Scaling Bio-Inspired Neural Features to Real-World Image Retrieval Problems

Gabriele Lagani[1]

[1]*University of Pisa, 56127, Pisa, Italy*

Abstract

In the last decade, approaches in feature extraction for content-based multimedia retrieval exploited neural feature representations to describe complex data types such as images. In particular, recent approaches proposed to leverage bio-inspired learning solutions, which have the advantage to offer better generalization from fewer training samples. However, scaling these solutions to real-world datasets is a challenging problem. In my recent research, I proposed a possible approach to achieve such scalability, based on translating bio-inspired learning models into matrix multiplications, which can efficiently be executed on GPU. In this way, for the first time, I was able to validate bio-inspired methodologies on large-scale datasets such as ImageNet.

Keywords

Hebbian Learning, Deep Learning, Bio-Inspired, Neural Networks

1. Introduction

Interpreting and retrieving multimedia data is difficult due to the high level of semantic abstraction with which information is represented. Deep Learning (DL) provides a valid aid to handle this type of information. For example, in the context of image data, Deep Neural Networks (DNNs), trained on supervised object recognition tasks, can provide highly abstract feature representations, which are useful for indexing and Content-Based Image Retrieval (CBIR) [1, 2, 3].

A pitfall of supervised training is the requirement of large amounts of labeled data, which are typically difficult to gather, as opposed to unlabeled data. In this light, previous work focused on *semi-supervised* learning approaches, by which labeled data could be enhanced by large amount of unlabeled data, in order to learn higher quality feature extractors with fewer labels [4, 5, 6, 7].

Recently, it was shown that bio-inspired unsupervised learning solutions based on the *Hebbian* principle [8, 9, 10] were able to achieve higher results than traditional counterparts such as Variational Auto-Encoder (VAE) based pre-training [11, 6], especially in learning regimes with very scarce label availability. Hebbian learning rules provide local mechanisms for synaptic adaptation, which are connected to data analysis operations such as clustering or principal component analysis [12], that enable training layers independently from the next ones, allowing a more effective exploitation of the available training set information. However, the application

CEUR Workshop Proceedings (CEUR-WS.org)

of such approaches has remained limited to simple datasets like CIFAR-10 and CIFAR-100 [13].

As part of my PhD work, I aimed at extending the analysis of such methodologies also to more realistic datasets such as ImageNet [14]. This is a collection of over 1.2 million real-world images crawled from the web, categorized into 1000 distinct classes. The problem of Hebbian approaches in such scenarios is that they do not scale well with the complexity of the problem at hand. To overcome this limitation, I developed *FastHebb* [15], an efficient solution for Hebbian feature learning and extraction that is based on re-expressing bio-inspired Hebbian synaptic equations in terms of matrix multiplications, which can be executed very efficiently by leveraging GPU acceleration. This solution permitted to achieve a speedup of Hebbian training up to 50x compared to other solutions, while producing feature representations that proved to be very effective in the context of neural feature-based CBIR, when validated on various benchmarks, including ImageNet.

The remainder of this document gives an overview of the developments mentioned above, organized according to the following structure: Section 2 discusses some related and background material; Section 3 describes more in detail bio-inspired Hebbian feature learning for retrieval, and, in particular, the FastHebb solution; Section 4 provides some empirical results to validate the proposed approaches; Finally, Section 5 presents some concluding remarks.

2. Related Work

Multimedia content retrieval, and, in particular, CBIR, has observed great benefits from transitioning from handcrafted feature representations learned ones. This is due to the *semantic* gap between these types of features: learned representations can encode highly abstract concepts [1]. Therefore, previous work has shown experimental evidence for the effectivenes of such features in image retrieval tasks [16, 17, 18, 19]. While these methods use features obtained from DNNs pre-trained on image classification tasks, in [20] the authors proposed an end-to-end training procedure specifically designed for CBIR. They used a siamese architecture with a triplet loss that pushes related images close in feature space (according to a given ground-truth), while pushing unrelated images away. Finally, Bai et al. [2] presented a comprehensive experimental comparison of various methods on modern computer vision datasets, including their proposed Optimized AlexNet for Image Retrieval (OANIR) approach, in which they applied an AlexNet-inspired [21] network architecture specifically modified and optimized for the retrieval task.

