Neurocomputing 144 (2014) 258–270

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Image receptive fields for artificial neural networks

Jean-Luc Buessler, Philippe Smagghe, Jean-Philippe Urban *

*Faculté des Sciences et Techniques, Université de Haute-Alsace, 4 rue des Frères Lumière, 68093 Mulhouse, France*

ARTICLE INFO

*Article history:*
Received 30 August 2013
Received in revised form
7 April 2014
Accepted 18 April 2014
Communicated by W.L. Dunin-Barkowski
Available online 29 May 2014

*Keywords:*
Computer vision
Visual object recognition
Artificial neural networks model
Supervised learning algorithm
Visual receptive fields
Extreme Learning Machine

ABSTRACT

This paper describes the structure of the *Image Receptive Fields Neural Network* (IRF-NN) introduced recently by our team. This structure extends simplified learning introduced by *Extreme Learning Machine* and *Reservoir Computing* techniques to the field of images.

Neurons are organized in a single hidden layer feedforward network architecture with an original organization of the network's input weights. To represent color images efficiently, without prior feature extraction, the weight values linked to a neuron are determined by a 2-D Gaussian function. The activation of a neuron by an image presents the properties of a nonlinear localized receptive field, parameterized with a small number of degrees of freedom.

A network composed of a large number of neurons, each associated with a randomly initialized and constant receptive field, induces a remarkable representation of the images. Supervised training determines only the output weights of the network. It is therefore extremely fast, without retro-propagation or iterations, adapted to large sets of images.

The network is easy to implement, presents excellent generalization performances for classification applications, and allows the detection of unknown inputs. The efficiency of this technique is illustrated with several benchmarks, photo and video datasets.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

The IRF-NN has been designed by our research team with the purpose of easing the use of images in supervised learning applications. Its algorithm, although very simple, allows us to work directly on photographs, or images, in color or gray-level, without any preprocessing or prior feature extraction. Training of a multi-class recognition task, for example, is simply achieved by presenting a collection of views and their labels.

The network uses the most elementary neural network architecture: a feedforward neural network with a single hidden layer. The novelty of the approach lies in the initialization of the weights. The weights connected to a neuron are not considered independent, but their values are determined by a 2-D Gaussian function discretized into an image size bitmap. The activation of a neuron is a function of the spatial integration of the input pixels weighted by the Gaussian function of their position. The gain factor and the sigmoid function of the neuron modulate this response strongly: some visual stimuli cause a response in the quasi-linear zone of the sigmoid while others provoke a response close to negative or positive saturation.

Albeit elementary, the IRF neuron model presents the properties of a receptive field: its response is mostly sensitive to a local region in the image and to specific stimuli; similar stimuli trigger activations of similar magnitudes. Considering the global network activation, we will show that the neighborhoods induced in the stimuli space are well adapted to the processing of images, since it can take lighting or color changes into account, as well as position, rotation, or shape variations, thanks to the spatial organization of the weights. It can be observed, for example, that moving an object in the image causes a gradual modification of the response.

The initialization of each receptive field depends on a few parameters that represent the neuron's *degrees of freedom* (DOF). A few configuration choices specify a few properties of the receptive field. For example the weights can be either of the same sign or centered with zero mean. This latter case favors a response to the contrast between the central and peripheral region of the receptive field.

A complementary novelty in the IRF-NN approach is the random initialization of the free parameters of the receptive fields. In practice, all free parameters of the receptive fields are initialized randomly: center and radius, and also magnitude and color sensitivity. Thus, each neuron has a specific and unique sensitivity. Unlike other approaches, no notion of convolution product or weight sharing is involved here. The receptive fields are not duplicated or iterated through the image. Our empirical studies establish that the activation vector of the internal layer forms an image encoding with interesting properties for learning algorithms.

* Corresponding author.
*E-mail address:* jp.urban@uha.fr (J.-P. Urban).

http://dx.doi.org/10.1016/j.neucom.2014.04.045
0925-2312/© 2014 Elsevier B.V. All rights reserved.

Supervised learning using a set of examples can be implemented in an extremely simple way. An interesting observation has been successfully introduced in the field of dynamic systems as *reservoir computing* (RC) [1,2], and more recently for classification with the *Extreme Learning Machine* (ELM) [3]. Endowing a network with a randomly initialized internal layer of large size avoids the strenuous adaptation phase of the input weights to the examples. Only the output layer is adapted through supervised learning and the determination of the output weights becomes a linear problem that can be solved using a simple algorithm, without iterations or local minima problems.

The ELM network uses randomly initialized and independent input weights. A theoretical approach shows that this network, like a Multi-Layer Perceptron, has universal approximation capability [4]. It can approximate any continuous target function and classify any disjoint regions. Why would it not work with images? There is no limitation in size or nature of the input vector. Experiments easily reveal the problem. The ELM network can be configured to achieve classification of photographs, whatever their sizes. Its recognition score can reach 100% on the training set. However, no generalization is observed even for photos that differ only slightly.

