive a significant volume of documents regarding different internal and external processes (Bostrom et al., 2016) and activities. These documents are issued by multiple organizations and have different sizes, designs, and forms (Bast & Korzen, 2017). Typically, these documents exist on electronic and paper forms; they can also be machine-generated or handwritten (Vani et al., 2014). The principal technique used to collect valuables information from these documents is called Optical Character Recognition (OCR). In this image recognition technique, computers recognize handwritten or machine-written characters with different approaches using deep learning-based techniques (Kotu & Deshpande, 2019). The information extracted using OCR can speed up greatly indexing and improve search speed drastically, which are essential features for the accessibility to these documents (Hamdi et al., 2019).

Before performing OCR, handwritten documents need a previous inspection for irregularities such as missing signature or stamp, empty customer/supplier name, etc. Because having these problems in business documents can bring severe consequences for the document processing flow and may induce a time loss and thus have to be avoided. This verification can be a tedious and time-consuming task, especially for many documents observed in large organizations with essential resources and processes.

This paper is structured as follows: the second section presents the paper context, defines the problem, and explores related works; the third section clears up the knowledge transfer concept;

DOI: 10.4018/IJWLTT.20220901.oa1

while the fourth section proposes the overall framework of the suggested approach and presents used techniques. The fifth section details the conducted experiment and discusses obtained results. In conclusion, the authors review the system features and present research perspectives.

PAPER CONTEXT

In this section, the authors present the basic concepts of this paper. Namely, defining ERP and its ecosystem (composed of different modules and systems), determining the boundaries of addressed problematic, and finally exploring related works to assess similar concepts and approaches that use deep learning and pattern recognition.

Enterprise Resources Planning (ERP)

In front of fierce competition for survival and expansion, companies shared the same ambitions of optimizing every step of their network of business processes by combining all activities, processes, and strategies into a single central computerized system called Enterprise Resource Planning (ERP). It includes all the business functions like planning, purchasing, capacity planning, production, quality control, inspection, maintenance, selling and distribution, finance, procurement, marketing, etc. The organization's entire data is stored in a unique database known as the manufacturing or central database (Addo-Tenkorang & Helo, 2011). ERP systems are built upon the concept of modules. Therefore, In addition to the Client Relationship Management (CRM) module which mission is to develop relationships with the organization's customers and the Supplier Relationship Management (SRM) module which support the company-supplier relationship during its entire lifecycle (Lang et al., 2002) (Huiping, 2009), different modules in the ERP ecosystem exist to address different needs for the organization, (e.g. MRP, WMS, etc.). Additionally, other functional modules can be implemented in ERP systems (e.g. document management module) to support this architecture.

Problematic

ERP implementations typically use a document management module to handle storing, indexing and retrieving scanned documents. These documents can come from various sources, have different layouts, and concern diverse transactions (purchase, delivery, etc.). Handwritten documents are a significant part of these documents and usually need additional verification for issues and missing data (missing signature or stamp, empty name, etc.). This step is usually done manually and can be a tedious and time-consuming task if the organization has many interactions with its surrounding actors. Using deep learning techniques to detect issues in the scanned handwritten documents automatically can speed up these documents' processing while maintaining decent accuracy.

Related Works

In image classification tasks, dataset size can be a decisive factor in the chosen model's performance because image classification usually requires a large volume of training data to output satisfactory results. Due to the limited availability of human-labeled training data, training data cannot be guaranteed to be sufficient. In practice, related data in a different domain can be included alongside the data from the target domain to expand the availability of our prior knowledge about the target future data (Shao et al., 2015). This is done by extracting useful information from data in a related domain and transferring them to target tasks. For example, deep learning models pre-trained on the ImageNet (Deng et al., 2010) dataset (containing 1.2M images of 1000 classes) are usually used for image classification experiences. (Yang et al., 2016) proposed a technique to determine whether a given source domain effectively transfers knowledge to a target domain and determines how much of that knowledge can be shared.

(Das et al., 2018) proposed a document structure analysis technique used for the rapid training of deep learning models for image segments. Finally, a stacked generalization-based assembling is utilized for combining the predictions of the base deep neural network models.

(Zhang, 2011) proposed a transfer learning approach that can improve classification accuracy by up to more than 10%, even when the connection between the auxiliary and the target tasks is not apparent.

In the context of document validation and object detection, (Dey et al., 2016) achieved an accuracy of 84.57% precision on stamp detection on scanned document images using an outlier detection and classification technique.

Several researched object detection techniques (stamp, signature, etc.) use color and geometric features. For instance, Micenkova et al. (Micenková & Beusekom, 2011) presented a method for stamp detection and differentiation from other color objects in the document using color clustering and geometrical features. The technique is limited to non-black colored stamps of many shapes and forms. Ahmed et al. (Ahmed et al., 2013) proposed an approach based on geometric features and key point descriptors that need a training to detect stamps in black and white and colored forms. However, it reported low precision and recall, as it could not handle severely overlapped stamps.

(Nandedkar et al., 2016) presented a novel spectral filtering based deep learning algorithm (SFDL) for detecting logos and stamps in a scanned document image. This paper presents a novel spectral filtering based deep learning algorithm (SFDL). It exploits that graphical regions such as stamps and logos have distinctive spectral characteristics and use it to identify symbols and seals in a color document. As it is based on deep CNN architecture, it requires a large training dataset for training.

TRANSFER LEARNING

Overview

In machine learning problems, the lack of labeled data can make supervised learning algorithms fail to build accurate classification models. Transfer learning has been developed to deal with such a lack of label problem (Pan & Yang, 2010). It aims to improve learning performance by transferring knowledge from several source domains to a target domain (Zhuang et al., 2019). Figure 1 presents a comparison between classical machine learning and knowledge transfer approaches.

For example, image classification can be modeled as a target learning task where there are only a few labeled training images even if the classification task is used in specialized areas (e.g. health, education, etc.).

Classifier-Based Knowledge Transfer

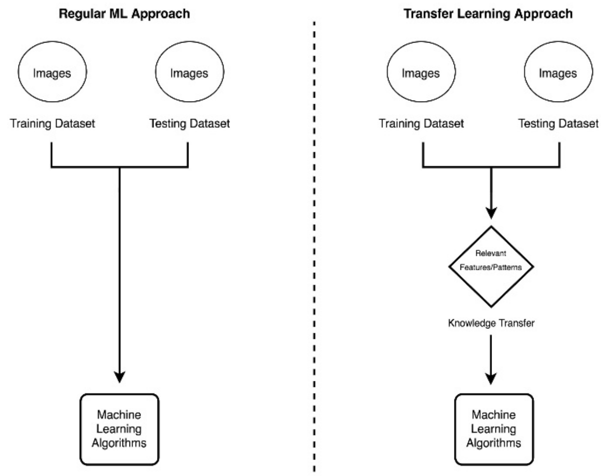
Classifier-based knowledge transfer is a significant part of existing visual transfer learning techniques, and it has attracted much attention in recent years. Instead of minimizing the cross-domain dissimilarity by updating instances representations, classifier based knowledge transfer methods aim to learn a new model that minimizes the generalization error in the target domain via provided training instances from both disciplines and the learned model (Shao et al., 2015).

PROPOSED FRAMEWORK

Techniques

The extraction of useful general features using the InceptionV3 model gives a large features vector that can be passed to multiple machine learning algorithms for further processing to shift the classification targets toward our specific needs. In this section, the authors present the techniques used in this framework.

Figure 1.



