IBM IEEE CAS/EDS
AI Compute Symposium 2021
October 13-14

A Hybrid Capsule Network-based Deep Learning Architecture for Deciphering Ancient Scripts with Scarce Annotations

Rodrigue Rizk
The Center for Advanced Computer Studies
University of Louisiana at Lafayette

UNIVERSITY of LOUISIANA LAFAYETTE

Introduction

- The advent of writing systems that turn sounds of spoken languages into something visual is considered a major milestone of human civilization. For thousands of years, **epigraphy** has been used as a prime tool in deciphering these writing systems from ancient inscriptions that supply valuable facts about ancient history.

- Most of these scripts have not been completely deciphered yet for many reasons ranging from the intrinsic complicated nature of the writing system and the time-consuming nature of the process to the poor quality of the text due to environmental and physical factors.

- Deciphering ancient scripts is a tedious process that can only be achieved manually by specialists.

- With the advancement in the computing field and in linguistics, it becomes possible for algorithms to begin deciphering ancient languages with the aid of **artificial intelligence**.

- In this work, a novel hybrid **capsule network**-based deep learning architecture for deciphering ancient scripts with scarce annotations is introduced.

UNIVERSITY of LOUISIANA LAFAYETTE

Introduction

- We have used the **Phoenician epigraphy** as a case study because the Phoenician alphabet is considered as the oldest verified alphabet known and most existing alphabets are derived from it which makes them inherit common features that could be leveraged for other character recognition tasks.

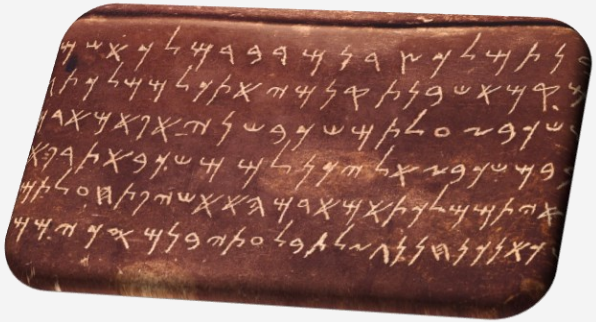
- We also present a **corpus of labeled data** of Phoenician alphabets that covers all different styles and stages. This corpus contributes to the digitization process of the Phoenician culture.

- This work can be replicated for any other ancient scripts with minor modifications considering the systematic methodology that we propose.

- To the best of our knowledge, this work is the first to propose such a framework that helps in overcoming the limitations of current epigraphic techniques.

Phoenician Language

- The Phoenician alphabet is only 22 letters based on sound (acrophony) in contrast to the plethora of symbols in cuneiform and hieroglyphics prevalent at that time.
- It is considered as the oldest verified alphabet known, and the mother of all alphabets.
- Phoenician was written from right to left horizontally, and vowels were omitted with no space or other marks between words and sentences (Scriptio continua), which makes deciphering Phoenician even harder.



Phoenician inscription, (5th century BC).

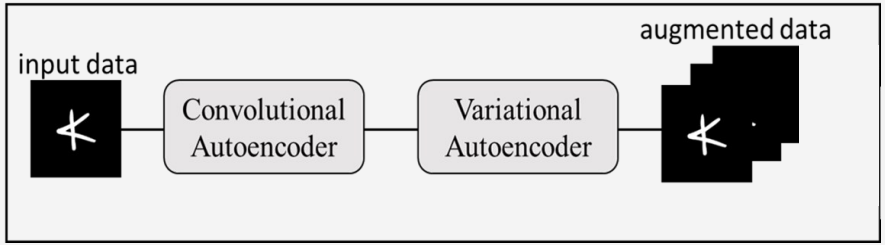
UNIVERSITY of LOUISIANA LAFAYETTE

Capsule Network

- A Capsule network is an emerging artificial neural network characterized by the capacity of preserving spatial information using better model hierarchical relationships.
- It is a variant architecture of convolutional neural networks with core elements called capsules that consist of grouped neurons in which their outputs represent different properties of the same feature.
- Capsule-capsule connections use a dynamic routing algorithm to recognize the part-whole relationships. This routing algorithm manages the way of routing between capsule layers.
- The proposed model uses a capsule network for a more efficient representation of the visuospatial features of characters and because it requires less training data.
- By using capsule networks, position and pose information are preserved (equivariance), which is useful for character segmentation and detection.
- The routing algorithm handles the problem of overlapping characters, which is a common characteristic of dense text structures like in inscriptions.

Data Acquisition and Preprocessing

- After building an image dataset of Phoenician letters with different styles for all the stages and dividing the letters into classes and categories, we perform conventional pre-processing and denoising techniques such as noise filtering and image binarization.
- A convolutional autoencoder is also used for denoising the image samples.
- To build an augmented dataset of images' characters that resemble human handwritten style, we build a generative model that consists of a variational autoencoder (VAE) in which it learns the latent variable model for our characters' images.

