Logodetect: One-shot detection of logos in image and video data

Jorge Davila-Chacon[1] and Max Pumperla[2, 3]

1 Heldenkombinat Technologies GmbH **2** IUBH Internationale Hochschule **3** Pathmind Inc.

Summary

`Logodetect` allows the detection of logos in images and videos after being trained with one example of the required logo. One-shot learning systems are not only great because they require little data, but also because they practically remove the need for specialists whenever you want to extend or re-purpose their functionality. This means that the benefits are manifold: the user requires little effor to collect and label training data, there is practically no time or economic costs for the training procedure, and it also provides a strategic benefit for business as they become self-sufficient for a large part of the system maintenance.

Statement of need

There is plenty of literature on the use of deep-learning for detecting logos, so, additionally to sharing with the community a couple of algorithms to get started with one-shot logo-detection, the aim of `Logodetect` is to provide a flexible architecture to facilitate the comparison of different algorithms for one-shot object recognition.

The inference pipeline of `Logodetect` supports architectures with one or two stages. **One-stage** architectures directly perform *object-recognition* on the raw images, and **two-stage** architectures first perform *object-detection* and then *object-recognition* in a second stage. The idea is that you can use a generic detector for a single class of objects (e.g. logos, traffic signs or faces) and then compare each of its detections with the exemplar, i.e., the sub-class that you are trying to recognize, to determine if both belong to the same sub-class (e.g. a concrete brand, a stop sign or the face of a loved one). To get started, we include two algorithms that you can play with. Both have a Faster-RCNN (Ren et al., 2015) in the first stage that performs object-detection and they differ in the second stage that performs object-recognition.

Architecture

As a baseline, we bring the exemplars and the detections from the first stage to the same latent space (this reduces the course of dimensionality) and then simply measure the Euclidian or the cosine distance between both embeddings for object-recognition. Both inputs are considered to belong to the same sub-class if their distance is below a threshold determined in a preliminary analysis of the training dataset. The code also provides functionality to add various transformations, so you have the option to augment each exemplar with different transformations if you want. Simply add one or more exemplars into the `data/exemplars` folder that is generated after you've followed the installation instructions below, and you are good to go.

As a first reference against the baseline, we also provide a modified ResNet (He et al., 2016) for object-recognition that directly takes the exemplars and the detections from the first stage and predicts if both belong to the same sub-class. Similarly to (Hsieh et al., 2019), this

network infers a distance metric after being trained with examples of different sub-classes, but instead of sharing the same weights and processing each input in a separate pass as in (Koch et al., 2015), it concatenates both inputs and processes them in one pass. This concept follows more closely the architecture proposed by (Bhunia et al., 2019), where the assumption is that the exemplars often have more high-frequency components than the detections, and therefore the model can increase its accuracy by learning a separate set of weights for each input.

Acknowledgements

References

Bhunia, A. K., Bhunia, A. K., Ghose, S., Das, A., Roy, P. P., & Pal, U. (2019). A deep one-shot network for query-based logo retrieval. *Pattern Recognition*, *96*, 106965. https://doi.org/https://doi.org/10.1016/j.patcog.2019.106965

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hsieh, T.-I., Lo, Y.-C., Chen, H.-T., & Liu, T.-L. (2019). One-shot object detection with co-attention and co-excitation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 32, pp. 2725–2734). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/92af93f73faf3cefc129b6bc55a748a9-Paper.pdf

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). *Siamese neural networks for one-shot image recognition*.

NullConvergence. (2018). *Mtorch*. Open-source software under MIT License. https://github.com/NullConvergence/mtorch

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-CNN: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28, pp. 91–99). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf