Keyword Spotting Feature Extraction for Historical Handwritten documents

Varsha Thakur[1], HimaniSikarwar[2]
[1,2]Department of Computer Science and Engineering, Rajasthan College of Engineering for Women (RCEW)

Abstract-Deep learning is presently an effective research area in machine learning technique and pattern classification association. This has achieved huge success in the areas of many applications. The feature extraction that is become the big part in a field of deep learning technique such as Convolutional Neural Network (CNN) has dramatically advanced challenging computer vision tasks, especially in object detection and object classification, achieving state-of-the-art performance in several computer vision tasks including text recognition, sign recognition, face recognition and scene understanding. In this paper we will see how it become useful for the keyword spotting task in historical documents which consists in retrieving information from documents based on a keyword query. The query can be done by-example by providing an image of the searched keyword or by-string by providing the searched keyword itself.
Keywords:Deep Learning, feature extraction, convolutional neural network (CNN), Keyword Spotting, Handwritten Documents, Query by Example, and Query by String.

I. INTRODUCTION

In the last years, we have seen the rise of different approaches for the handwritten recognition task in computer vision but it is still a widely unsolved problem. Under different conditions, such as large vocabularies, with different writing styles or degraded documents, keyword spotting solutions have been suggested instead of a complete transcription to spot words in documents images[6][7].
Convolutional neural networks (convents, CNNs) have been established as the state-of-the art models in a wide range of vision tasks. The tasks of searching for relevant word images given an image or text query, known in the related literature as keyword spotting or word spotting (KWS)[30], is no exception to this rule. Several variants of convolutional networks for keyword spotting have been proposed and employed with success [34].

1.1 Image Feature extraction

Feature extraction can be defined as the fact of reducing an algorithm input data when this data is considered too large or redundant for processing. In machine learning, Feature Extraction begins with the initial set of consistent data and develops the borrowed values also called as features, expected for being descriptive and non-redundant, simplifies the consequent learning and observed steps. In few cases it Feature Extraction associates the decreasing the amount of assets needed to define a huge set of information. The result of feature extraction process is called feature vector. The feature vector is supposed to include the most relevant information from the input data which will allow the use of the reduced information instead of the whole initial data without compromising the task process [1][2].
Historical documents as a subset of handwritten documents are valuable resources for scholars so their contents can be made available via the internet or other electronic media [4].The main problem is that such contents are only available in image formats, which makes them difficult to search. In this case, document image word spotting techniques can be used to search the textual information from the digitized document images and make this information accessible to users. Word spotting is the task of locating pecific words in a collection of document images [3].
Firstly, optical character recognition has been employed for indexing documents. However, this approach is useless if documents are degraded or noised. Recently, researchers focused on developing document retrieval systems that are based on the analysis of some words describing the content of the researched document [4].
This approach that is called Keyword spotting has gained an increasing amount of research interest lately. The goal in word spotting is to retrieve parts of a document image collection with respect to a given query. Often times, this query representation is either an image (Query-by-Example, QbE) or a string defining the sought after word (Query-by-String, QbS) [5]. Keyword spotting methods can be separated in two categories. Template-based methods are comparing template images of the keyword query with document images. This has the advantage that template images are easy to obtain and that no knowledge of the underlying language is necessary. However, at least one template image is necessary for each keyword query [6]. Moreover, these systems typically do not generalize well to unknown writing styles. Dynamic Time Warping (DTW) has been extensively studied to match template images with segmented word images based on a sliding window and different features, such as word profiles, closed

contours or gradient features. Recent segmentationfree methods match template images with whole document images. On the other hand, learning-based systems are using supervised learning to train keyword models. These methods are expected to generalize better to unknown writing styles but they require a considerable amount of labeled training data. Hidden Markov models (HMM) have been proposed for modeling words or characters [7]. The character based approach is inspired by systems for complete transcription. It does not depend on keyword images for training and can be used to spot arbitrary keywords. Another character based approach is proposed in using recurrent neural networks.