One of the challenges of retrieval tasks is that datasets have typically a very large scale. Manually providing ground-truth labels for training might become very expensive in these cases, suggesting that semi-supervised learning techniques might be exploited in these cases [5, 4, 6, 7]. Recently, bio-inspired *Hebbian* learning approaches have shown great promises for unsupervised [22, 23, 24, 25, 26, 12] semi-supervised learning [8, 9, 10], but their capability to scale to large datasets has remained limited.

In this contribution, I describe a recent solution, named *FastHebb*, that I developed as part of my PhD, which enabled scaling Hebbian-based solutions for image recognition and retrieval also to real-world and large scale datasets such as ImageNet[14].
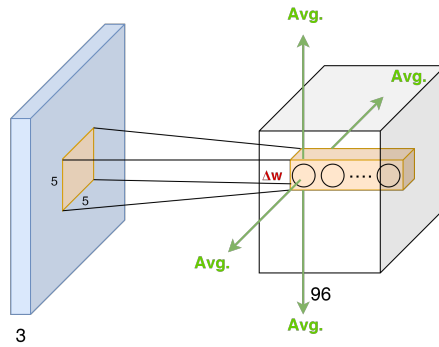
Figure 1: Weight update aggregation mechanism in convolutional layers. Updates are aggregated by averaging along horizontal and vertical dimensions.

3. Scalable Neural Features from Bio-Inpired Learning: FastHebb

I have explored different types of bio-inspired learning rules, but in the following I will focus on two types in particular: Hebbian Principal Component Analysis (HPCA) and Soft-Winner-Takes-All (STWA). I will not delve deep into the theoretical details of these approaches, but the interested reader can refer to [27, 26, 12].

These learning rules follow a *local* learning scheme, by which a neuron updates its weights based on information that is available at the neuron site, according to biological constraints. This is opposed to traditional *backprop* learning where an end-to-end error delivery mechanism takes place. However, when dealing with convolutional network layers, neurons at different horizontal/vertical offsets need to maintain shared weights, so they must undergo the same weight updates. This is achieved by aggregating weight updates at different offsets through a (weighted) averaging mechanism (Fig. 1).

This subdivision into phases of update computation and aggregation results in slow processing when dealing with large data streams. In order to overcome this limitation, the ideas of *FastHebb* are twofold: first, update computation and aggregation phases are merged in a unique phase; second, Hebbian update computations are translated into matrix multiplications. The resulting computation can leverage GPU processing much more efficiently, enabling Hebbian-based approaches at a much larger scale. The details can be found in [15].

4. Experiments and Results

This Section presents an experimental validation of the FastHebb method in the context of CBIR. In order to show the scalability of the FastHebb approach to complex scenarios, the ImageNet ILSVRC2021 [14] benchmark was used for evaluation. To evaluate the approaches in conditions of label scarcity, I considered various *sample efficiency* regimes, i.e. scenarios in which we assume that only a given percentage of the available training samples is labeled. We adopted a semi-supervised training protocol where a network model (shown in Fig. 2) was pre-trained using Hebbian methods on all the available training samples, and then fine-tuned using supervised backprop training on the labeled samples only. As a baseline for comparison,
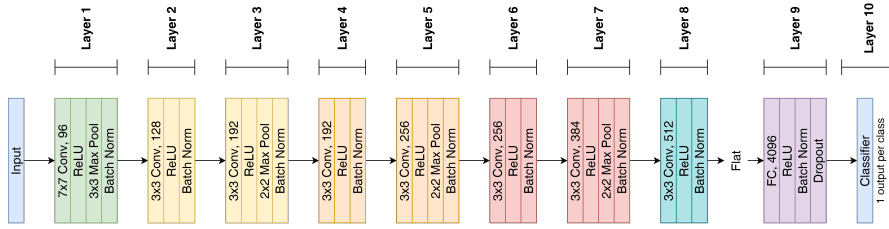
Figure 2: Network architecture used for the FastHebb experiments on the ImageNet benchmark.

Table 1

Analysis of algorithm performance for unsupervised pre-training on ImageNet, comparing VAE, Hebbian PCA (HPCA), Hebbian PCA with FastHebb (HPCA-FH), soft-WTA (SWTA), and soft-WTA with FastHebb (SWTA-FH) methods.