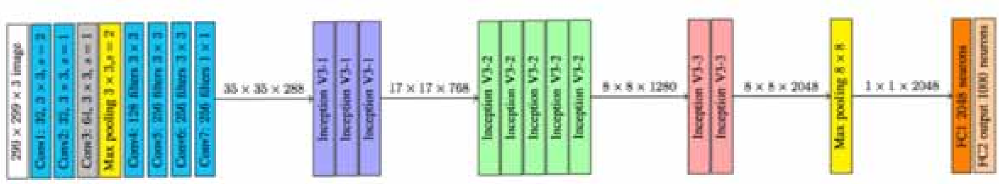
- **InceptionV3**: Convolutional neural networks have shown superior performance than other classifiers in image classification tasks (Wang et al., 2016). Even though they are not invariant to rotation and geometric distortions (Lecun et al., 2015), pre-trained CNN models trained on huge images dataset (i.e. ImageNet) can extract a deep feature vector invariant to rotation and form changes (Tarawneh et al., 2018). Inception V3 (Szegedy et al., 2016) is a pre-trained deep learning model for image classification into 1000 classes. This model was trained on the ImageNet dataset and has achieved an error rate of 3.5% and become the 1st Runner Up for image classification in ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2015. Figure 2 presents Inception V3 architecture.

In our case, the process will stop just before the last layer and recover a vector of 2048 features per image, which will be processed in the next step by chosen machine learning classifiers.

The InceptionV3 was chosen for this experiment because it is a free to use, publicly available deep model boasting a decent performance with significant training efficiency.

- **Naive Bayes**: NB is one of the simplest yet effective multi-class classification methods based on the Bayesian Rule. Given a document instance to be classified, represented by a vector x = (x1,...,xn) representing some n features (independent variables), it assigns to this instance probabilities p(Ck | x1,...,xn), for each of K possible outcomes or classes Ck. Using Bayes' theorem, the conditional probability can be decomposed as (1):

Figure 2. Outline of InceptionV3 Architecture (Ponti et al., 2017)



$$p(C_k \mid x) = \frac{p(x \mid C_k) \cdot p(C_k)}{p(x)} \tag{1}$$

The naive Bayes classifier combines the naive Bayes probability model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posteriori or MAP decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label y^ = Ck for some k as follows (2):

$$\hat{y} = argmax_{k\epsilon\{1,\ldots,K\}} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k) \tag{2}$$

- **Support Vector Machine**: SVM is a non-probabilistic classification method introduced by V. Vapnik, which constructs a hyper-plane in a high-dimensional feature space by empirical risk minimization. Because SVM is a binary classification algorithm, a traditional way of performing multi-class classification is by combining several binary "one-against-all" or "one-against-one" SVM classifiers (Sun et al., 2016).

  We have k (by the number of classes) similar "one-against-all" optimization tasks (3):

$$\begin{gathered} \min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{I} \xi_i \\ w^T \phi(x_i) + b \geq 1 - \xi_i, \quad if\ y_i = m, \\ w^T \phi(x_i) + b \leq -1 - \xi_i, \quad if\ y_i \neq m, \\ \xi_i \geq 0 \end{gathered} \tag{3}$$

  where yi– class of xi. φ– kernel function. C– penalty parameter. Thus the class of a document is derived by (4)

$$c_{SVM} = argmax_{c\epsilon C} ((w^c)^T \phi(x) + b^c) \tag{4}$$

- **Logistic Regression**: LR is a widely used classification algorithm in the industry. Compared with simple algorithms such as decision tree and naive Bayes classification, logistic regression has higher accuracy (Kang et al., 2018). And compared with the algorithms which have higher classification accuracy, such as support vector machine and neural network, the training speed of logistic regression is faster (Salazar et al., 2012). Because of its simplicity and efficiency, logistic regression still attracts wide attention from researchers. Many researchers believe that in many competitions related to big data, some algorithms with better classification effects have the bottleneck of training speed, so logical regression is still one of the most efficient and comprehensive evaluation algorithms. The logistic regression of multi-objective classification uses the formula (5) to calculate the probability of sample x_i belonging to category C_i.

