ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.3

Self Assessment System for Blind People in Library Management System

M.Karthick[1], M.Manikandan[2], J.K.Murali[3], M.Yuvarajkumar[4], A.R.Narendrakumar[5]
BE Student[1, 2, 3, 4], Assistant Professor[5]
Department of Computer Science and Engineering
University College of Engineering Thirukkuvalai, India

Abstract:
Personal assistant and navigator system for visually impaired people. Design a library system that integrates Object recognition and Computer Vision algorithms. Implement audio system to know about book with return details. Using Hidden Markov Model to convert text to voice and voice to text. Proposing to use object detection for blind people and give them audio vocal information of object and current location of blind user. Detecting book name using the camera and giving voice instructions about the direction of book in library mall.

Index Terms: Object recognition, Scene Text Recognition, Convolutional Neural Network, Optical Character Recognition, Computer Vision algorithms

1. INTRODUCTION

Recently, the community has seen a sturdy recovery of neural networks, which is mainly interested by the great success of deep neural network models‖, exactly DCNN, and voice recognition in various vision tasks. However, popular of the fresh works related to deep neural networks have devoted to detection or classification of object groups In this paper, we are troubled with a typical problem in computer vision: image-based sequence recognition. In real world, a stable of visual objects, such as scene text, handwriting and musical score, tend to occur in the form of sequence, not in isolation. Unlike overall object appreciation, recognizing such classification like objects regularly requires the system to expect a series of object labels, instead of a single label. Therefore, recognition of such objects can be naturally cast as a sequence recognition problem. Another unique property of sequence-like objects is that their lengths may vary drastically. For instance, English words can either consist of 2 characters such as ―OK‖ or 15 characters such as compliments. Then, the most common subterranean models like CNN cannot be straight practical to sequence intention, since CNN models often operate on inputs and outputs with fixed dimensions, and thus are incapable of producing example, the processes in initially detect single characters and then recognize these detected fonts with DCNN models, which are qualified using considered character images. Such methods often need training a strong character detector for exactly detecting and collecting each character out from the unique word image some other styles treat scene text recognition as an image arrangement problem, and assign a class label to each English word (90K words in total). It chances out a large qualified model with a enormous number classes. Which is difficult to be generalized to other types of sequence-like objects, such as Chinese texts, musical notches, etc., because the numbers of basic groupings of such kind of sequences can be greater than 1 million. In sudden, current organizations based on DCNN cannot be straight used for image-based sequence appreciation. For example, Graves et al. cutting a set of geometrical or image structures from handwritten texts, while Su and Lu convert word images into sequential HOG features. The step is unconventional of the consequent instruments in the channel, thus the existing systems based on RNN can be trained and enhanced in an end-to-end style. Some conventional scene text appreciation methods that are not based on neural networks also brought discerning ideas and novel representations into this field. For example, Almas a et al. and Rodriguez Serrano et al. Projected to embed word images and text strings in a common, and word acknowledgment is converted into a recovery problem. Though achieved promising presentation on standard ethics, these methods are generally outperformed by previous algorithms based on neural networks as well as the approach proposed in this paper. The proposed neural network model is named as Convolutional Recurrent Neural Network (CRNN), since it combines DCNN and RNN, and constructs an end-to-end system for sequence recognition. Although both DCNN and RNN are well-established techniques, to the best of our knowledge we are the first to combine them for sequence recognition. Note that integrating DCNN and RNN in an end-to-end manner has also been explored in image classification explored for describing the context of scene text images, but also used for predicting the structured outputs of sequence-like objects. As a result, fully-connected layers for producing structured outputs like are not essential in our method, and the model size is greatly reduced. For system like materials, saves some single advantages over conservative models: 1)It can be directly learned from sequence labels for occurrence, words needing no detailed reasons for instance, characters It has the same property of DCNN on learning informative representations directly from image data, requiring neither hand-craft features nor preprocessing steps, including linearization segmentation, component localization, etc.; 3) It has the same property of RNN, being able to produce a sequence of labels; 4) It is unrestrained to the lengths of classification-like objects, needing only height stabilization in both training and testing phases; 5) It achieves better or highly competitive

performance on scene texts (word recognition) than the prior arts 6) It contains much less limitations than a standard DCNN model, overwhelming less storage space