Both categories are relying on features extracted from the images. Such features are generally handcrafted and optimizing them for different data sets is often difficult. Deep Learning solutions have shown that it is possible to learn features directly from pixels. Restricted Boltzmann Machines (RBM have been extensively used to extract features from data sets. Once stacked into Deep Belief Networks (DBN), they are able to extract multi-layer features from images. Convolutional RBM shave provene specially successful on images. General Convolutional Neural Networks (CNNs) are also used to extract features on large data sets of images or videos [8][9].

The systems that are proposed used for all type of scripts, documents and the letters, these features have been tested on well-known benchmark data sets for keyword spotting (IAM offline database, George Washington database and Parzival database, Cenparmi) and are compared with benchmark feature sets.

The rest of this paper is organized as follows. The General Framework of Word Spotting System is introduced in Section II. Section III presents the spotting methods. The Data sets are detailed in Section IV and Literature Revieware discussed in Section V. Finally, conclusions are drawn in Section VI.

II. THE GENERAL FRAMEWORK OF WORD SPOTTING SYSTEM

The process of word spotting is divided into two parts document archival processing and query processing [10]. Both the process have some common steps such as pre-processing and feature extraction which are described as follow,

2.1 Preprocessing

Generally, preprocessing stage in document image word spotting has the following processes: binarization, noise removal,skewcorrectionandline/wordsegmentation.Pre-processing is a major step for word spotting. It converts the data into such form that features can be extracted easily.

2.2 Word image representation

When building a document image word spotting system, a key consideration is how to represent word images within document images. This is fundamental to providing acceptable performance results. One of the most important advantagesoffeatureextractionisthatitreducesthestoragerequired and hence the system becomes faster and effective

2.3 Feature Extraction

For measuring the necessary shape information contained in the pattern, feature extraction is used which makes matching patterns easy just by using the formal procedure. The majority of word spotting techniques uses shape based features. Shape based features can be of two types low level and high level. The low-level feature describes more specific information like gradient direction, Edges, corners, and ridge. High-level features describe information like strokes, blobs, reservoir, region, etc. [11]

2.4 Feature Matching

The Matching process identifies most related word images from the document image with respect to the query word image. Matching can be done in two ways complete word matching and incremental word matching. Dynamic Time Wrapping is one of the widely used techniques for incremental word matching [12]. There are various distance measures available for word matching technique such as Euclidean distance, Cosine Similarity and Normalization

2.5 Cross Correlation (NCC).

Indexing and Ranking documents are also an important part but here we only focus on feature extraction and matching techniques.

III. SPOTTING METHODS

Several Word spotting techniques review in this section based on the used extracted feature.

3.1 Profile-based features

Global shape features such as the lower and the upper profiles capture the outline of a word. A profile is represented by a one-dimensional vector corresponding to the column-wise distance from the top of the bounding box to the foreground pixel of a word[13]. The distance between two profiles can be computed by any distance measure such as Euclidian distance (ED) and Dynamic Time Wrapping (DTW).

3.2 Gradient, Structural features

These features are capable of measuring the characteristics of an image at global, intermediate and local ranges, respectively. Generally, GSC is suitable features for handwritten document word spotting since they are able to capture the shape of the written words.

Similarly GLBP is a gradient feature that improves the Histogram of Oriented Gradients (HOG) [14] by calculating the gradient information at transitions of the Local Binary [15] Pattern code. For the matching step, some of approaches use the Euclidian Distance and the Cosine Similarity.

3.3 Bag of features

Thestate-of-the-artbagoffeaturesmodelhasbeenusedfor word spotting. Visual words are referred as visterms[16]. Matching is carried out with the use of bag-ofwordspoweredbySIFTdescriptorswhichareextractedfrom word images [17].

3.4 Other Features

Some graph-based approaches are proposed in which Attributed graphs are constructed using a part-based approach. While graph nodes correspond to grapheme which is extracted from convex groups of the skeleton, representadjacencyrelations between graphemes nodes.

