Deep Learning Feature Extraction for Handwritten Keyword Spotting in Historical Documents

Varsha Thakur, Himani Sikarwar
Department of Computer Science and Engineering
Rajasthan College of Engineering for Women (RCEW)
Email: varshavbn@gmail.com

Abstract-Deep learning is presently an effective research area in machine learning technique and pattern classification association. This has achieved big success in the areas of many applications. The feature extraction that used in a 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 some of it uses for keyword spotting in handwritten 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.*

*****

I. Introduction

Conversion of given input data in to set of features are known as Feature Extraction. 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. An approach that decreases the amount of given data by extracting the detailed attributes is a procedure of assuming different features from the previously given features in order to decrease the cost of feature analysis, develop classifier accuracy and permit bigger classification efficiency [1].

Historical documents as a subset of handwritten documents are valuable resources for scholars so their contents can be made available via the in ternetor other electronic media. 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 specific words in a collection of document images [2].

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 [3].

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) [4].

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 [5]. 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 segmentation free 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 [6]. 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 proven especially successful on images. General Convolutional Neural Networks (CNNs) are also used to extract features on large data sets of images or videos [7][8].

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 results are 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 [9]. Both the process have some common steps such as pre-processing and feature extraction which are described as follow,

[1] Preprocessing

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

[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 advantages of feature extraction is thatit reduces the storage required and hence the system becomes faster and effective

[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. [10]

[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 [11]. There are various distance measures available for word matching technique such as Euclidean distance, Cosine Similarity and Normalization 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.

- 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[12]. The distance between two profiles can be computed by any distance measure such as Euclidian distance (ED) and Dynamic Time Wrapping (DTW).

- Gradient, Structural features

  These features are capable of measuring the characteristics of an image at global, intermediate and local ranges, respectively. Generally, GSC are 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) [13] by calculating the gradient information at transitions of the Local Binary [14] Pattern code. For the matching step, some of approaches use the Euclidian Distance and the Cosine Similarity.

- Bag of features

  The state-of-the-art bag of features model has been used for word spotting. Visual words are referred as visterms[15]. Matching is carried out with the use of bag-of words powered by SIFT descriptors which are extracted from word images [16].

- Other Features

  Some graph-based approaches are proposed in which Attributed graphs are constructed using a part-based approach. While graph nodes correspond to graphe me which is extracted from convex group softhe skeleton, represent adjacency relations between graphemes nodes.

  Some introduced the Pyramidal Histogram of Characters (PHOC) based at tribute representation which can 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 Character s(PHOC).

Convolution Neural Networks (CNNs) for feature extraction. This allows to build robust representations for word spotting. Training was performed using stochastic gradient descent algorithm

Introduced PHOCNet, a deep CNN architecture trained with PHOC representation.

IV. Data Sets

Development of robust document image word spotting systems requires data bases of a dequate size and diversity (many writers, multiple samples per writer, etc.) that contain an adequate amount of variations of several factors such as script, writing styles, font size and quality. In this section, were view several data bases that have been used in the literature for various document image understanding tasks including word

Spotting task.

- 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

- 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.

- Parzival database

  The Parzival database includes 45 pages written using German language in the thirteenth century. It is consideredasagoodchoiceforwordspottingtaskas itiswritten by three writers [17].

- CENPARMI

  TheCENPARMIdatabaseincludes137documentswritten 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. Result

In Dao Wu, XiaoweiTian [13] they used SSD directly recognize the characters in the CTW dataset. The input size of SSD is 512 ∗ 512 which differs from those in CTW dataset. With the performance of 65.54% mAP. Meanwhile, the YOLOv2 has the 71% mAP in the detection task. For Keyword spotting, from the CTW dataset, They choosed four keywords for experiment in the training and validation set of CTW which contains 25,887 images with 812,872 Chinese characters.Context Extractor Module can improve precision to 92%. Their average F-measure finally achieves 94% which performs extraordinarily well.