2. PROPOSED NETWORK ARCHITECTURE

The network planning of RNN, as shown in Fig. 2.0, consists of three apparatuses, containing the convolutional layers, the regular layers, and a transcription layer, from lowest to top.

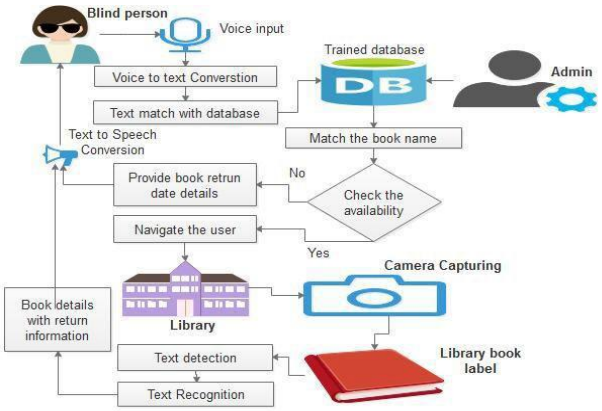

Figure.1. Architecture for Self-assessment system for blind people in library management system

At the bottom of CRNN, the convolutional layers automatically extract a feature sequence from each input image. On top of the convolutional network, a recurrent network is built for making control for each frame of the feature sequence, outputted by the convolutional layers. The transcription layer at the top of CRNN is adopted to translate the per-frame predictions by the recurrent layers into a label sequence. Though CRNN is collected of different types of network designs (e g. DCNN and RNN), it can be jointly trained with one loss function In CRNN model, the component of convolutional layers is constructed by taking the convolutional and max-pooling layers from a standard DCNN model (fully-connected layers are removed). Such component is used to extract a consecutive feature representation from an participation image. Beforehand being fed into the network, all the images need to be scaled to the same height. Then a structure of article vectors is removed from the feature maps produced by the component of convolutional layers, which is the input for the recurrent layers. Specifically, each feature vector of a feature sequence is generated from left to right on the feature maps by column. This means the i-the feature vector is the concatenation of the i-the columns of all the maps. The width of each column in our settings is fixed to single pixel. As the layers of difficulty, max-combining, and element wise beginning function control on local regions, they are conversion invariant. Then, each column of the feature maps matches to a square district of the original image (termed the amenable field), and such rectangle states are in the same order to their conforming columns on the article maps from left to right. As illustrated in Fig. 2.0, each vector in the feature sequence is associated with a receptive field, and can be considered as the image descriptor for that region.

2.2 Sequence Labeling

A deep ―Recurrent Neural Network‖ is made on the top of the convolutional layers, as the recurrent layers. The recurrent layers predict a label distribution for each frame in the feature sequence. The benefits of the recurrent layers are three-fold. Firstly, RNN has a tough ability of taking background material within an arrangement. Using circumstantial cues for image- based sequence recognition is more stable and helpful than treating each symbol independently. Taking scene text recognition as an example, wide characters may require several successive frames to fully describe. Besides, some ambiguous characters are easier to extricate when detecting their contexts, e.g. it is cooler to diagnose changed the charm heights than by recognizing each of them independently. Next, i.e. the convolutional layer, agreeing us to equally train. Thirdly, RNN is able to function on sequences of chance lengths, navigating from starts near ends. A outmoded RNN unit has a self-connected hidden layer between its input and output layers. Each time it receives a frame in the sequence, it updates its internal state.