Some introduced thePyramidalHistogramof Characters(PHOC)basedattributerepresentationwhichcan be used to represent both word images and strings. Fisher Vector representation of the images is used to the attribute representation. The spatial position of characters in word images is encoded using the Pyramidal Histogram of Characters(PHOC).

Convolution Neural Networks(CNNs)forfeatureextraction.Thisallowsbuilding robust representations for word spotting. Training was performed using stochastic gradient descent algorithmIntroducedPHOCNet, a deep CNN architecture trained with PHOC representation.

IV. DATA SETS

Development of robust document image word spotting systems requires databases of adequate size and diversity (many writers, multiple samples per writer, etc.) [4] That contains an adequate amount of variations of several factors such as script, writing styles, font size and quality. In this section, we will review several databases that have been used in the literature for various documents image understanding tasks including word spotting task.

4.1 George Washington database (GW)

This data set has become standard benchmarks for word spotting. It consists of 20 pages from a letter book by George Washington. The corresponding annotation contains word level bounding boxes and transcriptions for 4860 words. Written by George Washington in the year 1755

4.2 IAM database

Originally proposed as a handwriting recognition benchmark, the IAM-DB data set has recently enjoyed an increased use as word spotting benchmark as well. It contains a total of 115320 words from 657 different writers. It is divided into three sets: training, testing and validation. A good property of this data is that each set includestextlineswrittenbyseveralwriterswhichmake it a good choice for word spotting with different writing styles.

4.3 Parzival database

The Parzival database includes 45 pages written using German language in the thirteenth century. It is consideredasagoodchoiceforwordspottingtaskas itiswritten by three writers [18].

4.4 Cenparmi

The CENPARMI database includes 137 documents written by 13 writers. The database contains 2107 text lines. It is divided into two sets: testing and validation. The testing set contains 112 documents while the validation set contains 25 documents.

Table 1 shows the summary of the databases used in word spotting tasks.

Databases	Description	Writers
GW	20 Pages	2
CENPARMI	137 Documents	13
IAM Database	1539 Pages	657
Parzival database	45 Pages	3

V. LITERATURE REVIEW

5.1 Early Work for Word Spotting based on the Variousfeature Extraction Techniques

Ali Abidi et al. [19] have used the height, width and convex area of each partial word as scalar features along with the vertical and horizontal word profiles. The reported act rate was 72 % in terms of F-Allowance on 115 queries and 90 handwritten documents written by 90 writers.

Later Wie and Gao [20] imported a system for locating words in the Mongolian Kanjur documents. The authors used four profile-based features for depicting word images. Queries are synthesized using pre-extracted glyph. They figured their system using 200 Mongolian Kanjur document images where the average R-precision rate was 61.02%. The R-precision is the ratio of the number of admissible retrieved words in the top R results to the total number of admissible words.

To better handle intra-class variability, Giotis et al. [21] proposed a shape-based equivalent scheme that makes use of local contour features formed by continuous connected curves. The evaluation is done using the GRPOLY-DBI [22] and the GW datasets. A mAP of 60.04 and 37.86 % was obtained for the tow datasets, respectively.Rabaev et al. [23] propose to use a pyramid-based method where HOG features are extracted at each level of the pyramid. For query matching, hierarchical searchisperformedstartingfromthehighestlevelofthepyramid.Fourdatasetsareusedto figure theproposedmethod, the GW, Lord Byron, the Cairo Genizah collection and Arabic Manuscripts from Harvard University. The reported mAP value is 85.53 % on the GW dataset.

Retsinas et al. [24] proposed to use a combination of global and local descriptors based on projection of oriented gradients, POG. Euclidean distance is used for query matching. They figured their approach on the Bentham Dataset and reported a mAP value of 77.7 %.

Graph-based methods [25, 26] are proposed to provide a prosperous representation that is prosperous to the deformations of handwriting and to defeat the computational cost associated with the traditional graph matching algorithms.