Kolcz et al. proposed the use of a line-oriented approachforhandwrittendocumentswordspottingbasedon matching profile-oriented features. The researchers experimented with old Spanish manuscripts and did not report the performance rate.

Sigappietal.[21] proposedasystemforwordspottingof Tamil language.Thefeaturesusedareprojectionprofile,lowerand upper word profile and Background-to-ink transitions features from each segmented word. This approach makes use of HMM to characterize stroke's variations of handwrittencharacters.Thereportedaccuracywas80.75%.How ever, the test is done using small dataset containing only 400 words.

Hamzaghilas, meriemgagaoua[12] a template matching based word spotting system for Arabic historical documents is proposed and the results show that the proposed method outperforms DTW in both precision and time response for the majority of queries. An MAP of 77.63 was obtained for the proposed method with a mean time response of 5.94 seconds. On the other hand, an MAP of 0.60 with a mean time response of 15.78 seconds was the results of the well-known DTW method.

Christophe Choisy [6] it was based on the NSHP-HMMmodel hybrids an HMM and a Markov Field (MRF). It analyzes binary patterns column by column at the HMM level, and pixel by pixel considering a neighborhood θ at the MRF level with mean score is35.994%, and the mean balanced by the word sample numberis 28.41%.

In [33,34], Srihari et al. make use of binary GSC features to spot words in handwritten documents. The authors segmented text lines into connected components. They used correlation distance for matching two GSC binary feature

vectors.Themethodwastestedonadatasetwrittenby8writers and reported a precision rate of 70 % at a recall of 50 %.

Rodríguez and Perronnin [35] presented a statistical frameworkbasedontwotypesofHMMs,SC-HMMsandCHMMs. They used query by word-class instead of querying by string or by example to overcome the downsides of both approaches. Column features, pixel count features and LGH features areemployedinthis work.Evaluation is done using 630 images written in French. Experimentally, they showed a mean average precision value of 87 %. They also showed that the SC-HMM is superior to the C-HMM and is consistently superior to DTW.

Almazánetal.IntroducedthePyramidalHistogramof 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).Usingthisrepresentation,imagesandstarin g matching is reduced to a nearest neighbour problem which allowstoachievehighaccuracyincaseof segmented words. However, it can not be applied directly in a segmentationfree approach as it involves computation of costly Fisher Vector representation at query time. They evaluated this methodinfourpublicdatasets:TheIAM,GWdatasets.They reported 80.64 and 93.3 % mAP on the tow datasets, respectively.

Inarelatedresearch,thisworkisextendedtoasegmentation-free scenario in [100]. Indexing based on character bi-gram is used to overcome the computational cost. They reported a mAP of 48.57 and 73 % on the IAM and GW datasets, respectively.

VI. Conclusion

Feature Extraction is a technique that reduces the amount of input data by refining its representative expressive attributes; here we saw some of its techniques for handwritten keyword spotting.

In recent years, we have seen so many techniques that methods used for the handwritten documents keyword spotting. In spite of that a lot of work remains to be done.

In this part, we will see some diverse challenges and future research work.

- Many researchers have focused on single writer document but using multi writer documents will give more variation in this field. So future work can include such variation of words.
- In the earlier work many approaches cannot perform the out of vocabulary word spotting.
- Using gray scale images instead of binary images could also cause to rebust features.
- Runtime Performance can also be improving for future research.
- There are very few efforts on Arabic and Chinese historicalwordspotting.Therefore,moreresearchworkisn eededfor such documents.

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]. 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.

[3]. Rashad Ahmed, Wasfi G. Al-KhatibSabri Mahmoud “A Survey on handwritten documents word spotting” Int J Multimed Info Retr (2017) IEEE.