2.3 Transcription

Transcript is the process of changing the per-frame guesses made by RNN into a label sequence. Mathematically, dictation is to find the label sequence with the highest probability adapted on the per-frame predictions. In duplication, there exists two modes of dictation, namely the lexicon-free and lexicon based copies. A lexicon is a set of label sequences that prophecy is constraint to, e.g. a spell checking dictionary. In lexicon-free mode, expectations are made without any lexicon. In lexicon-based mode, predictions are made by choosing the label sequence that has the highest probability.

2.3.1 Probability of label sequence

We adopt the conditional probability defined in the Connectionist Temporal Classification (CTC) proposed by Graves et al. The probability is defined on the label sequence l conditioned on the per-frame predictions y and it ignores the specific position of each label. Accordingly, when we use the damaging log-likelihood of this chance as the objective function, we only need images and their agreeing label sequences, avoiding the labor of cataloging the positions of individual sequence elements.

3. IMPLEMENTATION DETAILS

The network conformation we use in our experimentations is abridged in. The manner of the convolutional layers is resulting from the VGG Very Deep constructions. A yank is made in order to create it right for knowing English texts. In the 3rd and the 4th max-pooling layers, we adopt pooling strides instead of the conventional strides. This tweak yields feature maps with larger widths, hence longer feature sequence. For example, an image containing 10 characters is typically of size 100 32, from which a feature sequence with 25 frames can be generated. This length exceeds the lengths of most English words. On top of that, the rectangular pooling strides yield rectangular receptive fields (illustrated in Fig.2.0), which are beneficial for recognizing some characters that have narrow shapes, such as 'i' and 'l'.

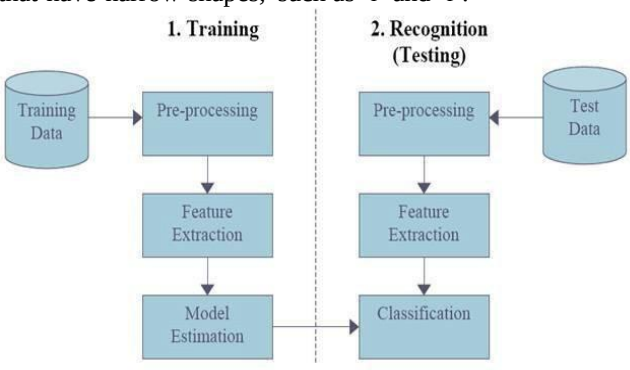

Figure.2. Architecture for training and recognition testing

The network not only has deep convolutional layers, but also deep recurrent layers. Both are known to be hard to train. We find that the batch normalization technique is critical for training a network of such depth. Two batch normalization layers are inserted after the 3rd, 5th, and 7th convolutional layers respectively.

3.1 Comparative Evaluation

All the acknowledgment accurateness on the above four community datasets, gotten by the proposed CRNN model and the recent state-of-the- arts techniques including the approaches based on deep models. In the constrained-lexicon case, our method consistently outperforms most state-of-the-arts approaches, and in aver-age beats the best text reader proposed in. Specifically, we obtain superior performance on IIIT5k, and SVT com-pared to only achieved lower performance on IC03 with the Full lexicon. Note that the model in is trained on a specific wordlist, namely that each word is associated to a class label. Unlike, CRNN is not controlled to establish a word in a known thesaurus, and able to grip casual strings (e.g. telephone numbers), sentences or other writings like Chinese words. Therefore, the outcomes of CRNN are inexpensive on all the challenging datasets.

