ICONIP 2017

A preliminary approach to semi-supervised learning in convolutional neural networks applying "sleep-wake" cycles

Mikel Elkano[1,2], Humberto Bustince[1,2], and Andrew Paplinski[3]

[1] Department of Automatics and Computation, Public University of Navarre, Campus Arrosadia s/n 31006, Pamplona, Spain
[2] Institute of Smart Cities, Public University of Navarre, Campus Arrosadia s/n 31006, Pamplona, Spain
{mikel.elkano, bustince}@unavarra.es
[3] Monash University, 25 Exhibition Walk, Clayton 3800, Melbourne, Australia
andrew.paplinski@monash.edu

Abstract. The scarcity of labeled data has limited the capacity of convolutional neural networks (CNNs) until not long ago and still represents a serious problem in a number of image processing applications. Unsupervised methods have been shown to perform well in feature extraction and clustering tasks, but further investigation on unsupervised solutions for CNNs is needed. In this work, we propose a bio-inspired methodology that applies a deep generative model to help the CNN take advantage of unlabeled data and improve its classification performance. Inspired by the human "sleep-wake cycles", the proposed method divides the learning process into sleep and waking periods. During the waking period, both the generative model and the CNN learn from real training data simultaneously. When sleep begins, none of the networks receive real data and the generative model creates a synthetic dataset from which the CNN learns. The experimental results showed that the generative model was able to teach the CNN and improve its classification performance.

Keywords: Semi-supervised learning, Sleep-wake cycles, Variational autoencoders, Convolutional neural networks, Generative models, Deep learning

1 Introduction

Deep learning has revolutionized the field of machine learning in the last decade. Among existing techniques, convolutional neural networks (CNNs)[8] have been shown to be the best performing approach in image processing [4,7]. These networks are based on bio-inspired architectures that capture important characteristics of the mammalian visual system, such as hierarchical organization and receptive fields. However, CNNs requires vast amounts of training data due to the large number of model parameters that need to be adjusted. Moreover, this drawback is even more accentuated by the scarcity of labeled data. Although a number of unsupervised solutions have been proposed to alleviate this problem [11,12,14], further research is needed to improve supervised classification performance using unlabeled data [12,14].

In this work, we propose the synergy of deep generative models and CNNs to improve supervised learning using unsupervised techniques. This combination has been motivated by the so-called *sleep-wake cycles* and the interaction between the hippocampus and the neocortex that takes place in the human memory consolidation process. It is now well established that sleep plays a key role in human memory performance by stabilizing memory traces and protecting them against interference [1,2,13]. According to [2], one of the functions of dreams might also be to create a virtual environment in which the human brain reinforces and tests certain behaviors. During the memory consolidation process, the hippocampus is responsible for the acquisition and integration of new information that will then be transferred to widespread high-order neocortical areas [3]. McClelland et al. suggest that after the initial acquisition, the hippocampal system serves as a teacher to the neocortex, allowing for the reinstatement of representations of past events in the neocortex [10]. In this manner, this information may be gradually acquired by the cortical system via interleaved learning.

The synergy proposed in this work takes the aforementioned concepts and puts them all together to train a variational autoencoder (VAE) [5] that allows supervised CNNs to take advantage of unlabeled data. Therefore, our model comprises two different neural networks: the VAE and the supervised CNN. The way in which the VAE helps the CNN to deal with mostly unlabeled datasets is (vaguely) inspired by the human sleep-wake cycles. During the waking period, both networks learn from real training data simultaneously. When sleep begins, none of the networks receive real data and the VAE creates a synthetic (virtual) dataset from which the CNN learns. During sleep, only the CNN modifies the parameters of the model. In this manner, the VAE "acts" as a hippocampal system that helps the CNN to reinforce the patterns received during the waking period. Results obtained from experiments carried out on MNIST handwritten digit recognition dataset show the effectiveness of this preliminary approach.