[4]. Sebastian Sudholt and Gernot A. Fink, “Evaluating Word String Embeddingsand Loss Functions for CNN-based Word Spotting”, in 14th IAPR International Conference on Document Analysis and Recognition, 2017 IEEE

[5]. 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

[6]. Christophe Choisy, “Dynamic Handwritten Keyword Spotting based on the NSHP-HMM” Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) IEEE

[7]. Kien Nguyen, Clinton Fookes, SridhaSridharan “ Improving deep convolutional neural networks with unsupervised Feature learning” in 2015 IEEE

[8]. Himanshu M. Kathiriya, Mukesh M. Goswami “Word Spotting Techniques for Indian Scripts” , International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017]

[9]. AlonKovalchuk, Lior Wolf, and NachumDershowitz, “A Simple and Fast Word Spotting Method”, Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on. IEEE, 2014.

[10]. 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.

[11]. 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

[12]. Hamzaghilas, meriemgagaoua, abdelkameltari, mohamedcheriet , “Arabic Word Spotting Based on Key-Points Features” , in 2017 IEEE

[13]. Dao Wu, Rui Wang, XiaoweiTian, Dong Liang, Xiaochun Cao , “The Keywords Spotting with Context for Multi-

Oriented Chinese Scene Text" in 2018 IEEE Fourth International Conference on Multimedia Big Data (Big MM)

[14]. 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

[15]. 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

[16]. Hongxi Wei, GuanglaiGao, "Visual Language Model for Keyword Spotting on Historical Mongolian Document Images", 2017 IEEE

[17]. 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

[18]. Nicholas R. Howe Andreas Fischer Baptiste Wicht, " Inkball Models as Features for Handwriting Recognition",2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 96-101

[19]. K. Terasawa and Y. Tanaka, "Slit style HOG feature for document image word spotting,"in Proceedings of the IEEE Int. Conf. on Document Analysis and Recognition. IEEE, 2009, pp. 116–120.

[20]. M. Rusinol, D. Aldavert, R. Toledo, and J. Llad´os, "Browsing heterogeneous document collections by a segmentation-free word spotting method," in Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011, pp. 63–67.

[21]. SharmaA,SankarKP(2015)Adaptingoff-the-shelfcnnsforword spotting and recognition.In:13the internationalconferenceondocument analysis and recognition (ICDAR), 986–990, x

[22]. Sigappi A, Palanivel S, Ramalingam V (2011) Handwritten document retrieval system for tamil language. Int J ComputAppl 31:42–4

[23]. Retsinas G, Louloudis G, Stamatopoulos, Gatos B (2016) Keyword spotting in handwritten documents using projections of orientedgradients.In:12thIAPRworkshopondocumentanalysis systems (DAS), 411–416

[24]. Zhang H, Wang D-H, Liu C-L (2010) Keyword spotting from online chinese handwritten documents using one-vs-all trained characterclassifier.IntConfFrontHandwritRecognit2010:271 – 276 89.

[25]. Zhang H, Zhou X-D, Liu C-L (2013) Keyword spotting in online chinese handwritten documents with candidate scoring based on semi-CRF model. In: Document analysis and recognition (ICDAR), 2013 12th international conference on, 567–571

[26]. Liu C-L, Yin F, Wang D-H, Wang Q-F (2011) Casia online and offlinechinesehandwritingdatabases.IntConfDocAnalRecognit 2011:37–41 87.

[27]. 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

[28]. 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.

[29]. 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

[30]. 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

[31]. Wei H, Gao G (2014) A keyword retrieval system for historical mongolian document images. IJDAR 17(1):33–45 27.

[32]. 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

[33]. Srihari S, Srinivasan H, Babu P, Bhole C (2006) Spotting words in handwritten arabic documents. In: Document recognition and retrieval XIII: Proceedings SPIE .

[34]. SrihariS,SrinivasanH,BabuP,BholeC(2005)Handwrittenarabic word spotting using the cedarabic document analysis system. In:Proceedings2005symposiumondocumentimageunderstanding technology

[35]. Rodríguez-Serrano JA, Perronnin F (2009) Handwritten wordspottingusinghiddenmarkovmodelsanduniversalvocabularies. Pattern Recognit 42(9):2106–2116.