

Novelty Detection in image recognition using IRF Neural Networks properties

Philippe Smaghe, Jean-Luc Buessler, Jean-Philippe Urban

Université de Haute-Alsace – MIPS
4, rue des Frères Lumière, 68093 Mulhouse, France

Abstract. Image Receptive Fields Neural Network (IRF-NN) is a variant of feedforward multi-layer perceptrons adapted to image recognition. It shows very fast training as well as robust and accurate results on supervised classification tasks. This paper presents another property of IRF-NN: responses of trained networks can be analysed to detect unknown images. Several discriminative and efficient novelty criteria are introduced and tested successfully on the ALOI image dataset. A combination of novelty detection and object recognition is illustrated with a robust, pose invariant application of multi-object localization in various backgrounds.

1 Introduction

Novelty detection is an important addition to supervised classification algorithms. A trained classifier should be able to distinguish an unknown input for which an attempt at classification would not be pertinent. We recently introduced an efficient method for image recognition and classification, named Image Receptive Fields Neural Network (IRF-NN) [1]. In this article, we show that basic measures on the neural output allow the detection of unknown images, without any alteration of the training algorithm.

Artificial Neural Networks have been widely used in the machine learning field and have yielded a high number of applications in pattern recognition, from handwriting [2] to object and face recognition [3]. Recently, a new approach of neural networks known as *reservoir computing* [4] brought a regain of interest in the field, solving efficiently otherwise difficult tasks such as dynamic system identification. A similar technique was recently introduced in feedforward networks as Extreme Learning Machine (ELM) [5]. The innovation relies on a constant hidden layer, with numerous stochastically initialized neurons. The adaptation is only applied on the output layer and can be expressed as a linear regression, thus speeding up and simplifying the training process considerably.

IRF-NN is an adaptation of this approach for images and introduces a new organization of the hidden layer. The input layer weights are not independent but rather computed as Gaussian functions with randomly initialized parameters. Such neurons can be compared to biological receptive fields with local and relatively specific responses. The network can therefore use images without prior feature extraction. The adaptation of the output is handled by a supervised training on a set of examples. IRF-NN shows very high performance for image classification.

We focus in this paper on an additional property of this algorithm. The output of the neural network for an image can be used not only to classify it but also to detect its novelty if it is not included in the learning set. We propose and evaluate a few statistical detection criteria, based on a basic analysis of the output vector.

The rest of the article is as follows: part 2 describes the IRF-NN and shows some interesting results on three known image datasets. In part 3, several criteria for novelty detection are proposed, discussed and tested. The last part presents an application of novelty detection in object localization.

2 Image Receptive Fields Neural Network

The proposed neural structure presented in figure 1 is a classic multilayer perceptron (MLP) [6] with one hidden layer.

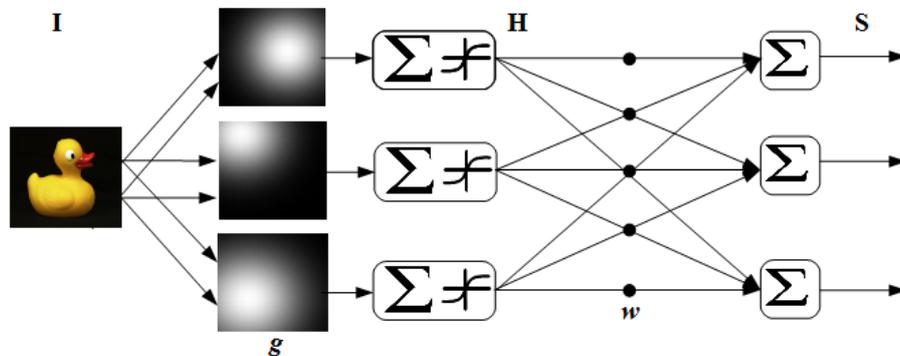


Figure 1 : Structure of the neural network

The neuron activation vector \mathbf{H} is computed by taking the dot product of the raw image \mathbf{I} with weight vectors \mathbf{g}_i , designed as regular functions of the image space. As shown in figure 1, each vector \mathbf{g}_i is defined as bidimensional Gaussian fields with random parameters such as position and standard deviation. The colour is selected with random amplitude α_c for each RGB component. A sigmoid output function ensures the non-linearity of the response. For a neuron i , with x,y the coordinates of the pixels of the image, the activation is:

$$\mathbf{H}_i = \tanh\left(\sum_c \sum_{x,y} \alpha_{ic} I_c(x,y) \cdot \mathbf{g}_i(x,y)\right)$$

The non-linearity combined with the Gaussian structure of the weights results in neurons that are very selective to stimuli, similarly to receptive fields in biology. Every set of images can be accurately represented due to the random initializations and a very high number of neurons.