The paper is organized as follows. Section 2 briefly describes the generative model used in this work, i.e., the variational autoencoder. The proposed synergy of variational autoencoders and convolutional networks is presented in Section 3, while Section 4 shows the effectiveness of this method on MNIST dataset. Finally, we conclude the paper in Section 5 and suggest a possible future line to extend this methodology to deeper architectures and larger datasets.

2 Preliminaries: variational autoencoders

Variational autoencoders (VAEs) [5] have become one of the most popular frameworks for building generative models. The reason behind their success is a fast backpropagation-based learning process that does not need strong assumptions. The word *autoencoders* comes from the fact that the neural network built by this technique is composed of an *encoder* and a *decoder*. An autoencoder is a neural network that tries to build an approximate copy of its input that resembles the training data. To this end, the encoder learns a low-dimensional *code* or *internal representation* z of the input x, while the decoder is responsible for reconstructing the original data from the internal code. This model allows one to extract useful properties from training data.

In the case of VAEs, the hidden code z represents a probability distribution that is learned during training, instead of single values. Therefore, the encoder becomes a variational inference network that maps the data to the distribution of the hidden code ($q_\phi(z|x)$), and the decoder becomes a generative network that maps the hidden code back to the distribution of the data ($p_\theta(x|z)$). In this manner, the data generation process starts by sampling z from its distribution. More specifically, VAEs assume that samples of z can be drawn from a simple distribution, i.e., $z \sim N(0,I)$, where I is the identity matrix. This is reasonable because any distribution in d dimensions can be generated by taking d variables that are normally distributed and mapping them through a sufficiently complicated function, such as a Multi-Layer Perceptron (MLP).

In this work, we use a semi-supervised VAE introduced by Kingma et al. [6] that is able to learn from both unlabeled and labeled data. This method is a combination of two different VAEs:

- M1 model: provides a low-dimensional latent representation z_1 of the original data using the following generative model:

$$p(z_1) = N(z_1|0,I); \qquad p_\theta(x|z_1) = f(x;z_1,\theta) \tag{1}$$

  , where $f(x;z_1,\theta)$ is a suitable likelihood function (e.g., a Gaussian or Bernoulli distribution) whose probabilities are formed by a non-linear transformation, with parameters θ, of a set of latent variables z_1. This non-linear transformation is given by a deep neural network.
- M2 model: describes the data as being generated by a latent class variable y plus a continuous latent variable z_2 as follows:

$$p(y) = Cat(y|\pi); \qquad p(z_2) = N(z_2|0,I); \qquad p_\theta(x|y,z_2) = f(x;y,z_2,\theta) \tag{2}$$

  , where $Cat(y|\pi)$ is the multinomial distribution, class labels y are treated as latent variables if no class label is available, and the input data x is given by the latent representation z_1 provided by M1. When y is unobserved, the inferred posterior distribution $p_\theta(y|x)$ predicts the class label, performing classification as inference.

3 Proposal: semi-supervised learning based on the synergy of variational autoencoders and convolutional neural networks

The semi-supervised methodology proposed in this work consists in the interaction of two different deep learning models: variational autoencoders (VAEs) and supervised convolutional neural networks (CNNs). More specifically, our proposal is based on a sequence of "sleep-wake cycles" in which a VAE serves as a teacher to the CNN. This scheme allows the supervised CNN to take advantage of the internal representation created by the VAE from unlabeled data.

The learning process of our model applies the following procedure:

1. The VAE is first trained using the whole dataset (which typically contains a small amount of labeled data) in order to obtain a robust internal representation of the data.