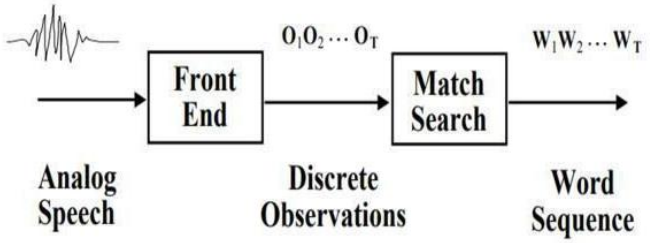

Figure.3. Algorithm for voice to word sequence diagram

Torch7/CUDA), the transcript layer (in C#) and the BK-tree data structure (in C#). Experimentations are carried out on a workspace with a 2.50 GHz Intel(R) Xeon(R) E5-2609 CPU, 64GB RAM and an NVIDIA(R) Tesla(TM) K40 GPU. Networks are qualified with setting the constraint to 0.9. For training efficiency, we first train our model on rescaled training images, whose sizes are 100 32, for about 250k iterations. Then, we continue the training with variable-size images for another 50k iterations. In this stage, we first sort all training images by their aspect ratios. In each iteration, we randomly pick a batch of consecutive images and rescale them to have height 32 and their median aspect ratio. By doing this, we achieve efficient variable-size training, with ineligible distortions on training images. The whole training process takes about 5 days. Tough images are scrabbled to have height 32. Widths are equivalently climbed with elevations, but at least 100 pixels. The unvarying testing time is 0.16s/sample, as unhurried on IC03 deprived of a lexicon. The estimated lexicon search method is applied to the 50k-lexicon of IC03, with the parameter set to 3. Testing each sample takes 0.53s on average. In the unconstrained-lexicon case, our method achieves the best performance on SVT and the second best on IC13. Note that the blanks in the ―none‖ columns of Table 2 denote that such methods are impotent to be practical to recognition without vocabulary or did not report the recognition exactitudes in the unconstrained cases. Particularly, our model outer-forms Photos which used 7.9 million of real word images with character-level annotations for training, while our method uses only synthetic text with word level labels as the training data. Yet, our method is still behind on IC13. Benefits from its large dictionary, however, it is not a model strictly unconstrained to a lexicon as mentioned before.

3.2.Screenshot

Figure.4. Conversion of image to text

Goal:

Given acoustic data $A = a_1, a_2, ..., a_k$

Find word sequence $W = w_1, w_2, ... w_n$

Such that $P(W \mid A)$ **is maximized**

Baves Rule:

$$P(W \mid A) = \frac{P(A \mid W) \cdot P(W)}{P(A)}$$

acoustic model (HMMs) — language model

P(A) is a constant for a complete sentence

For example, simplifying the network by removing the 4th and 6th convolutional layers ends up with 86.5% accuracy on the validation set of Synth, while for the proposed network, the accuracy is 92.3%. Using a network structure similar to the one proposed in results in 89.2% accuracy, also lower. On the other hand, increasing the depth.

3.2.1. Screenshot

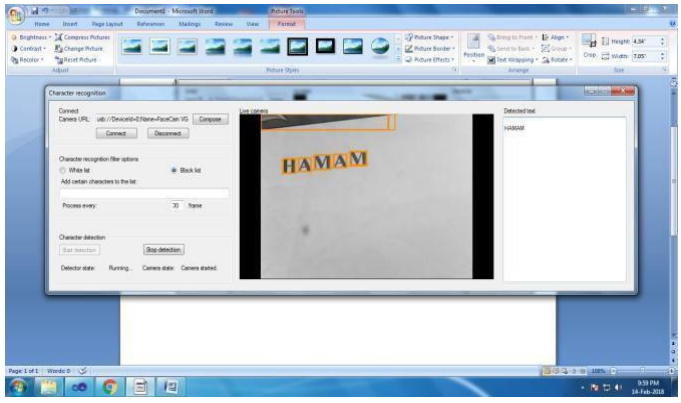

Figure.5. Detecting the text for image

3.3. Classic Extent:

This pilaster is to report the storage space of the academic model. In CRNN, all layers have weight- sharing connections, and the fully-connected layers are not needed. Therefore, the quantity of strictures of CRNN is much less than the models educated on the alternatives of CNN later in a much reduced model compared with clearly displays the differences between different methods in details, and fully establishes the advantages of CRNN over other opposing methods. Several variants of the architecture specified in Table 1 are tested in order to evaluate the impact of each part. We find that the configuration of the convolutional

layers is critical to the overall performance.

