IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709

Deep Learning Based Content Retrieval for Recognition and Classification in Historical Document

Abhishek[1], Bharath S[2], Pavan Kumar S[3], Madhu R[4], Shruthi K.R[5]

[1-4]Student, Global Academy of Technology, Bengaluru

[5]Professor, Department of Information Science and Engineering, Global Academy of Technology,Bengaluru

Abstract: Due to the recent rapid expansion in the number of digitised historical files, this is vital. It provides efficient methods for information extraction and statistics retrieval to allow access to data. It makes use of optical character recognition to convert document images into textual representations (OCR). OCR techniques today frequently do not belong in the historical domain. Additionally, they typically need a substantial volume of annotated documents. This paper will therefore show you a few ways to allow OCR on past data.

Authentic, hand-labeled coaching information should be added to the image. OCR with all features OCR and page structure analysis, which comprises text blocking and line segmentation, are the device's two primary functions. While the OCR approach is based on a convolution neural network, our delineation method uses on recurrent neural network. Both approaches are state-of-the-art in the concerned field. developed a novel authentic dataset for the Protonium Portal for OCR.

This information, which is openly available on this corpus, will be used to evaluate all suggested strategies. We illustrate it using some actual examples of annotated data so that both categorization and OCR tasks may be carried out. The experiment seeks to achieve this. If your information is limited, decide how to accomplish it properly in a satisfactory manner. We also demonstrate that the rating we conducted is on par with or superior to the results of certain modern systems. The study's findings demonstrate how to create a successful OCR engine for historical documents even in the lack of substantial training data.

INTRODUCTION

For a very long time, codicology and paper background have considered the identification and recovery of ancient watermarks to be an important research area. When classifying non-dated mediaeval manuscripts, for instance, watermark identification is primarily used to date historical documents. The identification of watermarks can also answer more general research problems, such as those pertaining to economic history. In the area of pattern recognition, there is a lot of research being done on topics like content-based feature extraction (CBIR) and feature spotting (PS). The primary driving force behind this work is the growing need for solutions that can retrieve specific items from images stored in vast digital libraries in the last few decades of current culture.

The requirement to complete the recall task without any prior knowledge of the images or structures to be retrieved has been a fresh and fascinating problem on CBIR and PS. The goal is to create general solutions that can operate on various digital image collections. The complexity of the task depends is significantly raised by this unique challenge. Therefore, a strong solution must take into account not just the standard variations in colour, shape, conservation, richness, and setting but also the absence of prior knowledge regarding potential picture library inquiries.

The definition of a reliable representation for the image nominees and the requests, as well as the interpretation of an adequate parameter capable of estimating the similarity among a given query as well as the available image candidates, are essential for the success of developing such flexible solutions. Deep models, particularly Convolutional Neural Networks (CNNs), provide an appealing alternative to develop a robust representation automatically, avoiding the difficult technical labour associated with defining handmade features (Lecun et al., 2010).

Recent deep criteria have been applied in several applications for the estimation of picture similarity. A deep metric is typically made up of two deep models (CNN-based), which are arranged in a Siamese design, in contrast to traditional architectures. In several applications, such as image representation and motion tracking to determine either two input photos are identical or not, such a deep network has shown good results.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709