The training of the weights \mathbf{W} is computed without recurrence, using a linear regression such as $\mathbf{W} = \mathbf{S} \cdot \mathbf{H}^+$, with \mathbf{S} the desired outputs and \mathbf{H}^+ the Moore-Penrose pseudoinverse of the neuron activation \mathbf{H} of the training set. \mathbf{S} is set with a 1-of-n binary encoding in classification applications. The class response R is obtained by taking the maximum neural output such as: $R(\mathbf{I}) = \operatorname{argmax}(\hat{\mathbf{S}}) = \operatorname{argmax}(\mathbf{W} \cdot \mathbf{H}(\mathbf{I}))$.

The neural network described above can be used to solve a supervised classification task from a dataset of image. The images are not pre-processed, and no feature extraction is needed. The algorithm has the advantage of a very fast training compared to other algorithms, with few and easily determined parameters.

We tested our algorithm on several image datasets. MNIST is composed of images of isolated handwritten digits. ALOI-1000 and COIL-100 are composed of various objects captured under different angles of view. The classification results are presented in table 1.

Dataset	Training Images	Neurons	Test Images	Classification Rate	State of Art Results
MNIST	60000	7500	10000	98.6 %	99.8 % [7]
ALOI-1000	18000	9000	54000	99.8 %	89.6 % [8]
COIL-100	1800	900	5400	99.4 %	99.9 % [9]

Table 1 : Classification results on three datasets

The algorithm shows very good performance in all classification tasks and excellent generalization skills for angle-of-view datasets like ALOI and COIL, with many classes and few training images. Training time for ALOI is only 20 minutes, compared to 200 minutes in [8].

3 Novelty detection criteria

Relatively few studies address novelty detection with neural networks. They are summarized in Marsland's [10] interesting synthesis. Some approaches are based on a particular and often hard to define training, but most try to evaluate a distance to known examples. Creating a simple and reliable measure in the input space is a challenging problem, especially in the case of images, and specific algorithm tricks or representation properties are often used to detect novelty. For example, it is possible to measure the convergence time of a dynamic network [11], or the inner distance of a self-organizing map (SOM) [12] to evaluate an input novelty. Zidelmal [13] uses SVMs to integrate input rejection based on distances from separating planes.

We introduce criteria that are simple and can be measured on any classifier response vector for which the structure contains one element per class, but accuracy is mainly based on neural networks properties. Thus, LeCun [14] reports similar criteria on a convolutional network with a large number of neurons for character recognition.

It is well known that neural networks generate results that are consistent for entries similar to the training ones, and inconsistent for unknown entries. The specific configuration of IRF-NN allows the detection of the randomness of the response vector to detect novelty. It takes advantage of two characteristics:

- The inner vector H is designed as a generic high-dimensional representation for any image input, without training on specific subsets. It is also continuous, with neighbour activations for similar images.
- Using the network for classification into many classes with a 1-of- n representation gives a high dimensional and high rank output space.

We propose and study several criteria \mathbf{C} that are relevant to discriminate novel inputs, using the statistical features of the $\hat{\mathbf{S}}$ response vector. Considering $\mathbf{S}^*_i \in \mathbb{R}^m$ the components of the neural response sorted in descending amplitude, we have:

- $C_1 = \left| 1 - (S_1^* - S_2^*) \right|$, an empirical classification margin;
- $C_2 = \left| S_m^* \right|$, the magnitude of the last response;
- $C_3 = \left| 1 - S_1^* \right|$, the magnitude of the first response;
- $C_4 = \sigma(\mathbf{S}) = \sqrt{\frac{1}{m-1} \sum_{i=2}^m (S_i^* - \overline{S_i^*})^2}$, the standard deviation of the response.

Note that each C_x is near 0 for the images of the training set, as the expected responses are $S_1^* = 1, S_i^* = 0 \forall i \in [2..m]$. For an unknown input, C values are greater because S_i is very stochastic and $E(S_i) = 1/m \forall i \in [1..m]$.