Buietal. [27] Proposed a graph-based method to describe the structural properties of the word images using in substitutes as a descriptor. The graph is constructed using the strokes extracted from the word images. The vertexes of the graph represent the strokes while the edges are corresponding to the spatial relationship between strokes. Distance between two word images is delineated as edit distance between their corresponding graphs.

Another graph-based approach is proposed in Riba et al. [28]. Attributed graphs are constructed using a part-based approach. While graph nodes correspond to grapheme which is extracted from convex groups of the skeleton, graph edges represent adjacency relations between graphemes nodes. The proposed approach is figured on a subset of the Barcelona Historical Handwritten Marriages Database (BH2M). They reported a mAP value of 51.62 %

Hongxi, Guanglai proposed a Bag-of-Visual-Words (BoVW) model for the Mongolian image Documents which consists of LLC method, query likelihood model and KL divergence, have been tested and compared with the baseline. The best act is 13.43% when the number of clusters increases to 5,500 for BoVW, the best act of LLC is reached at 27.99% when the number of neighbors is 2 and the number of clusters is 500. Infer aptly, the act of the LLC method is twice as much as the original BoVW approach. [29]

Rusiñol et al. [30] proposed a segmentation-free word spotting method based on a bag of features model that makes use of SIFT. First, the document image is divided into a set of local patches. The local descriptors are extracted for each patch as well as for the query image. Then, the distance between the query image descriptors and patch descriptors is computed to find the distance between the query image and each patch. Then, the document patches with a higher probability of containing instances of the query are retrieved.

Czunietal. [31] Used local features based on different variations of SIFT descriptor to spot handwritten words. In his method, each feature point (q) of the query is only compared to the candidate points of the candidate words which lies inside a circle centered at the point q. Then, matching points set is constructed from all feature point pair shaving the least Euclidean distance and the least difference in orientation. The distance is computed with the use of the matching points for the query and candidate words. The authors reported 85 % accuracy on 22 pages.

5.2 Works on CNN for Feature Extraction

Sehla, Afef [32] gave the two approaches for feature extraction and image encoding which are Bag of Feature and deep learning and especially Convolutional Neural Networks (CNN). Concretely, they have used the “AlexNet” CNN model trained to perform well on the Image Net dataset. The large act gap between these two families of approaches make the BoF image encoding technique useless in result.

Sharma and Sankar [33] adapted Convolution Neural Networks (CNNs) for feature extraction. This allows building prosperous representations for word spotting. Training was performed using stochastic gradient descent algorithm. High and low learning rates are assigned for the classification and convolution layers, respectively.

The proposed method is figured on IAM dataset and reported a mAP value of 46.53 %.The advantage of this approach is that it reduces data demand and computational cost.

Sudholt and Fink [34] importedPHOCNet, a deep CNN architecture trained with PHOC representation. The authors figured their work using three datasets; GW, IAM, IFN/ENITdatasets.ThereportedmAPvalueswere92.64%, 82.97 and 96.11 on the three datasets, respectively. As claimed by the authors, the proposed method outperforms current state-of-the-art artapproaches and has short training and test time.

Table 2 Early Work for Word Spotting based on the variousfeature Extraction Techniques

References	Features	Language	Database size	Document	Accuracy type
Ali Abidi et al. [19]	DTW	Urdu	90 Pages	Handwritten	F: 72 %
Wei and Gao [20]	DCT	Mongolian	200 Pages	Handwritten	P: 61.02 %
Giotis et al. [21]	PAS	dissimilarity Greek	GRPOLY-DBI	Handwritten	P: 60.04 %
		English	GW	Historical	P: 37.86 %
Rabaev et al. [23]	HOG	English	GW	Historical	mAP: 85.53 %
Retsinas et al. [24]		English	Bentham	Handwritten	mAP: 57.7 %
BUI et al. [25]	Graph-based	German	Parzival	Historical	
Riba et al. [28]	Graph-based	English	BH2M	Handwritten	mAP: 51.62 %
Rusiñol et al. [30]	SIFT	English	GW	Historical	mAP: 61.35 %
Czúnietal. [31]	SIFT		22 Pages	Handwritten	mAP: 85%

