**ORIGINAL PAPER**

# Pho(SC)-CTC—a hybrid approach towards zero-shot word image recognition

Ravi Bhatt[1] · Anuj Rai[1] · Sukalpa Chanda[2] · Narayanan C. Krishnan[1]

## Abstract

Annotating words in a historical document image archive for word image recognition purpose demands time and skilled human resource (like historians, paleographers). In a real-life scenario, obtaining sample images for all possible words is also not feasible. However, zero-shot learning methods could aptly be used to recognize unseen/out-of-lexicon words in such historical document images. Based on previous state-of-the-art method for zero-shot word recognition "Pho(SC)Net", we propose a hybrid model based on the CTC framework (Pho(SC)-CTC) that takes advantage of the rich features learned by Pho(SC)Net followed by a "connectionist temporal classification" (CTC) framework to perform the final classification. Encouraging results were obtained on two publicly available historical document datasets and one synthetic handwritten dataset, which justifies the efficacy of Pho(SC)-CTC and Pho(SC)Net.

**Keywords** PHo(SC)Net · CTC · Zero-shot word recognition · Historical documents · Zero-shot learning · Word recognition

## 1 Introduction

Historical documents narrates condition about human societies in the past. Easy availability and usability of image acquisition devices have led to the digitization and archival of such historical documents. Searching for important and relevant information from the large pool of images in those digital archives is a challenging task. Earlier, end-to-end transcription of the text using OCR was a popular way to achieve this goal. However, the performance of OCR often depends on character-segmentation accuracy, which is error prone, spe-

cially in the context of cursive handwritten text. Moreover, end-users of such digital archives (historians, paleographers, etc.) are often not interested in an end-to-end transcription of the text, rather they are interested in specific document pages where a query incident, place name, person name has been mentioned. To cater to this requirement, word-spotting and recognition techniques play an important role as they help directly in document indexing and retrieval.

Deep learning models have been quite successful in many document analysis problems. Thus, it is a natural fit for word recognition in historical documents as well. However, training deep networks for this problem is a challenging task due to many reasons. The lack of a large corpus of word images, partly due to the changes in the appearance of characters and spellings of words over centuries, makes it difficult to train a deep model. The complexity is further compounded by the requirement of learning large number of word labels, using only a small set of samples. In addition, the historical documents exhibit many undesirable characteristics such as torn and blurred segments, unwanted noise and faded regions, handwritten annotations by historians and artefacts; all of them contributing to the difficulty of the task.

This work is focused on work recognition and extends it to the zero-shot learning (ZSL) setting.

Classical word recognition involves training a machine learning model to recognise the words given the images con-

---

Ravi Bhatt and Anuj Rai have contributed equally to this work.

✉ Ravi Bhatt
2020aim1008@iitrpr.ac.in

Anuj Rai
2019aim1003@iitrpr.ac.in

Sukalpa Chanda
sukalpa@ieee.org

Narayanan C. Krishnan
ckn@iitrpr.ac.in

1 Department of Computer Science and Engineering, Indian Institute of Technology Ropar, Rupnagar, Punjab 140001, India

2 Department of Computer Science and Communication, Østfold University College, 1757 Halden, Norway