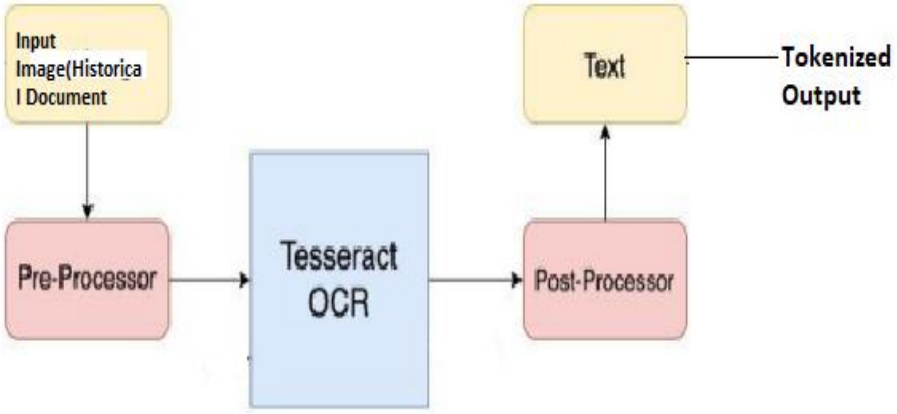


Figure 1: System Model

The workload handled and strategies taken into consideration has significantly increased during the previous several years. Recent applications span innovative subjects including manuscript dating and text localization in maps, while previous studies on DIAR for ancient documents mostly concentrated on script identification and keyword identification. Since the Deep Learning explosion has affected many study topics, including the classification and recognition of ancient writings, various efforts currently focus on the identification of specific text lines. The division of regions in predominantly printed materials was the subject of a great series of projects on this area in layout analysis. Many deep learning techniques are now used, with a focus on architectures built on convolutional layers, while examining this topic from a methodological angle . In particular, a number of models are built to generate an output that is the same size as the input image, enabling pixel labelling that may be applied to semantic segmentation in various situations. Before CNNs, pixel tagging was accomplished via a sliding window that fed the rnn at various points. On the other hand, by training the neural network on entire pages or significant portions of images, the use of fully connected layers enables the achievement of the necessary pixel labelling.

IMPLEMENTATION

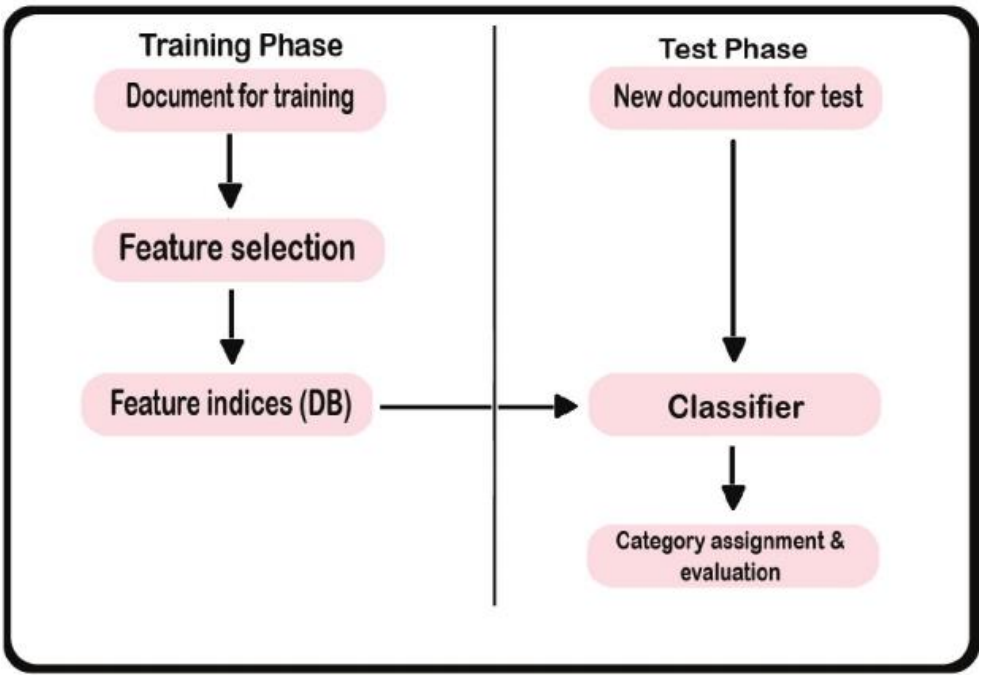


Figure 2: Data Flow Diagram

The basic processes in an image processing-based currency recognition technique are image acquisition, image pre-processing, and currency recognition. Three steps are typically involved in image processing:

- Insert an image using a direct digital camera shot or an optical scanner.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709

- Change or otherwise alter the image.
- Publish the outcome.