P precision, mAP mean average precision, F-measure, A- accuracy

Table 3various word spotting techniques using CNN for features Extraction

References	Features	Language	Database size	Document	Accuracy type
Sharma and Sankar [33]	CNN	English	IAM	Handwritten	mAP: 46.53 %
Sudholt and Fink [34]	CNN	English	GW	Historical	mAP: 92.64 %
	CNN	English	IAM	Handwritten	mAP: 82.97 %

VI. CONCLUSION

We have studied a number of important aspects concerning the models and the class of methods used for the Keyword spotting. Feature Extraction is a technique that reduces the amount of input data by refining its representative expressive attributes; here in our work

We have presented the experiment that helps us to draw some insightful conclusions.

In most deep learning applications, CNNs are used for huge amount of training samples to learn powerful classifiers, so same in our work a well-designed CNN structure has been proposed, which can handle limited training samples for the images by high quality features.

In our Experiment the Dataset is extracted, all the features are extracted firstly and then convolutional neural networking is done.

To summarize, our proposed model allows users to see the result with convolutional layers for different databases and these will give the better results than their traditional usage.

As future work, we would like to examine the act of more advanced ensemble strategies.

VII. REFERENCES

[1] Suresh Dara and PriyankaTumma, “Feature Extraction By Using Deep Learning: A Survey” ,Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record

[2] SehlaLoussaief, AfefAbdelkrim “Deep Learning vs. Bag of Features in Machine Learning for Image Classification”, 978-1-5386-4449-2/18/$31.00 ©2018 IEEE

[3] Mohamed Lamine, HassibaNemmour, “New Gradient Descriptor for Keyword Spotting in Handwritten Documents”. 3rd International Conference on Advanced Technologies for Signal and Image Processing - ATSIP' in 2017 May 22-24, Fez, Morroco.