2. Both the VAE and the CNN evolve (learn) simultaneously throughout a sequence of "sleep-wake cycles". Each of these cycles is composed of "sleep" and "waking" periods:
   - Wake: the supervised CNN is trained using only labeled data from the training set. Simultaneously, the VAE learns from the training set using both labeled and unlabeled data.
   - Sleep: the CNN does not receive real data anymore (in this cycle). Instead, it is trained using synthetic data generated and labeled by the VAE, which does not carry out any learning process during sleep (only the CNN learns during this period). We refer to these synthetic samples as "dreams". In order to generate new data, the encoders of M1 and M2 are dropped from the computations, since only the reconstruction path is needed for this process. Instead of having a real image as the input of the network, the decoder of M2 receives a vector sampled from a normal distribution directly and generates the input of the layer z_1 of M1. Finally, the decoder of M1 generates a new image from z_1. In this work, the values taken by the variable y during the generation process are not based on any previously learned parameter. Instead, the VAE generates the same number of samples for all labels in each cycle. It is worth noting that dreams vary from one cycle to another to prevent the CNN from overfitting. This variation is given by a Gaussian diffusion process defined by the following expression:

$$\begin{aligned} df &= \sqrt{(1-\gamma)} * \varepsilon_1 + \sqrt{\gamma} * \varepsilon_2 \\ input\_z_2 &= \varepsilon_1 + s * (df - \varepsilon_1), \end{aligned} \tag{3}$$

     where $\varepsilon_1 \sim N(0,1)$ and $\varepsilon_2 \sim N(0,1)$, $\gamma \in \mathbb{R}$ ranges from 0 to 1, $s \in \mathbb{R}$ sets the smoothing factor, and $input\_z_2$ is the input of the layer z_2 of M2. Both γ and s control the trajectory variations of $input\_z_2$, which determines how dreams vary from one cycle to another. In this work we set γ to 0.8 and s to 1.

As the VAE becomes more reliable, the amount of synthetic data used by the CNN increases. In this manner, the model will become more confident in its own internal representation as it evolves. We must remark that the amount of synthetic data can be kept constant if the CNN varies the weight assigned to synthetic samples, obtaining a similar scenario. A diagram of the proposed methodology is shown in Fig. 1.

From a bio-inspired point of view, we suggest the following scheme. As described, the VAE is responsible for the construction of a robust internal representation of the real data received during the waking period, while the CNN focuses on maximizing discrimination capabilities. Therefore, there are two different neural circuits that are specialized in different tasks. We speculate that the VAE can be vaguely serving as a "hippocampal system" that might reinforce certain patterns in neocortical areas that would be represented by the CNN, which would be acting as a dedicated visual system. This reinforcement comes from a virtual environment created by the VAE during sleep [2], where the CNN learns from data generated from the input representation of past events [10]. In this manner, the VAE would be responsible for the acquisition of new information coming from unlabeled data and the subsequent reinstatement of these patterns in the visual system (CNN).

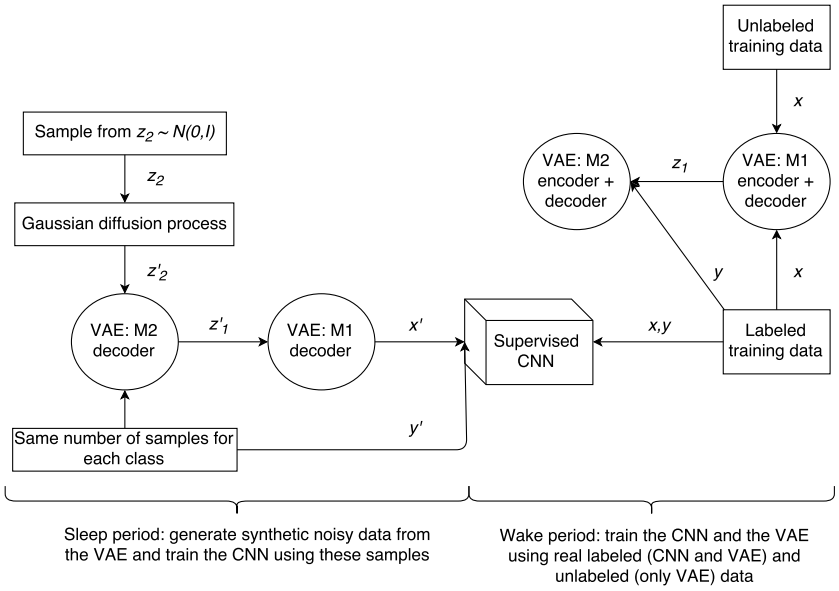


Fig. 1: Proposed sleep-wake cycle.