The end result could be an edited version of the original image or a report based on image analysis. The following is a flowchart of the methodology's steps:

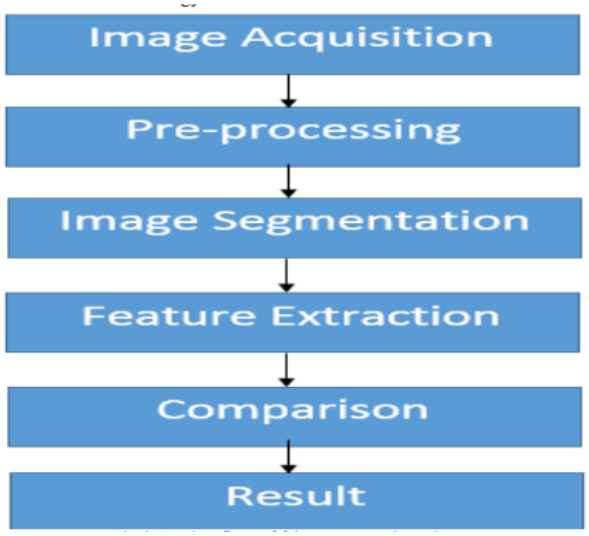


Figure 3: Methodology Steps

Image Acquisition

In general, image acquisition in the context of image processing refers to the action of acquiring a picture from a source—typically one that is hardware-based—so that it can be transmitted through subsequent processing steps. The initial stage in the following conclusions can be drawn for computer vision is always picture capture because processing is impossible without an image. The picture that is captured is entirely unedited and the output of the scanner that produced it, which might be crucial in some industries to have a constant starting point from which to develop. One of its ultimate objectives of this approach is to have an input source that runs within such precise and regulated parameters that the very same vision can, if appropriate, be nearly precisely duplicated under the same circumstances, making it simpler to identify and remove abnormal variables.

Preprocessing:

Pre-processing a picture entail enhancing or enhancing certain aspects of the image that are crucial for later analysis and processing. With the use of a median filter, noise from the split image is removed. The median filter works by sliding a window across the entire frame and determining the resulting number of pixels as the median intensity value in the current window. The channels of the resulting smoothed images are recovered. Another pre-processing technique used was to maintain the same aspect ratio while normalising the size of different currency notes. The size of the note divided by the width of the network is known as the aspect ratio.

Image Segmentation

In an image, it establishes region borders. It can investigate a wide range of image segmentation and thresholding strategies. The best global threshold: 1. When there are the fewest possible misclassified pixels, a threshold is considered to be globally optimum. 2 Bimodal histogram (object and background) 3 EITHER the scatterplots of the item and the environment are known OR the ground truth is known.

Feature Extraction

A method of dimensionality reduction known as feature extraction effectively depicts the visually appealing portions of an image as a short feature set. When a reduced feature set is needed to efficiently fulfil tasks like picture matching and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709

retrieval when image sizes are big, this method is helpful. Among an image's characteristics are: Area or Size. Every denomination has a different size parameter from the next. As a result, size is a property that can be used to identify currencies. The picture fluctuates depending on the angle at which the snapshot of the image was taken, which is the feature's main drawback. A new criteria called aspect ratio was introduced to categories the denominations in order to solve this difficulty.

Comparison

The traits that we collected from the photos of the bank notes are quite important in our comparison. In actuality, comparing the qualities is what allows us to tell false notes from real ones. We have segmented the image to compare performance, and then we create a second binary image by removing any connected components (objects) with fewer than P pixels from the first binary image. To create the binary picture that can be compared, the previous procedure is performed three times. The difference is then stored after we have compared the two photos.