Generalization is based on the proximity of the vectors in an appropriate space. It is necessary to design the mapping of the inputs into the internal space for a better representation of images. In particular, an image should not be considered as a table of independent components. Pixel values should be interpreted in correlation with their position and their neighborhood.

The notion of receptive fields in the IRF-NN takes this neighborhood implicitly into account. The main difference with an ELM network resides in the initialization of the input layer, not in a supplementary algorithmic step. The weights of the input layer are not independent random variables since their initialization is based only on a dozen DOF by neuron. Empirical results confirm that generalization is effective and appropriate for various image characteristics.

The IRF-NN approach allows us to process images with an efficiency and simplicity similar to the one that made the success of the RC and ELM approaches. Their fundamentals are identical. An ordinary linear regression is sufficient for an efficient adaptation of the output weights, even with a large number of images. The architecture implemented very closely follows the one of a *single layer feedforward network* (SLFN). The weights of the input layer are determined in a generic way, independently of the images to be processed, using a random draw for the free parameters of the neurons. There is only one global coefficient (noted $q$) that takes the dynamics of the image set into account to optimize the nonlinear response globally. Once initialized, the input layer remains constant. It can be stored either as a weight matrix or as a table of free parameters of the neurons. The weight vectors depend only on these parameters and on the size of the images.

The properties of the IRF-NN appear to be remarkable. The examples presented in this paper verify that the algorithm can process a large number of images (several tens of thousands) to distinguish and recognize 1000 objects, and learning time takes only a few minutes. The network is able to generalize efficiently from only a few views, as well as identify a particular view within a very similar set of images. Photograph processing for object recognition is fast, compatible with real-time video applications.

Several recent conference papers [5–7] have presented some of the properties of the IRF-NN. The purpose of this paper is to describe further the principle of the neural network with the basic algorithms and the setting of the parameters to discuss a number of its characteristics and to establish the main network properties.

Section 2 gives a short state of the art and presents the main artificial neural network techniques for image recognition. Section 3 details the IRF-NN architecture, weight initialization for gray-level and color images and the algorithms of the network. Section 4 discusses our concept of *image receptive fields* (IRFs) induced by the organization of the weights. It first gives a general interpretation of the network in terms of random and sparse sampling of a continuous scale-space representation of the input image. Then it provides some detailed information about the network's functioning and its configuration. Section 5 illustrates some IRF-NN properties with various image datasets. It shows that the internal representation forms an efficient encoding of the images. A single linear classifier can then be used to perform classification or recognition of large sets of images.

## 2. Related works

The neural network developed in this paper presents two characteristics that are atypical in the field of image processing. The IRF-NN uses the image directly without prior feature extraction and yet is based on a very simple neural feedforward architecture with only one internal layer. Over the years, the state of the art tends to associate pixel-array inputs with very large (or deep) multilayer networks. Increasing in size is, however, not the only path to improvement. Our work is consistent with a few recent publications showing that simple networks can produce stunning and competitive performances.

### 2.1. Learning techniques applied to images

Object recognition in computer vision has been intensively investigated in the last four decades. When it comes to identifying objects from previously selected examples, the task can be assimilated to supervised classification, a field in which nonlinear approaches like neural networks or SVMs have proven themselves very successful (e.g. [8,9]). An image, however, differs quite a lot from the signals for which artificial neural networks excel. It is a large 2-D pixel array, subject to many fluctuations like lighting or angle of view, and in which the object to be recognized represents only a part of the data vector in variable environments. The region representing the object is itself variable in size, position, and even angle. And it gets even more complex when the context of the scene is taken into account, where several objects can be present and induce shadows, reflections, occultations, etc.

An obvious approach is to reduce the number of variables of the problem. The classifier is not presented with an array of pixels but with a vector of descriptors resulting from various processing steps. Invariant feature extraction eases the classification task; if necessary, the dimension of the vector is further reduced, e.g. by principal axis projection with *principal component analysis* (PCA). The feature-based approach is very general and associated with various decision algorithms like nearest neighbors, classification trees, and bag-of-words. The advantages of using neural networks in this context are notably discussed in the interesting review of Egmont-Petersen et al. [10].

The success of these approaches depends chiefly on appropriate feature selection, which is task specific and subject to the designer skills. There are no generic features available and no theory for feature selection [11]. The improvements of detectors and the evaluation of their performances are active research areas (e.g. [12]). A recent paper by Andreopoulos and Tsotsos [13] presents a very detailed overview of the object recognition literature. It emphasizes that the role of learning algorithms has become much more important in recent years, with more advanced techniques.

In some works, feature selection is an integrant part of learning. The cascade of classifiers approach of Viola and Jones [14] uses a very large number of elementary features, initially far larger than

Download English Version:

https://daneshyari.com/en/article/412200

Download Persian Version:

https://daneshyari.com/article/412200

Daneshyari.com