4. CONCLUSION AND FEATURE ANALYSIS

In this paper, we have obtainable a novel neural network architecture called―Convolutional Recurrent Neural Network‖ (CRNN), which integrates the advantages of both Deep Convolutional Neural Networks (DCNN) and Recurrent Neural Networks (RNN). CRNN is able totaled input images of varying dimensions and produces predictions with different lengths It directly runs on coarse level labels (e.g. words), demanding no thorough comments for each distinct element(e.g. characters) in the training phase. Moreover, as ―CRNN abandons fully connected layers used in conventional neural networks, it results in a much more compact and efficient model‖. All these things make CRNN an brilliant approach for image based sequence recognition. The experiments on the scene text recognition bench-marks demonstrate that CRNN achieves superioror highly competitive performance, compared with conventional methods as well as other DCNN and RNN based algorithms. This confirms the advantages of the proposed algorithm. In adding,Really, CRNN is a common background, thus it can be applied to other domains and problems (such as Chinese atmosphere recognition), which contain sequence calculation in images A hardware/software system for guiding a visually impaired character via a building to a library to discover book was once placed. We showed an integration of imaginative and prescient-founded with book with author, return details Keep record of altered groups like; Books, Journals, Newspapers, Magazines, etc. Classify the books subject wise. Easy way to enter new books. Preserve record of widespread material of a book like; Book forename, Journalist name, Publisher's name, Date/ Year of publication, Cost of the book, Book buying date/ Bill no. Easy way to make a check-out. Easy way to make a check-in. Automatic fine calculation for late returns. Different criteria for searching a book. Different kind of reports like; total no. of books, no. of issued books, no. of journals, etc. Easy way to distinguish how many books are supplied to a specific student. Easy way to know the status of a book. Experience calendar for librarian to reminisce their dates. My Notes segment for librarian to write any note. Online access for registered user to see the status of their books. Completely cloud based Library Management System. No need to invest heavily on Hardware. SAAS based pricing. and much more.

5.REFERENCES

[1]. R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semanticsegmenta-tion,‖ in Proc. Int. Conf. Comput. Vision Pattern Recog., 2014, pp. 580–587.

[2]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ―Imagenet classifica- tion with deep convolutional neural networks,‖ in Proc. Advances in Neural Info. Processing Sys., 2012, pp. 1106–1114.

[3]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition,‖ Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[4]. T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, ―End-to-end text recognition with convolutional neural networks,‖ in Proc. Int. Conf. Pattern Recog., 2012, pp. 3304–3308.

[5]. A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, Photoocr: Reading text in uncontrolled conditions,‖ in Proc. Int. Conf. Com-put. Vision, 2013, pp. 785–792.

[6]. M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, Read- ing text in the wild with convolutional neural networks,‖ Int. J. Comput. Vision (Accepted), 2015.

[7].A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, ―A novel connectionist system for unconstrained handwriting recognition,‖ IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 5, pp. 855–868, 2009.

[8]. B. Su and S. Lu, ―Accurate scene text recognition based on recur- rent neural network,‖ in Proc. Asian Conf. Comput. Vision, 2014, pp. 35–48

[9]. Zhu, Yingying, Cong Yao, and Xiang Bai. "Scene text detection and recognition: Modern advances and upcoming leanings." Borderlines of Computer Science 10.1 (2016): 19-36.

[10]. Wei, Yuanwang, et al. "Text detection in scene images based on exhaustive segmentation." Signal Processing: Image Communication 50 (2017): 1-8

[11]. C. Yao, X. Bai, B. Shi, and W. Liu, ―Strokelets: A learned multi- scale representation for scene text recognition,‖ in Proc. Int. Conf. Comput. Vision Pattern Recog., 2014, pp. 4042–4049.