[4] Rashad Ahmed, Wasfi G. Al-KhatibSabri Mahmoud “A Survey on handwritten documents word spotting” Int J Multimed Info Retr (2017) IEEE.
[5] Sebastian Sudholt and Gernot A. Fink, “Evaluating Word String Embedding’s and Loss Functions for CNN-based Word Spotting”, in 14th IAPR International Conference on Document Analysis and Recognition, 2017 IEEE
[6] Baptiste Wicht, Andreas Fischer, Jean Hennebert “Deep Learning Features for Handwritten Keyword Spotting” in 23rd International Conference on Pattern Recognition (ICPR) Cancún Center, Cancún, México, December 4-8, 2016
[7] Christophe Choisy, “Dynamic Handwritten Keyword Spotting based on the NSHP-HMM” Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) IEEE
[8] Kien Nguyen, Clinton Fookes, SridhaSridharan “ Improving deep convolutional neural networks with unsupervised Feature learning” in 2015 IEEE
[9] HimanshuM. Kathiriya, Mukesh M. Goswami“Word Spotting Techniques for Indian Scripts” , International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017]
[10] AlonKovalchuk, Lior Wolf, and NachumDershowitz, “A Simple and Fast Word Spotting Method”, Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on. IEEE, 2014.
[11] H. Toselli and E. Vidal, “Fast HMM-filler approach for key word spotting in handwritten documents,” in Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013, pp. 501–505.
[12] George Retsinas, GeorgiosLouloudis, NikolaosStamatopoulos and Basilis Gatos ,“ Efficient Learning-Free Keyword Spotting” ,Citation information: DOI 10.1109/TPAMI.2018.2845880, IEEE Transactions on Pattern Analysis and Machine Intelligence
[13] HamzaGhilas, meriemgagaoua, abdelkameltari, mohamedcheriet,, “Arabic Word Spotting Based on Key-Points Features” , in 2017 IEEE
[14] Dao Wu, Rui Wang, XiaoweiTian, Dong Liang, Xiaochun Cao , “The Keywords Spotting with Context for Multi-Oriented Chinese Scene Text” in 2018 IEEE FourthInternational Conference on Multimedia Big Data (Big MM)
[15] KonstantinosZagoris, Member, IEEE, IoannisPratikakis, Senior Member, “Unsupervised Word Spotting in Historical Handwritten Document Images Using Document-Oriented Local Features” IEEE transactions on image processing, vol. 26, no. 8, august 2017
[16] Hongxi Wei, Hui Zhang, GuanglaiGao , “Representing word image using Visual word embeddings and RNN for keyword spotting on historical document images”, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2017
[17] Hongxi Wei, GuanglaiGao, “Visual Language Model for Keyword Spotting on Historical Mongolian Document Images”, 2017 IEEE
[18] Giotis A, Sfikas G, Nikou C, Gatos B (2015) Shape-based word spotting in handwritten document images. In: 13th international conference on document assay and recognition (ICDAR),2015, 561–565
[19] Gatos B, Stamatopoulos N, Louloudis G, Sfikas G, Retsinas G, Papavassiliou V, Sunistira F, Katsouros V (2015) Grpoly-db: an old greekpolytonic document image database. In: 13th international conference on document assay and recognition (ICDAR), 2015, 646–650
[20] Giotis A, Sfikas G, Nikou C, Gatos B (2015) Shape-based word spotting in handwritten document images. In: 13th international conference on document assay and recognition (ICDAR),2015, 561–565
[21] Rabaev I, Kedem K, El-Sana J (2016) Keyword retrieval using scale-space pyramid. In:12th IAPR workshop on document assay systems (DAS), 2016, 144–149
[22] Retsinas G, Louloudis G, Stamatopoulos, Gatos B (2016) Keyword spotting in handwritten documents using projections of oriented gradients. In:12th IAPR workshop on document assay systems (DAS), 411–416
[23] Bui QA, Visani M, Mullot R (2015) Unsupervised word spotting using a graph representation based on in substitutes. In: 13th international conference on document assay and recognition(ICDAR), 2015, 616–620
[24] 102. Riba P, Lladãs J, Fornés A (2015) Handwritten word spotting by inexact matching of grapheme graphs. In: 13th international conference on document assay and recognition (ICDAR),2015, 781–785
[25] Hongxi Wei, GuanglaiGao, “Visual Language Model for Keyword Spotting on Historical Mongolian Document Images” , IEEE 2017
[26] Leydier M, Aldavert D, Toledo R, Llados J (2011) Browsing heterogeneous document collections by a segmentation-free word spotting method. In: International conference on document assay and recognition, 63–67
[27] Czuni L, Kiss P, Gal M, Lipovits A (2013) Local feature based word spotting in handwritten archive documents. In: 2013 11th international workshop on content-based multimedia indexing (CBMI), 179–184
[28] SehlaLoussaief, AfefAbdelkrim, “Deep Learning vs. Bag of Features in Machine Learning for Image Classification”, IEEE 2018
[29] Sharma A, Sankar KP (2015) Adapting off-the-shelf cnnsforword spotting and recognition. In: 13th international conference on document assay and recognition (ICDAR), 986–990, x
[30] Sudholt S, Fink GA (2016) PHOCNet: a deep convolutional neural network for word spotting in handwritten documents. https:// arxiv.org/pdf/1604.00187.pdf (ICDAR), 2013 12th international conference on, 567–571
[31] Liu C-L, Yin F, Wang D-H, Wang Q-F (2011) Casia online and offline chinese handwriting databases . IntConf Doc Anal Recognit 2011:37–41 87.
[32] Perronnin F, Rodriguez-Serrano JA (2009) Fisher kernels for handwritten word spotting. In: Proceedings of the 2009 10th international conference on document analysis and recognition, ICDAR ’09, (Washington, DC, USA), 106–110, IEEE computer society
[33] Roy PP, Rayar F, Ramel J-Y (2015) Word spotting in historical documents using primitive code book and dynamic programming. Image Vis Comput 44:15–28 41.
[34] Giotis A, Sfikas G, Nikou C, Gatos B (2015) Shape-based word spotting in handwritten document images. In: 13th international conference on document analysis and recognition (ICDAR),2015, 561–565
[35] Abidi A, Jamil A, Siddiqi I, Khurshid K (2012) Word spotting based retrieval of Urdu handwritten documents. In: Proceedings of the 2012 international conference on frontiers in handwriting recognition, ICFHR’12,(Washington, DC,USA),331–336,IEEE Computer Society
[36] Wei H, Gao G (2014) A keyword retrieval system for historical Mongolian document images. IJDAR 17(1):33–45 27.
[37] Kesidis A, Galiotou E, Gatos B, Lampropoulos A, Pratikakis I, Manolessou I, Ralli A (2009) Accessing the content of greek historical documents. In: Proceedings of the third workshop on analytics for noisy unstructured text data, AND ’09, (New York, NY, USA), 55–62, ACM
[38] Srihari S, Srinivasan H, Babu P, Bhole C (2006) Spotting words in handwritten arabic documents. In: Document recognition and retrieval XIII: Proceedings SPIE .
[39] SrihariS, SrinivasanH, BabuP, BholeC(2005)Handwritten Arabic word spotting using the cedArabic document analysis system. Proceedings 2005 symposium on
[40] Rodríguez-Serrano JA, Perronnin F (2009) Handwritten word spotting using hidden markov models and universal vocabularies. Pattern Recognit 42(9):2106–2116.