Dataset	Method	Epoch Duration	Num. Epochs	Total Duration
ImageNet	VAE	2h 59m 19s	16	47h 49m 4s
	SWTA	105h 13m 24s	3	315h 40m 12s
	SWTA-FH	3h 38m 6s	3	**10h 54m 18s**
	HPCA	155h 41m 39s	3	467h 4m 57s
	HPCA-FH	3h 39m 18s	3	10h 57m 54s

we considered Variational Auto-Encoder (VAE) based semi-supervised training [6].

In order to assess the improvements in scaling properties of FastHebb, compared to traditional Hebbian learning, we measured duration and total number of training epochs required by each approach, for the unsupervised pre-training phase (while the successive supervised fine tuning phase is comparable across methods). The number of epochs is counted as the epoch at which weights stop changing. The results, shown in Tab. 1 , demonstrate significant improvements (up to 50x) due to the FastHebb solution. Moreover, when the number of epochs is also considered, Hebbian pre-training results to be faster than the backprop-based VAE counterpart.

Trained networks were then used to extract feature representations from the last hidden layer (before the final classifier), to be used for retrieval. The neural feature based retrieval process works as follows: given a query image, taken from the test set, its feature representation is computed by feeding it to the network model. The feature representation of dataset images is also pre-computed in the same way, and stored for indexing and retrieval purposes. Given the query feature representation, we search for the closest dataset elements in feature space, and rank them according to Euclidean distance. Finally, we evaluate the mean Average Precision (mAP) over all the queries, where a retrieved image is considered a positive if its label, according to the ground-truth, corresponds to the query label.

Tab. 2 shows the retrieval results, in terms of mAP, achieved with the different methods, in various regimes of sample efficiency. It should be noticed that the results without FastHebb are not reported in this Table, because they coincide with the FastHebb results. Indeed, FastHebb is a reformulation of weight updates that improves the computational efficiency, but it does not change the synaptic dynamics itself. From the results we can see that Hebbian approaches exhibit better results than the backprop-based VAE method, especially in regimes with fewer

Table 2
Retrieval results on ImageNet, comparing VAE-based and FastHebb-based (HPCA-FH and SWTA-FH) semi-supervised methods.

Regime	Method	mAP
1%	VAE	0.95
	SWTA-FH	1.94
	HPCA-FH	**4.88**
2%	VAE	1.39
	SWTA-FH	2.54
	HPCA-FH	**5.72**
3%	VAE	2.02
	SWTA-FH	3.21
	HPCA-FH	**6.56**
4%	VAE	3.83
	SWTA-FH	3.77
	HPCA-FH	**7.19**
5%	VAE	5.49
	SWTA-FH	4.19
	HPCA-FH	**7.75**
10%	VAE	**13.27**
	SWTA-FH	9.14
	HPCA-FH	10.56
25%	VAE	**24.50**
	SWTA-FH	20.75
	HPCA-FH	21.24
100%	VAE	**44.21**
	SWTA-FH	39.60
	HPCA-FH	43.81

labels (below 10%). The latter is able to improve only when a larger number of labels is available for the fine-tuning phase.

5. Conclusions

The problem of sample efficient training for DNNs is of strong practical interest, due to the difficulty to gather labeled training samples. In this contribution, I discussed some semi-supervised training approaches based on bio-inspired Hebbian learning methods, which are promising in these scenarios. Scaling these approaches to real-world CBIR settings, which are well captured by benchmarks such as ImageNet, is a significant challenge, that I proposed to address through the FastHebb solution. Results showed a significant performance increase during training thanks to FastHebb. Moreover, evaluation of Hebbian neural features in retrieval settings showed promising results, especially in scenarios of label scarcity.

As possible future work directions, I suggest to explore further Hebbian rules which can be used for feature extraction and unsupervised pre-training, for example derived from Independent Component Analysis (ICA) [28]. Finally, in the context of semi-supervised learning, Hebbian

approaches can also be combined with pseudo-labeling and consistency-based methods [29, 30].

Acknowledgments

References

[1] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, J. Li, Deep learning for content-based image retrieval: A comprehensive study, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 157–166.

[2] C. Bai, L. Huang, X. Pan, J. Zheng, S. Chen, Optimization of deep convolutional neural network for large scale image retrieval, Neurocomputing 303 (2018) 60–67.

[3] G. Amato, F. Falchi, C. Gennaro, F. Rabitti, Yfcc100m-hnfc6: A large-scale deep features benchmark for similarity search, in: L. Amsaleg, M. E. Houle, E. Schubert (Eds.), Similarity Search and Applications, Springer International Publishing, Cham, 2016, pp. 196–209.