In comparison with unified objective functions [14] and self-training techniques [12], the advantage of training two separate neural networks (VAE and CNN) is that each network is specialized in a different task. As a consequence, a wide variety of generative models and convolutional neural networks can be applied in this scheme. These properties allow us to build bio-inspired models that might capture certain interactions between human neural circuits.

4 Experimental results on MNIST Handwritten Digits dataset

We tested the effectiveness of our method using the MNIST database of handwritten digits [9]. The architecture and hyperparameters used throughout the experiments are the following:

– Variational autoencoder (VAE): we used the code published by Kingma at GitHub (`https://github.com/dpkingma/nips14-ssl`), which is written in Python with Theano library. Both M1 and M2 models were built considering the configuration recommended by the authors. For M1 we used a 50-dimensional latent variable z. The Multi-Layer Perceptrons (MLPs) of the generative and inference models were composed of two hidden layers, each with 600 hidden units, using softplus $log(1+e^x)$ activation functions. M2 also used 50-dimensional z and softplus activation functions, but in this case the MLPs had one hidden layer, each with 500 hidden units. The likelihood functions for $p_\theta(x|z_1)$ and $p_\theta(x|y,z_2)$ were given by Bernoulli and Gaussian distributions, respectively.

– Convolutional neural network (CNN): we used a simple CNN with two convolutional layers and two fully-connected layers. The convolutional layers had 3x3 receptive fields, 2x2 max-pooling, and 20 and 50 filters, respectively. The fully-connected layers were composed of 512 and 10 units. Dropout ratio was set to 20% and 50% in convolutional and fully-connected layers, respectively. All layers applied ReLU activation functions, except for the output layer, which used a softmax non-linearity.

Both models were trained with the Adam optimizer using the following default settings: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$.

For the experiment, a random subset of the original MNIST dataset is treated as unlabeled data by discarding its labels. Regarding sleep-wake cycles, the VAE and the CNN ran for 10 and 20 epochs in each cycle, respectively, for a total of 20 cycles. Note that the VAE ran for less epochs than the CNN, since it does not perform any learning process during sleep. The number of epochs for the initial training of the VAE was set to 300. In each cycle, the number of synthetic images created by the VAE increases according to the following equation:

$$n_s(i) = n_s(i-1) + \frac{(n_l * 4 - n_s(0))}{c} \quad \text{for all } i \in \mathbb{N} \quad with \quad i > 0, \tag{4}$$

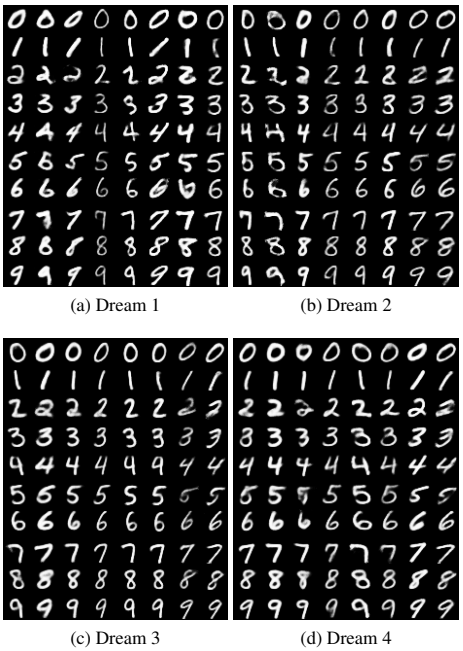
where $n_s(i)$ is the number of synthetic images in the cycle i, $n_s(0)$ is the number of synthetic images in the initial cycle, n_l is the number of labels, and c is the number of cycles. Due to the small amount of labeled data, the batch size was set to 20 for the CNN.