[41] P. Krishnan, K. Dutta, and C. V. Jawahar, “Deep feature embedding for accurate recognition and retrieval of handwritten text,” in Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016, pp. 289–294.

[42] Yushi Chen, Hanlu Jiang, Chunyang Li, XiupingJia, and PedramGhamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232– 6251, 2016.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”. In Advances in neural information processing systems, pp. 1097-1105, 2012.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei, “Imagenet: A large-scale hierarchical image database”. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248-255, 2009.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla and M. Bernstein, “Imagenet large scale visual recognition challenge”. arXiv preprint arXiv:1409.0575, 2014.

[46] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding”. arXivpreprint arXiv:1408.5093, 2014.

[47] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines”. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807-814, 2010.

[48] P. Domingos, “A Few Useful Things to Know about Machine Learning,” Communications of the ACM, vol. 55, no. 10, p. 78, 2012.

[49] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” European Conference on Computer Vision, pp. 346–361, 2014.

[50] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,” in Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 2169–2178.

[51] C. Wieprecht, L. Rothacker, and G. A. Fink, “Word Spotting in Historical Document Collections with Online-Handwritten Queries,” in International Workshop on Document Analysis Systems, 2016.

[52] U. V. Marti and H. Bunke, “The IAM-database: An English Sentence Database for Offline Handwriting Recognition,” International Journal on Document Analysis and Recognition, vol. 5, no. 1, pp. 39–46, 2002.

[53] V. Romero, A. Forn´es, N. Serrano, J. A. S´anchez, A. H. Toselli, V. Frinken, E. Vidal, and J. Llad´os, “The ESPOSALLES database: An ancient marriage license corpus for off-line handwriting recognition,” Pattern Recognition, vol. 46, no. 6, pp. 1658–1669, 2013.

[54] Pratikakis, K. Zagoris, B. Gatos, J. Puigcerver, A. H. Toselli, and E. Vidal, “ICFHR2016 Handwritten Keyword Spotting Competition (HKWS 2016),” in International Conference on Frontiers in Handwriting Recognition, 2016, pp. 613–618.

[55] P. Krishnan, K. Futta and C.V. Jawahar, “Deep feature embedding for accurate recognition and retrieval of handwritten text”, International Conference on Frontiers in Handwriting Recognition, pp. 289-294, China, 2016.

[56] P. Krishnan and C.V. Jawahar, “Matching Handwritten Document Images”, European Conference on Computer Vision, pp. 766-782, Netherlands, 2016.