[4] H. Larochelle, Y. Bengio, J. Louradour, P. Lamblin, Exploring strategies for training deep neural networks., Journal of machine learning research 10 (2009).

[5] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, in: Advances in neural information processing systems, 2007, pp. 153–160.

[6] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, M. Welling, Semi-supervised learning with deep generative models, Advances in neural information processing systems 27 (2014) 3581–3589.

[7] Y. Zhang, K. Lee, H. Lee, Augmenting supervised neural networks with unsupervised objectives for large-scale image classification, in: International conference on machine learning, 2016, pp. 612–621.

[8] G. Lagani, F. Falchi, C. Gennaro, G. Amato, Hebbian semi-supervised learning in a sample efficiency setting, Neural Networks 143 (2021) 719–731.

[9] G. Lagani, F. Falchi, C. Gennaro, G. Amato, Evaluating hebbian learning in a semi-supervised setting, in: International Conference on Machine Learning, Optimization, and Data Science, Springer, 2021, pp. 365–379.

[10] G. Lagani, D. Bacciu, C. Gallicchio, F. Falchi, C. Gennaro, G. Amato, Deep features for cbir with scarce data using hebbian learning, arXiv preprint arXiv:2205.08935 (2022).

[11] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[12] G. Lagani, F. Falchi, C. Gennaro, G. Amato, Comparing the performance of hebbian against backpropagation learning using convolutional neural networks, Neural Computing and Applications 34 (2022) 6503–6519.

[13] A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images, 2009.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, Ieee, 2009, pp. 248–255.

[15] G. Lagani, C. Gennaro, H. Fassold, G. Amato, Fasthebb: Scaling hebbian training of deep neural networks to imagenet level, in: International Conference on Similarity Search and Applications, Springer, 2022, pp. 251–264.

[16] A. Babenko, A. Slesarev, A. Chigorin, V. Lempitsky, Neural codes for image retrieval, in: European conference on computer vision, Springer, 2014, pp. 584–599.

[17] A. Babenko, V. Lempitsky, Aggregating deep convolutional features for image retrieval, arXiv preprint arXiv:1510.07493 (2015).

[18] J. Yue-Hei Ng, F. Yang, L. S. Davis, Exploiting local features from deep networks for image retrieval, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 53–61.

[19] A. Gordo, J. Almazán, J. Revaud, D. Larlus, Deep image retrieval: Learning global representations for image search, in: European conference on computer vision, Springer, 2016, pp. 241–257.

[20] A. Gordo, J. Almazan, J. Revaud, D. Larlus, End-to-end learning of deep visual representations for image retrieval, International Journal of Computer Vision 124 (2017) 237–254.

[21] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems 25 (2012) 1097–1105.

[22] A. Wadhwa, U. Madhow, Bottom-up deep learning using the hebbian principle, 2016.

[23] Y. Bahroun, A. Soltoggio, Online representation learning with single and multi-layer hebbian networks for image classification, in: International Conference on Artificial Neural Networks, Springer, 2017, pp. 354–363.

[24] T. Miconi, Multi-layer hebbian networks with modern deep learning frameworks, arXiv preprint arXiv:2107.01729 (2021).

[25] M. Gupta, S. K. Modi, H. Zhang, J. H. Lee, J. H. Lim, Is bio-inspired learning better than backprop? benchmarking bio learning vs. backprop, arXiv preprint arXiv:2212.04614 (2022).

[26] G. Lagani, F. Falchi, C. Gennaro, G. Amato, Training convolutional neural networks with competitive hebbian learning approaches, in: International Conference on Machine Learning, Optimization, and Data Science, Springer, 2021, pp. 25–40.

[27] G. Amato, F. Carrara, F. Falchi, C. Gennaro, G. Lagani, Hebbian learning meets deep convolutional neural networks, in: International Conference on Image Analysis and Processing, Springer, 2019, pp. 324–334.

[28] A. Hyvarinen, J. Karhunen, E. Oja, Independent component analysis, Studies in informatics and control 11 (2002) 205–207.

[29] A. Iscen, G. Tolias, Y. Avrithis, O. Chum, Label propagation for deep semi-supervised learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5070–5079.

[30] P. Sellars, A. I. Aviles-Rivero, C.-B. Schönlieb, Laplacenet: A hybrid energy-neural model for deep semi-supervised classification, arXiv preprint arXiv:2106.04527 (2021).