We tested these criteria on novelty detection with the ALOI dataset. We used 4 sets. The neural network was trained to recognize object with a *training set* of 9000 images from the first 500 ALOI objects. The *generalization set* contains the 27000 remaining views of the same objects. These images are samples of different inputs but not novelty, as they are very well classified by IRF-NN. The *related set* is composed of 27000 views of the 500 remaining ALOI objects. Objects from these two sets can be alike or different, but the image composition is identical with a centred object in a dark background. Although otherwise useful generalization can be observed, we hope to discriminate these images as novelty. Finally, the *unrelated set* is a collection of 27000 very diverse images extracted from unrelated high-resolution pictures.

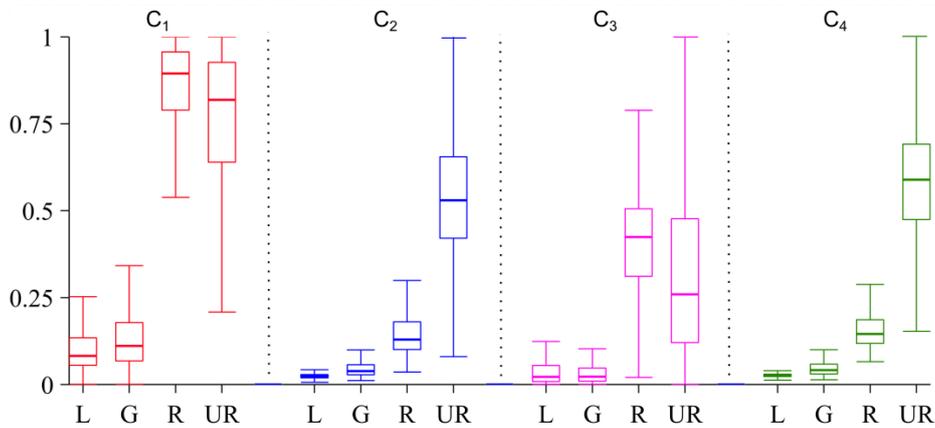


Figure 2: Statistical distributions of criteria C_{1-4} , for Learning, Generalization, Related and UnRelated sets, with quartiles and 2nd and 98th percentiles.

For comparison, the statistical distributions of each criterion for each set are shown in figure 2: these criteria seem significant and sufficient to detect novelty. ROC curves for each criterion are shown in figure 3 to assess thresholding efficacy.

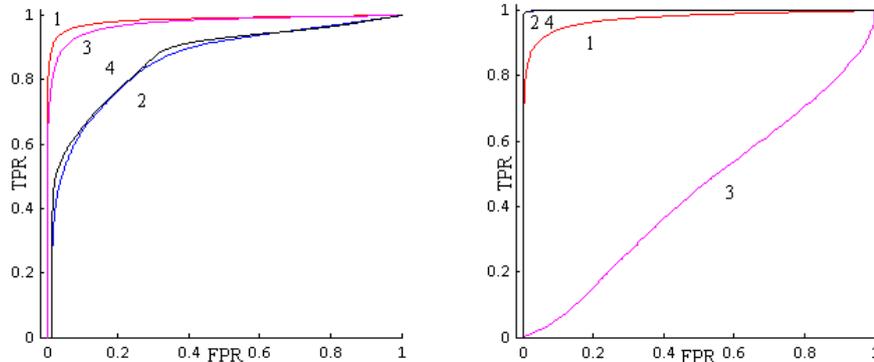


Figure 3 : ROC curves for the discrimination between *generalization set* against a) *related set* or b) *unrelated set*, for criteria 1-4 (numbers)

Figure 3.a evaluates the ability to distinguish images from the *related set* against those of the *generalization set*; and figure 3.b the images from the *unrelated set* against those of the *generalization set*. C_2 and C_4 are very discriminative for unrelated images. For example, C_2 achieves 100% true positive rate with a false positive rate of 0.22%. Meanwhile, C_1 and C_3 are better suited in the case of related images: C_1 achieves a 95.7% true positive rate with a false positive rate of 5%.

This suggests a combination of several criteria depending on the application, in case one needs to discriminate objects that can either be different or similar to the *learning set* in the same application. Novelty can be defined as either *unrelated* or *related* and *unrelated set* and accurately detected.

4 Object localization using novelty detection

We illustrate briefly here the use of the IRF-NN to localize known objects in larger pictures as shown in figure 4.a. The algorithm is efficient and can easily be used to score sub-images extracted from a large picture at regular intervals. By combining the previously described novelty detection criteria and the classification of known objects, the algorithm can simultaneously detect, localize and recognize one or more known objects in a large picture. The mapped results with a 2-pixel interval for the C criteria are shown in figure 4.b.

