Modelling fonts with convolutional neural networks

Philip Bouman
10668667

***Abstract*—**

***Index Terms*—**

OVERVIEW

The goal of the project is to find relations between characters of different types of writing systems (scripts). In the first part of the research a deep convolutional neural network will be trained to classify (images of) characters. A deep convolutional neural network is a machine learning architecture consisting of (multiple) stacks of layers. Depending on the availability of (training)data the training/learning process can either be supervised, using phonemes to label characters or unsupervised, using just the characters. The second part of the project will focus on getting a better understanding of the representations learned by the network. This will mainly be done by investigating and visualizing features learned in individual layers and smaller combinations of layers. The results can be compared to existing image recognition networks or between methods (for example if supervised and unsupervised are both used).

LITERATURE REVIEW

The relation between word forms and their meanings in natural language is commonly referred to as sound symbolism. A well studied phenomenon of sound symbolism is the *bouba-kiki* effect, in which participants label abstract shapes with non-word names. Rounded shapes are systematically labelled bouba over kiki, this effect is often ascribed to cross-sensory mappings between acoustic properties of sound and shape. Cuskley et al. [1] suggest that this phenomenon is also heavily mediated by the symbolic, culturally acquired shapes of letters. This similarity between orthography and abstract shapes forms the basis of this research, in which the aim is to find if this relation can also be found between graphemes and phonemes. Graphemes represent the smallest individual characters of a writing system and phonemes relate to the sound of particular graphemes [2]. Phonemes represent a standardized collection of spoken language and are collected in the International Phonetic Alphabet (IPA) [3]. How these IPA symbols represent the exact pronunciation of a phoneme can be found in a phonological encoding scheme[1]. By using machine learning to classify graphemes to phonemes, a wide variety of writing systems[2] can be explored. These include Alphabets, Abjads (consonant alphabets) and Abugidas (symbols consisting of compositions of consonant and vowel). The desired machine learnng architecture will be a convolutional neural network (CNN), since CNNs are among the most suitable architectures for handwritten digit and character recognition [4] and not yet used for this specific problem. Neural networks learn intermediate representations of their inputs, which can be useful for subsequent classification tasks. The structure of CNNs, mimicking receptive fields, is particularly useful for handling data that can be represented as two- or higher-dimensional input, such as images [5]. Deep CNNs also incorporate many layers and thus many intermediate representations, while keeping the number of free parameters small [6]. Recent work with CNNs on recognition of natural images and 3D objects resulted in higher accuracies than previously achieved by different methods. CNNs seem to work well for supervised and unsupervised learning tasks and can be succesfull using relatively small datasets [4].

RESEARCH QUESTION

Can a CNN succefully be used to classify graphemes of different writing systems to related general phonemes?

METHOD AND APPROACH

The first step of the project is collecting more data and converting it to the proper file format. The current dataset needs to be extended with more writing systems, as to have enough data to train the CNN. The writing systems used in the dataset are gathered from Omniglot[3]. The letters and corresponding IPA symbols are then extracted as Unicode sequences. With the Python Image Library (PILLOW) the letters are drawn, this allows for easy scalability of the size of the images. Google NOTO fonts[4] will be used as the default font, as it supports many different scripts and uses a consistent layout throughout them. The data will be split up into a training, validation and test set. For creating the architecture of the CNN an existing machinle learning API will be used. This is most likely gonna be Keras, which is a high-level overlay API for either the TensorFlow or Theano library. Keras allows for fast prototyping, modularity and extensibility especially for CNNs. Next the architecture needs to be trained on part of the data and validated on another part of the data. In turn this will lead to a recurrent process of altering the architecture, tweaking the parameters and training again to achieve desired results.

[1] Appendix encoding scheme
[2] Appendix list of writing systems
[3] http://www.omniglot.com/charts/#xls
[4] https://www.google.com/get/noto/

Evaluation

The evaluation will be carried out by measuring the accuracy of the predicted phonemes for previously unseen graphemes using the test set. Depending on whether actual phoneme-grapheme relations exist, evalutation can be done on individual graphemes, simplified graphemes or grapheme clusters. Another part of the evaluation will consist of investigating the feature representations learned by the network. By visualising these representations they can be compared to other research concerning shape-sound relations.

I. Approach

A. Data

B. Design

C. Implementation

II. Experiments and Results

III. Discussion

IV. Conclusion

V. Appendix

References

[1] Christine Cuskley, Julia Simner, and Simon Kirby. Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research*, 81(1):119–130, 2017.

[2] Mark S Seidenberg. Beyond orthographic depth in reading: Equitable division of labor. *Advances in psychology*, 94:85–118, 1992.

[3] International Phonetic Association. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[4] Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, and Jurgen Schmidhuber. Convolutional neural network committees for handwritten character classification. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 1135–1139. IEEE, 2011.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[6] Théodore Bluche, Hermann Ney, and Christopher Kermorvant. Feature extraction with convolutional neural networks for handwritten word recognition. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 285–289. IEEE, 2013.