$$p(C_i \mid x) = \frac{e^{w_i^T x + w_{0i}}}{\sum_{j=1}^{K} e^{w_i^T x + w_{0j}}}, \quad i = 1, \ldots, K \tag{5}$$

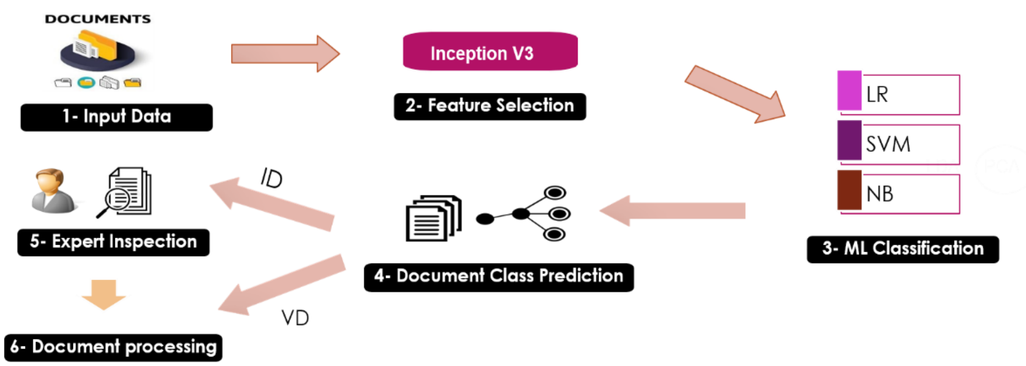
The weight matrix and the bias vector are the parameters ofthis model. These parameters can be obtained by minimizing a loss function. This function can be defined as (6):

$$l(\theta = w, w_0, D) = -\sum_{i=1}^{|D|} \log p(y^{(i)} \mid x^{(i)}) \tag{6}$$

Framework Overview

A general framework is suggested in this paper for issues detection in handwritten documents, in which a transfer learning approach is used for pattern recognition. Figure 3 presents the proposed architecture.

Figure 3. Proposed Framework Architecture



The proposed method is based on using a pre-trained deep convolutional neural network (e.g. InceptionV3) to extract useful features from scanned documents. This step is followed by a classification task using multiple machine learning algorithms like Logistic Regression, Support Vector Machine, and Naive Bayes classifier. Scanned documents can be classified using the extracted relevant patterns into two (2) classes based on if a stamp/signature specific features vector is detected:

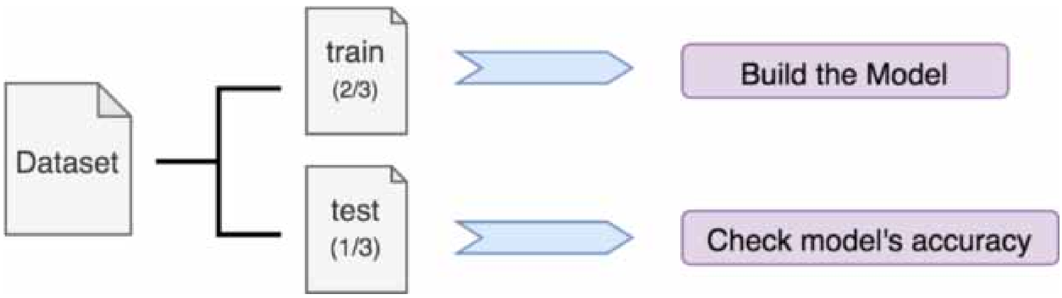
- Valid Documents (VD)
- Invalid Documents (ID)

If a document is labeled as valid (VD), it is forwarded for eventual processing in the system (OCR, indexing, etc.); otherwise a further inspection by a human expert to validate the document.

Model Construction

Using the proposed architecture, we can test different combinations for an optimal transfer learning model. The extracted features using InceptionV3 are forwarded to the NB, SVM, and LR for multi-class classification. In machine learning, the algorithm model needs to be trained to update each parameter in the model. Therefore, it is necessary to provide a training sample. Simultaneously, to validate the constructed model on unseen records, a test set is also needed (Figure 4).