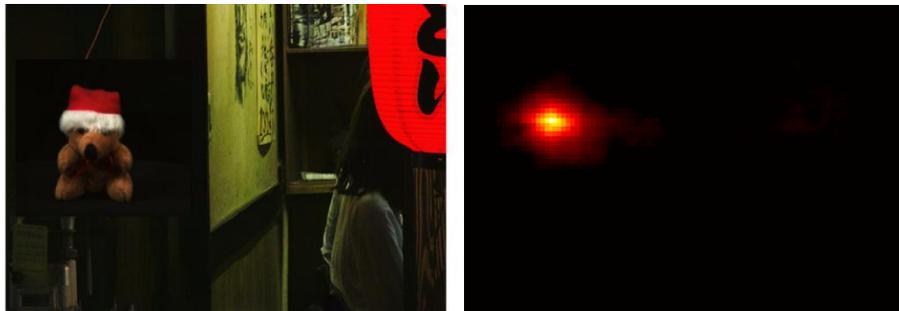


Figure 4: a) Example of localization task b) Mapped results of the criteria for this image

The training is typically done with a hundred objects of ALOI. The object views are duplicated to include distortions such as shifts and background changes. This improves the robustness of the algorithm to the position and context of the object, and allows a bigger scanning interval. The technique is detailed in [15], with results such as a 99.52% detection rate and perfect localization accuracy.

5 Conclusion

IRF-NN introduces a specific weight structure to adapt MLP algorithms to image classification. This approach has shown to be very fast and efficient in tasks of image classification and handwritten digits recognition. IRF-NN also allows the use of the statistical properties of the neural output to detect novelty, and we have successfully tested several criteria on large scale applications with tens of thousands of images. Novelty detection can be exploited in numerous applications, such as multi-object recognition and localization in images.

References

- [1] P. Daum, J.-P. Urban and J.-L. Buessler, Image Receptive Fields Neural Networks for Object Recognition, In *Artificial Neural Networks and Machine Learning (ICANN 2011)*, 2011.
- [2] P. Simard, D. Steinkraus and J. Platt, Best Practice for Convolutional Neural Networks Applied to Visual Document Analysis. In *International Conference on Document Analysis and Recognition. IEEE Computer*, Los Alamitos, 2003.
- [3] F. Yang, et al., Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity Verification. *IEEE Transactions on Neural Networks*, vol. 14(5), 2003.
- [4] M. Lukoševičius et H. Jaeger, Reservoir computing approaches to recurrent neural network training, *Computer Science Review*, vol. 3, pp 127-149, 2009.
- [5] G. Huang, Q. Zhu, C. Siew, Extreme Learning Machine: Theory and Applications. *Neurocomputing*, pp. 489-501, vol. 70(1-3), 2006.
- [6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, New York: Macmillan College Publishing Compagny Inc, 1994.
- [7] D. Ciresan et al., Multi-column Deep Neural Networks for Image Classification, *CVPR*, 2012.
- [8] L. Elazary and L. Itti, A Bayesian model for efficient visual search and recognition, *Visual Search and Selective Attention*, vol. 50(14), p. 1338-1352, 2010.
- [9] J. Matas and S. Obdrzalek, Object Recognition Methods Based on Transformation Covariant Features. In *12th European Signal Processing Conference*, 2004.
- [10] S. Marsland, Novelty detection in learning systems, *Neural Computing Surveys*, pp. 3:157-195, 2003.
- [11] H. Tuong Vinh and J. Rouat, A novelty detector using a network of integrate and fire neurons. *Artificial Neural Networks (ICANN'97)*, 1997.
- [12] A. Muñoz, J. Muruzábal, Self-organizing maps for outlier detection. *Neurocomputing*, vol.18(1-3), p.33-60, 1998.
- [13] Z. Zidelmal, A. Amirou, A. Belouchrani, Heartbeat classification using Support Vector Machines (SVMs) with an embedded reject option, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 26, No. 1, 2012.
- [14] Y. LeCun et al., Handwritten Digit Recognition with a Back-Propagation Network, *Advances in Neural Information Processing Systems*, p.396-404, 1990.
- [15] P. Smaghe, J.-P. Urban, J.-L. Buessler, Déjà-Vu Object Localization using IRF-NN properties, under review for *IJCNN 2013*, 2013