Sequence Diagram

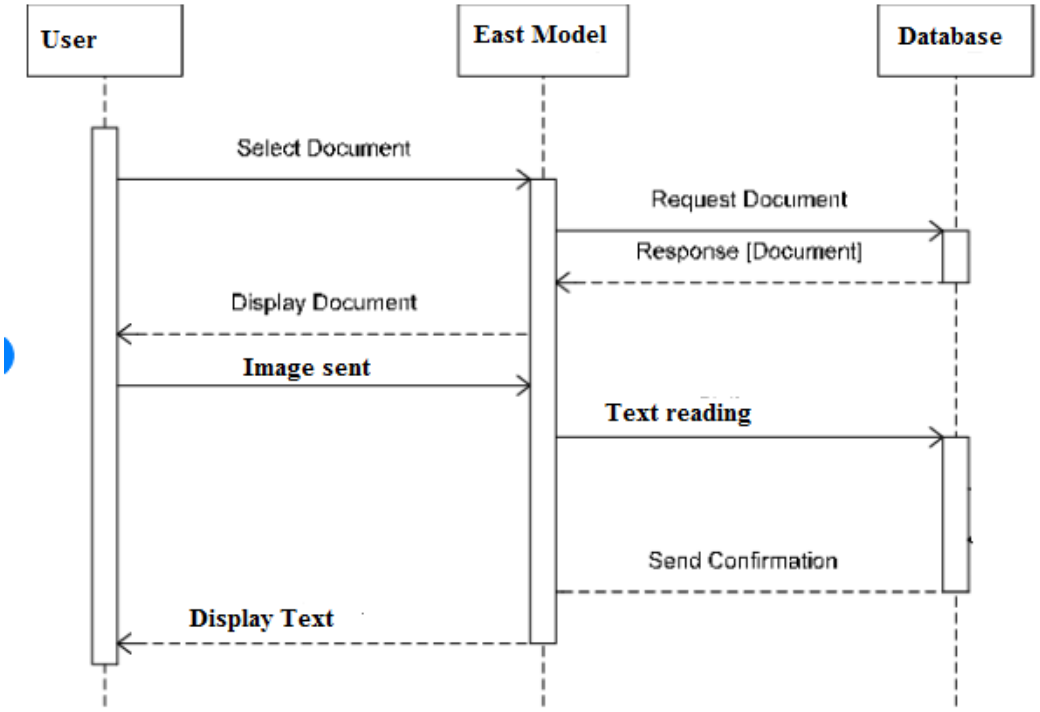


Figure 4: Sequence Diagram

Object interactions are arranged in temporal sequence in a sequence diagram. It shows the classes and objects involved inside the scenario as well as the flow of messages that must be exchanged for the objects to work as intended. In the Functional View of both the system being developed, sequence diagrams are often connected to use case realisations.

EAST (Efficient Accurate Scene Text Detector)

Currently, text identification in natural settings is a prominent topic in research. The Fast and Reliable Scene Text (EAST) detectors model has a narrow receptive field, which prevents it from being successful in identifying large text portions despite its quick detection speed and strong performance. In this study, we extended the EAST model by modifying the bounding box shrinking algorithm to increase the model's prediction accuracy short sides of text regions, changing the loss function from stable cross-entropy to Focal loss, enhancing the model's learning capability on challenging, encouraging examples, and adding a function enhancement device (FEM) to expand the EAST model's receptive field and improve its detection capability for text message regions. The EAST model's total flow is displayed.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709

OpenCV

A set of programming tools called OpenCV is primarily focused on real-time computer vision. Because of its modular design, the package contains a number of shared or static modules. We are employing an image processing module with features like histograms, colour space conversion, affine and aspect warping, geometric picture transformations, linear and non-linear noise removal, and more. Libraries like the Viola-Jones or Hartley classifier, the LBPH face recognition algorithm, and the Histogram of Image Gradients are all included in our project (HOG).