Figure 4. Overview of model construction

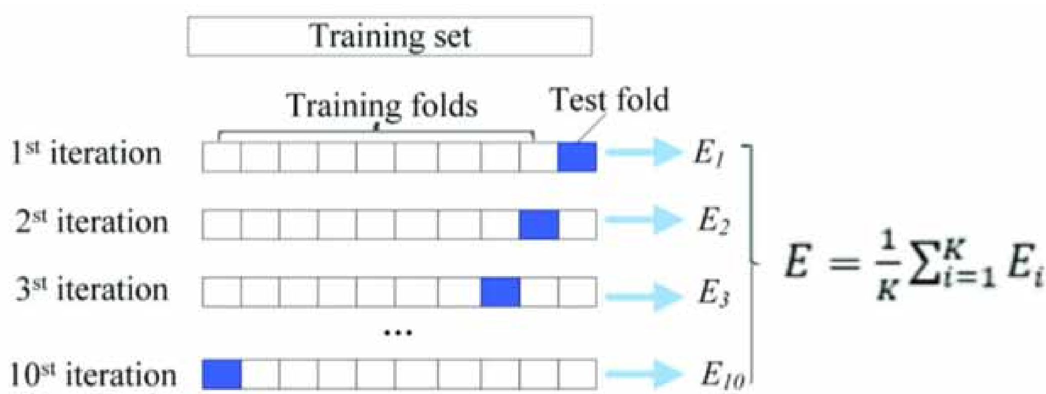


The partitioning ratio is an essential factor, but we must also ensure that the training dataset should include all possible patterns used to define the problem and extend to the modeling domain's edge.

Model Validation

To assess the generalization ability of the model, a cross-validation procedure is usually executed against the chosen model. By varying the number of folds (k), we can overcome the dataset's unbalancing and test our conclusions on every part of the dataset. Figure 5 shows a ten-fold cross-validation operation diagram.

Figure 5. K-fold cross-validation diagram (k=10)



The dataset is divided into K parts (usually 10), and then one part is taken as the test set, the other parts are considered as the training set. The process is repeated until all of the dataset has been processed and tested.

Dataset

The authors collected the data used in this paper from the archive of a sponsoring company. The data is real-world documents issued by various organizations and covering different transactions. Samples of documents were selected and labeled manually into two (2) categories:

- Valid Document (VD): documents containing a stamp and/or signature.
- Invalid Documents (ID): documents not containing stamp and/or signature.

Dataset's class distribution is summarized in Table 1.

Table 1. Classes of documents and the number of samples

Document Class	Sample Count
VD	68
ID	52
Total	120

Figure 6. Sample documents from the dataset

To generalize the results of our experiments on different situations, the images were selected from various sizes, layouts, orientation, and from different sources. Figure 6 presents samples of different classes.

This dataset is considered as small in CNNs standards (usually CNNs need thousands of images if the training is started from scratch), but the use of pre-trained deep models reduces the need for large datasets while maintaining high performances due to transfer learning.