Figure 2 shows a subset of synthetic samples generated by the VAE during sleep. As we can observe, the model was able to construct new images from its internal representation, modifying the trajectory of digits in each cycle to prevent the CNN from overfitting.

In order to assess whether the sleep period was beneficial for the CNN, we ran our method (VAE+CNN) along with a supervised CNN in isolation (baseline). According to Fig. 3, the classification performance of the CNN when using 100 labels shows 10% improvement after 20 sleep-wake cycles. When we used 200 and 300 labels, our approach yields 6% and %4 improvement, respectively, after 20 sleep-wake cycles. These results suggest that the rate of improvement increased as the proportion of labeled data decreased, and thus the CNN was able to take advantage of the data generated by the VAE to improve its classification performance. Moreover, the plots in Fig. 3 show that the CNN of our method kept learning throughout the sleep-wake cycles, while the baseline CNN stopped improving its performance in a few epochs. This behavior suggests that the proposed methodology could be an interesting solution to improve incremental or online learning in CNNs, since the VAE and the CNN are trained simultaneously and the interaction between both neural networks takes place gradually.

5 Discussion

In this work, we have proposed a bio-inspired methodology to improve semi-supervised learning in convolutional neural networks (CNNs) using variational autoencoders (VAEs).

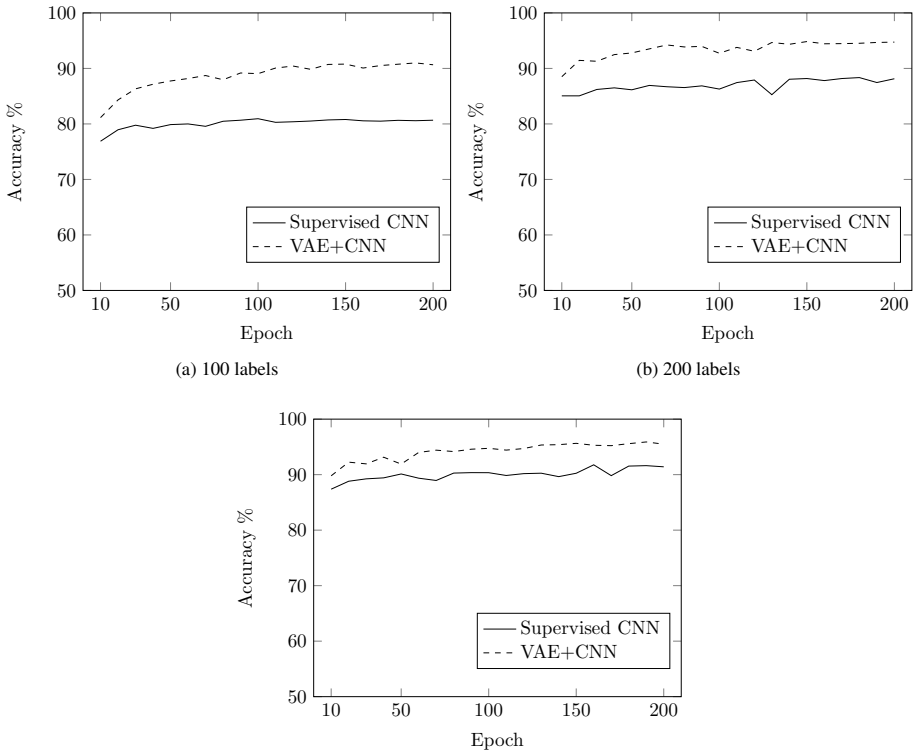


(a) Dream 1 (b) Dream 2

(c) Dream 3 (d) Dream 4

Fig. 2: Synthetic data generated by the VAE during sleep.