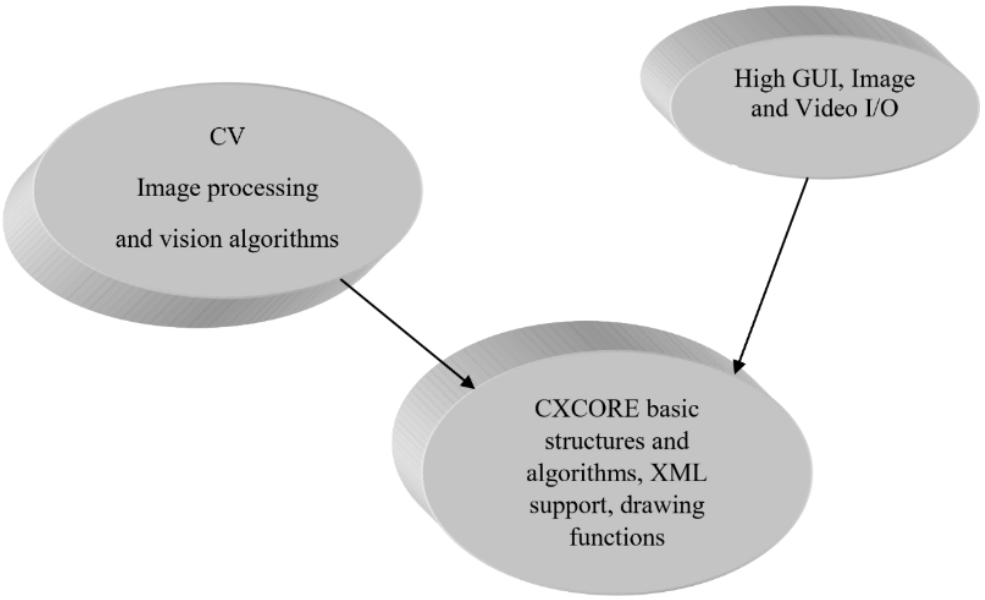


Figure 5: Structure of OpenCV

OCR:

The initial method of character recognition, optical character recognition, frequently has a low recognition rate because to character segmentation errors. Any OCR system must perform segmentation. The image written reports are divided into lines, words, and characters. The algorithm based being employed has a significant impact on the OCR system's accuracy. Given the higher character set and structural complexity of handwritten language traditional print English and any other produced document, segmenting it is challenging. Consonants and vowels are present. There may be some character overlap. Projections profiles form the foundation of the suggested method. According to experimental findings, overlapping lines, words, and characters can still achieve 100%-line segmentation and roughly 98%-character segmentation accuracy.

Image to text conversion

The preceding illustrates how Text-To-Speech works. The OCR and picture pre-processing components make up the first block. The or before image, which is in.png format, is changed into a.txt file. Tesseract OCR is what we're employing.

RESULT

- Recognition of the Document using OCR
- Classifying the Document Image based on

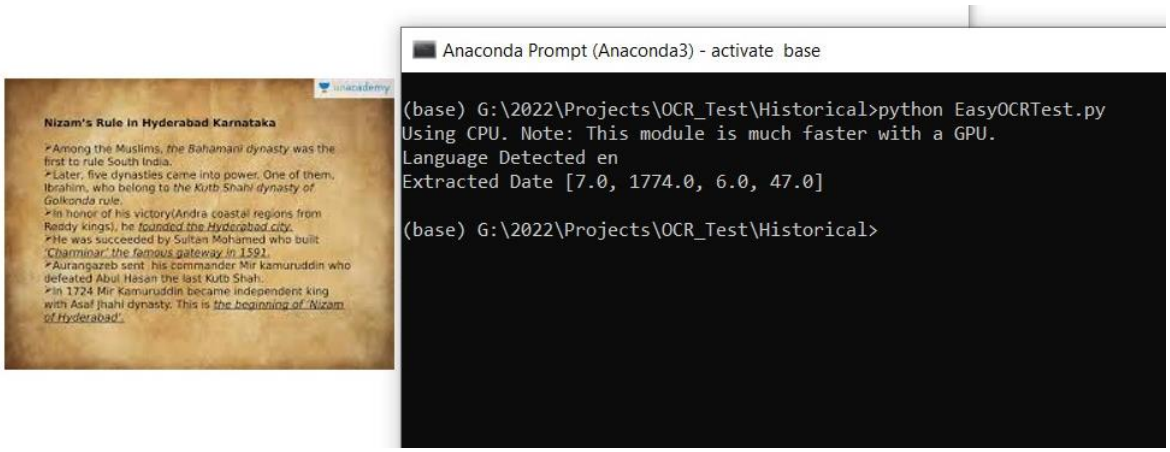
I. Language Detected in the document Image
II. Numbers Detected in the document Image