Results and Discussion

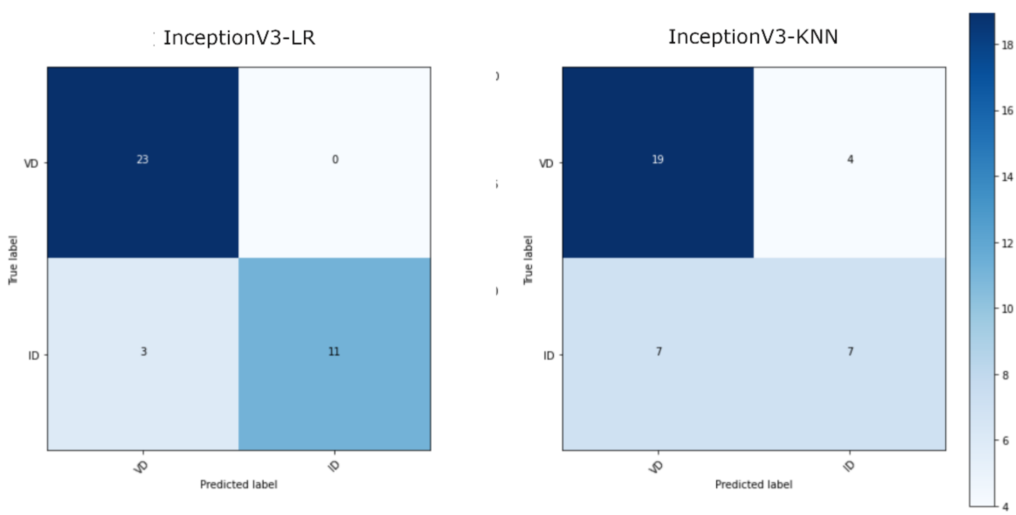
The results of the conducted experiment are shown in Table 2.

Based on overall results, it is clear that transfer learning in a document classification can be a practical approach because all models performed well (accuracy over 60%) without any optimization routine or special hyper-parameters tuning. Furthermore, LR based models got an excellent performance (91.8%), surpassing the classifiers in this task used in the last step. Figure 7 presents the confusion matrix of these top 2 models.

Table 2. Experimental Results

Model	Accuracy %	Precision		Recall	
		VD	VI	VD	VI
InceptionV3-LR	91.8	0.89	0.99	0.98	0.80
InceptionV3-KNN	70.2	0.73	0.64	0.83	0.50
InceptionV3-NB	62.1	0.66	0.50	0.83	0.30

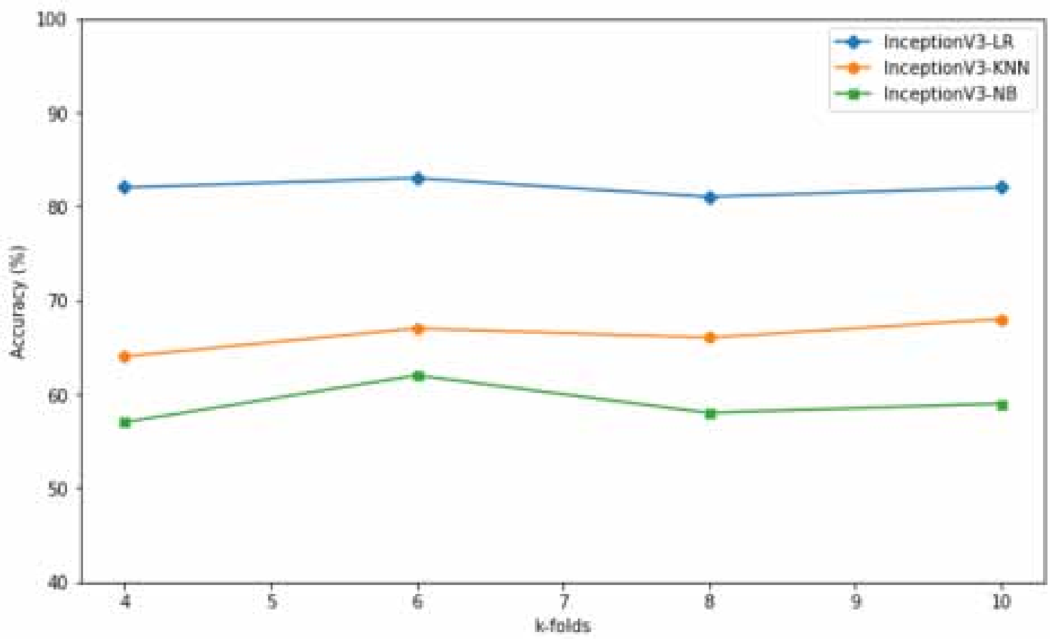
Figure 7.