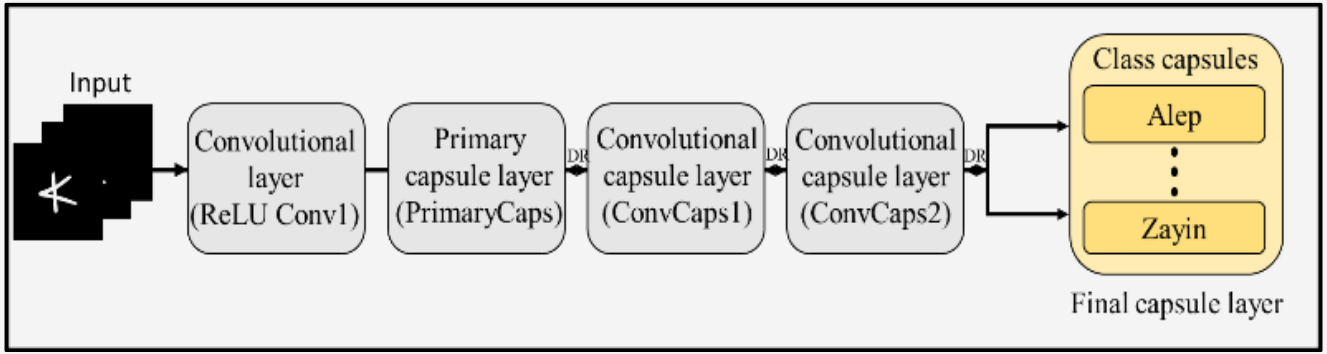


Hybrid autoencoder-based preprocessing architecture.

A sample of images from our dataset for the Phoenician letter A for different stages.

Proposed Capsule Network-based Architecture

- Since our dataset consists of characters' images, we build a capsule network that preserves the visuospatial features' information with less amount of data.
- It consists of capsules that perform computation of the inputs and encapsulate the results with the information of the character into a vector in order to decipher the Phoenician character and classify it into one of the 22 letters.
- The architecture of the capsule network consists of a set of three layers: a convolutional layer, primary capsule layer (PrimaryCaps), and a class capsule layer (ClassCaps).



Proposed capsule network-based architecture.

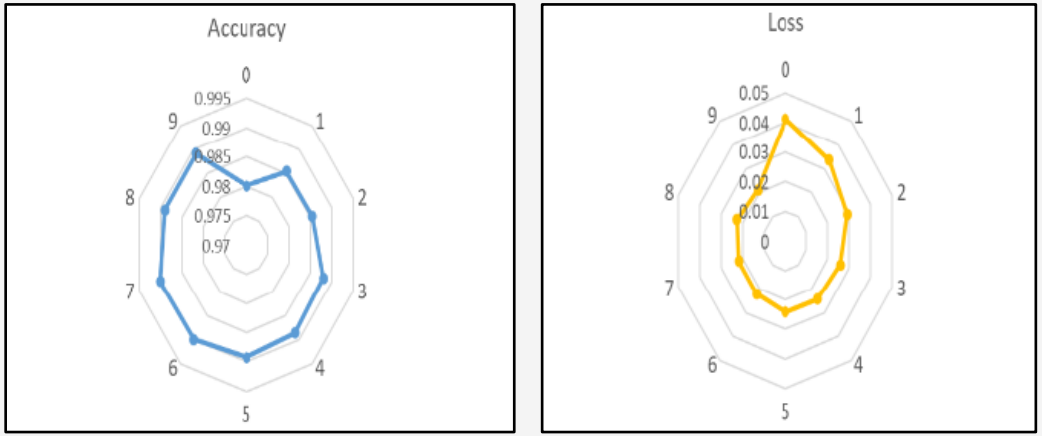
UNIVERSITY of LOUISIANA LAFAYETTE

Proposed Capsule Network-based Architecture

- For each layer, the vector representation of images' characters of the previous layer is multiplied by the weight matrices that capture the visuospatial features of the character and its different writing styles.
- The resulting outputs are then multiplied by the network weights that are assigned by the dynamic routing (DR).
- This routing algorithm manages the way of routing between capsule layers in which the lower-level capsule will send its input to the higher-level capsule that agrees with its input.
- The weighted sum of the inputs vectors is then obtained by summing up all the resulting outputs. Finally, the output is obtained by applying a non-linear activation function (Squash).
- The ClassCaps layer contains one capsule per class of character resulting in 22 classes that represent our 22 characters.

Experimental Setup and Results

- We implement the model in Python using the deep learning framework Keras with TensorFlow backend.
- Training is conducted with early stopping callback using the adaptive learning rate optimization algorithm ADAM.
- Our model achieves a **high accuracy** of **0.9891** and a **loss** of **0.021**.



Radar plots of the accuracy and the loss for different epochs.

UNIVERSITY of LOUISIANA LAFAYETTE

Conclusion

- This work presents a hybrid capsule network-based deep learning architecture for deciphering ancient scripts with scarce annotations with a case study on the Phoenician epigraphy.
- We present a corpus of labeled data of Phoenician alphabets that covers all different styles and stages. This corpus contributes to the digitization process of the Phoenician culture.
- Our model achieves an overall accuracy of 0.9891 and a loss of 0.021 in recognizing ancient scripts characters.
- This work can be replicated for any other ancient scripts with minor modifications considering the systematic methodology that we proposed. It can help develop an automated deciphering system to save epigraphists' valuable time and effort in deciphering the Phoenician epigraphy in a short period.
- Since the Phoenician alphabet is considered as the mother of all existing alphabets, this makes most of them to inherit common features that could be leveraged for other character recognition tasks. It can be employed as a transfer learning backbone for recognizing other existing alphabets which suffer from a lack of annotated data.