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11709



CONCLUSION

The optimum of image retrieval is achieved by selecting an appropriate distance measure and producing effective feature descriptors from the numerous features that the OCR has fetched. to clearly separate the classes' intra- and inter-class heterogeneity. It is still difficult to find a means to decrease the dimension into a lower subspace without sacrificing accuracy. The majority of existing content-based image retrieval systems operate on the most fundamental levels of image attributes, such as colour, texture, and shape.

REFERENCES

[1]. Pascanu R, Mikolov T, Bengio Y (2013) On the difficulty of training recurrent neural networks. In: International conference on machine learning, pp 1310–1318

[2]. Marinai, S.; Gori, M.; Soda, G. Artificial Neural Networks for Document Analysis and Recognition. IEEE Trans. Pattern Anal. Mach. Intel. 2005, 27, 23–35. [CrossRef] [PubMed]

[3]. Graves A, Ferna´ndez S, Gomez F, Schmidhuber J (2006) Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd international conference on machine learning (ACM), pp 369–376

[4]. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv: 1712.04621

[5]. Sabir E, Rawls S, Natarajan P (2017) Implicit Language Model in LSTM for OCR. In: 2017 14th IAPR international conference on document analysis and recognition (ICDAR), vol 7. IEEE, pp 27–31

[6]. Karpinski, R.; Belaïd, A. Semi-Synthetic Data Augmentation of Scanned Historical Documents. In Proceedings of the 2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, 20–25 September 2019; pp. 268–273.

[7]. Jaehyun An, Sang Hwa Lee, Nam Ik Cho, "Content based image retrieval using color features of salient regions," International Conference on Image Processing, IEEE, 2014.

[8]. R N Muhammad Ilyas, S Pannirselvam, "An enhanced technique for texture-based image retrieval using framelet transform with GLCM," International Journal of Computer Science and Mobile Computing, IEEE, Vol.6, pp. 150-157, 2017.

[9]. Pushpalatha Srikanth Nikkam, B.Eswara Reddy, "A key point selection shape technique for Content based image retrieval system," International journal of computer vision and Image processing.(IJCVIT),Vol.6, pp.54-70, 2016.

[10]. Nilima R kharsan, Sagar S Badnerke, "A review paper on content-based image retrieval technique using color and texture feature," International journal of engineering trends and technology (IJETT), 2017.

[11]. Zahid Mehmood, Fakhar Abbas, Toqeer Mahmood, Muhammad Arshad Javid, Amjad Rehman, Tabassam Nawaz," Content Based Image Retrieval Based on Visual Words Fusion Versus Feature Fusion of Local and Global Features," Arabian Journal of science and Engineering, Springer, Vol.43, pp.7265- 7284, 2018

[12]. Mahantesh.K.et.al, "A Study of Subspace Mixture Models with Different Classifiers for Very Large Object Classification," In the proceedings of International Conference on Advances in Computing, Communications and Informatics, PRIP-IEEE, India, pp. 540-544, 2014.

[13]. Ruifeng Zhu, Fadi Dornaika, Yassine Ruichek," Joint graph-based embedding and feature weighting for image classification," Elsevier, 2019.

[14]. Lingling Zhang, Jun Liu, Minnan Luo, Xiaojun Chang, Qinghua heng, Alexander G. Hauptmann, "Scheduled sampling for oneshot learning via matching network," Elsevier, 2019.