The second experiment was about executing a cross-validation procedure for every proposed model. Cross-validation techniques are used to assess the generalization ability of the model, in other words, its capacity to classify an unseen document correctly. By varying the number of folds (k), the effects of the dataset's unbalancing can be overcome, and the hypothesis can be tested on every part of the dataset. Figure 8 shows the results of the cross-validation routine on constructed models.

Cross-validation results confirmed the superior performance of the combination of InceptionV3 and logistic regression for stamp and signature recognition on scanned document images. The results

Figure 8. Cross-validation results



showed similar performances to the first experiment for the other models (InceptionV3-KNN and InceptionV3-NB).

Deep pre-trained models in computer vision and pattern recognition are a popular approach for object detection on images using awards-winning deep convolutional neural networks. In this paper, the authors explored the validation of handwritten documents in an ERP system environment using transfer learning; this validation is based on detecting the stamp and signature pattern on the scanned document image. For instance, InceptionV3 deep model performed very well when combined with logistic regression with an accuracy of 91.8% while other models also got decent performances (over 60%).

Such performances using a small dataset size (120 samples) are interesting because CNNs are generally known for their needs of massive training data. Their performances depend on the quality of the dataset. Still, the deep model has prior knowledge from training on the ImageNet dataset, which can come in handy even in a different context.

The K-fold Cross-validation routine was executed on the dataset using different k factors (4, 6, 8, and 10), and the results correlated with the classification results (which can be considered a cross-val with k=3).

CONCLUSION AND PERSPECTIVES

This paper introduces a framework for automatic validation of handwritten documents using knowledge transfer. It describes the main steps of feature selection using a pre-trained deep learning model, applying machine learning algorithms, validating constructed model, and combining this in a simple architecture, providing the system to evolve. Transfer learning reduces the dataset's size needed to implement such a strategy while keeping a high-level accuracy on detecting issues on scanned documents. Documents can be classified using a constructed model on different kinds of problems depending on recognized patterns.

The model constructed using InceptionV3 and Logistic regression got an accuracy of 91.8% without special tuning or parameters optimization; the model's cross-validation further confirmed this result on the dataset at different fold sizes.

In our future work, we plan to gather a bigger dataset containing various issues to perform a more versatile experiment and explore new feature selection techniques (e.g. VGG19, DenseNet, etc.) and more specialized machine learning algorithms. Other approaches are also in consideration to improve dataset variance, such as data augmentation or features engineering.

REFERENCES

Addo-Tenkorang, R., & Helo, P. (2011). Enterprise Resource Planning (ERP): A Review Literature Report. *Lecture Notes in Engineering and Computer Science, 2194*(1), 1126–1134. https://www.uwasa.fi/tuotanto/english/staff/

Ahmed, S., Shafait, F., Liwicki, M., & Dengel, A. (2013). A generic method for stamp segmentation using part-based features. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 708–712. doi:10.1109/ICDAR.2013.145

Bast, H., & Korzen, C. (2017, July 25). A Benchmark and Evaluation for Text Extraction from PDF. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. doi:10.1109/JCDL.2017.7991564

Bostr¨om, C., Herelius, J., Hugosson, M., & Maleev, S. (2016). *Automatic reading and interpretation of paper invoices ADC invoice.* www.it.uu.se

Das, A., Roy, S., Bhattacharya, U., & Parui, S. K. (2018). Document Image Classification with Intra-Domain Transfer Learning and Stacked Generalization of Deep Convolutional Neural Networks. *Proceedings - International Conference on Pattern Recognition,* 3180–3185. doi:10.1109/ICPR.2018.8545630

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F-F. (2010). *ImageNet: A large-scale hierarchical image database*. 10.1109/CVPR.2009.5206848

Dey, S., Mukherjee, J., & Sural, S. (2016, June 10). Stamp and logo detection from document images by finding outliers. *2015 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2015*. doi:10.1109/NCVPRIPG.2015.7489947