In order to improve the classification performance of the CNN with the knowledge extracted by the VAE from unlabeled data, our model runs a sequence of "sleep-wake cycles" composed of sleep and waking periods. These cycles define the way in which both networks interact with each other and allow the CNN to learn from the VAE. During the waking period, both networks are simultaneously trained using real training data. When sleep begins, none of the networks receive real data and the VAE creates a number of synthetic noisy images from which the CNN learns. Based on this procedure, we speculate that the function of the VAE would be twofold: 1) during the waking period, it acts as a "hippocampal system" that is responsible for acquiring unlabeled data and building an internal representation that integrates new information; 2) during sleep, it serves as a teacher to the CNN (that would represent high-level neocortical areas) by creating a virtual environment in which the CNN reinforces the patterns received in the waking period and reinstates the representations of past stimuli. The experiments carried out on MNIST dataset show that the CNN was able to learn from images created by the VAE. More specifically, the classification performance was improved by up to 10% over the purely supervised CNN. The advantage of our approach over simpler semi-supervised solutions is that one could apply any type of generative model or convolutional neural network. Consequently, the usage of deeper models should allow our method to tackle more complex problems, since each network would specialize in either "hippocampal" or "visual" tasks.



(a) 100 labels

(b) 200 labels

(c) 300 labels

Fig. 3: Accuracy of the supervised CNN and the proposed synergy of VAE and CNN on MNIST dataset.

However, this work is a preliminary approach to the proposed methodology and further experiments are needed to assess its performance on large-scale datasets. Future work involves adding mechanisms that allow the VAE to learn how to create "useful dreams" for the CNN. This could be done by adding an extra branch to the decoder of M2, which would specialize in learning the distribution from which "useful dreams" are drawn. In this context, "useful dreams" might be the set of synthetic images that are generated with a high confidence level and force the CNN to make doubtful predictions. The learning process would consist in maximizing the norm of $p_\theta(y|x')$ (which implies confident labeling), x' being the synthetic image, while minimizing the norm of the CNN's output vector (which implies a doubtful prediction). This branch would be active only during sleep (replacing the original branch of the decoder) and would not affect the learning process of the VAE in the waking period.

Acknowledgments. This work has been partially supported by the Spanish Ministry of Science and Technology under the project TIN2016-77356-P (AEI/FEDER, UE).

References

1. Diekelmann, S., Born, J.: The memory function of sleep. Nat. Rev. Neurosci. 11(2), 114–126 (2010)
2. Franklin, M.S., Zyphur, M.J.: The role of dreams in the evolution of the human mind. Evolutionary Psychology 3(1), 59–78 (2005)
3. Geib, B.R., Stanley, M.L., Dennis, N.A., Woldorff, M.G., Cabeza, R.: From hippocampus to whole-brain: The role of integrative processing in episodic memory retrieval. Hum Brain Mapp 38(4), 2242–2259 (2017)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. vol. 2016-January, pp. 770–778 (2016)
5. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: ICLR (2014)
6. Kingma, D., Rezende, D., Mohamed, S., Welling, M.: Semi-supervised learning with deep generative models. In: NIPS. vol. 4, pp. 3581–3589 (2014)
7. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. vol. 2, pp. 1097–1105 (2012)
8. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2324 (1998)
9. LeCun, Y., Cortes, C., Barges, C.: The mnist database of handwritten digits
10. McClelland, J.L., McNaughton, B.L., O'Reilly, R.C.: Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102(3), 419–457 (1995)
11. Rasmus, A., Valpola, H., Honkala, M., Berglund, M., Raiko, T.: Semi-supervised learning with ladder networks. In: NIPS. vol. 2015-January, pp. 3546–3554 (2015)
12. Shinozaki, T.: Semi-supervised learning for convolutional neural networks using mild supervisory signals. Lecture Notes in Computer Science 9950, 381–388 (2016)
13. Wamsley, E.J.: Dreaming and offline memory consolidation. Curr Neurol Neurosci Rep 14(3), 433 (2014)
14. Zhang, Y., Lee, K., Lee, H.: Augmenting supervised neural networks with unsupervised objectives for large-scale image classification. vol. 2, pp. 939–957 (2016)