Hamdi, A., Jean-Caurant, A., Sidere, N., Coustaty, M., & Doucet, A. (2019). An analysis of the performance of named entity recognition over OCRed documents. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries,* 333–334. doi:10.1109/JCDL.2019.00057

Huiping, C. (2009). An integration framework of ERP, SCM, CRM. *Proceedings - International Conference on Management and Service Science, MASS 2009*. doi:10.1109/ICMSS.2009.5301692

Kang, K., Gao, F., & Feng, J. (2018). A new multi-layer classification method based on logistic regression. *13th International Conference on Computer Science and Education, ICCSE 2018*, 52–55. doi:10.1109/ICCSE.2018.8468725

Kotu, V., & Deshpande, B. (2019). Lecture 04: Training Neural Networks. In Data Science (pp. 307–342). Elsevier. doi:10.1016/B978-0-12-814761-0.00010-1

Lang, A., Paravicini, D., & Revaz, E. (2002). From Customer Relationship Management (CRM) to Supplier Relationship Management (SRM). *Management*, *81*. Advance online publication. doi:10.1088/0004-637X/731/1/34

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In Nature (Vol. 521, Issue 7553, pp. 436–444). doi:10.1038/nature14539

Micenková, B., & Van Beusekom, J. (2011). Stamp detection in color document images. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1125–1129. doi:10.1109/ICDAR.2011.227

Nandedkar, A. V., Mukherjee, J., & Sural, S. (2016, June 10). A spectral filtering based deep learning for detection of logo and stamp. *2015 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2015*. doi:10.1109/NCVPRIPG.2015.7490053

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. In IEEE Transactions on Knowledge and Data Engineering (Vol. 22, Issue 10, pp. 1345–1359). doi:10.1109/TKDE.2009.191

Ponti, M. A., Ribeiro, L. S. F., Nazare, T. S., Bui, T., & Collomosse, J. (2017). Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask. *Proceedings - 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials SIBGRAPI-T 2017,* 17–41. doi:10.1109/SIBGRAPI-T.2017.12

Salazar, D. A., Salazar, J. C., & Vélez, J. I. (2012). Comparison between SVM and logistic regression: Which one is better to discriminate? *Revista Colombiana de Estadistica, 35*(2), 223–237.

Shao, L., Zhu, F., & Li, X. (2015). Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(5), 1019–1034. doi:10.1109/TNNLS.2014.2330900 PMID:25014970

Sun, X., Liu, L., Wang, H., Song, W., & Lu, J. (2016). Image classification via support vector machine. *Proceedings of 2015 4th International Conference on Computer Science and Network Technology, ICCSNT 2015*, 485–489. doi:10.1109/ICCSNT.2015.7490795

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2818–2826. doi:10.1109/CVPR.2016.308

Tarawneh, A. S., Celik, C., Hassanat, A. B., & Chetverikov, D. (2018). *Detailed Investigation of Deep Features with Sparse Representation and Dimensionality Reduction in CBIR: A Comparative Study*. 10.3233/IDA-180895

Vani, B., Beaulah, M. S., & Deepalakshmi, R. (2014). High accuracy optical character recognition algorithms using learning array of ANN. *2014 International Conference on Circuits, Power and Computing Technologies, ICCPCT 2014*, 1474–1479. doi:10.1109/ICCPCT.2014.7054772

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). CNN-RNN: A Unified Framework for Multi-label Image Classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2285–2294. doi:10.1109/CVPR.2016.251

Yang, L., Jing, L., Yu, J., & Ng, M. K. (2016). Learning Transferred Weights from Co-Occurrence Data for Heterogeneous Transfer Learning. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(11), 2187–2200. doi:10.1109/TNNLS.2015.2472457 PMID:26353384

Zhang, J. (2011). Deep transfer learning via restricted Boltzmann machine for document classification. *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011, 1*, 323–326. doi:10.1109/ICMLA.2011.51

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2019). *A Comprehensive Survey on Transfer Learning*. https://arxiv.org./abs/1911.02685