Reconnaissance d'objets par apprentissage d'images - Réseaux de neurones à champs récepteurs aléatoires

Paméla Daum, Jean-Luc Buessler, Jean-Philippe Urban

Reconnaissance d'objets par apprentissage d'images – Réseaux de neurones à champs récepteurs aléatoires

P. Daum J.L. Buessler J.P. Urban

Laboratoire MIPS, Université de Haute Alsace

4 rue des Frères Lumière – F68093 Mulhouse
jp.urban@uha.fr

Résumé

Cet article présente une technique originale pour l'apprentissage dans le domaine de l'analyse d'images.
Des travaux récents ont montré qu'un apprentissage rapide et efficient pouvait être obtenu en adaptant uniquement les poids de sortie d'un réseau neuronal lorsqu'il est composé d'un grand nombre de neurones artificiels. Nous avons adapté cette stratégie à l'image où elle s'avère d'une surprenante efficacité.
Pour travailler avec des images plutôt qu'avec un vecteur de données, les poids d'entrée du réseau sont considérés comme les composantes d'un champ récepteur. Chaque neurone répond à un champ récepteur dont les paramètres, tels que position, taille ou couleur, sont aléatoires La couche adaptative est ajustée par une régression linéaire, sans aucune itération, sans boosting, ou sélection de variables.
Ce réseau, très simple, s'avère étonnamment performant en classification comme pour l'approximation de fonctions par apprentissage supervisé. Il exploite directement des images sans extraction préalable de caractéristiques. La reconnaissance, quasiment sans erreur, de 1000 objets en rotation à partir de 72000 photographies de la base ALOI illustre les remarquables propriétés de cette approche.

Mots Clef

réseaux de neurones artificiels, traitement d'image, classification, extreme learning machine, echo state networks, apprentissage supervisé.

Abstract

Image Learning for Object recognition – Random Receptive Fields Neural Networks

This paper extends a recent and very appealing approach of computational learning to the field of image analysis. Recent works have demonstrated that the implementation of Artificial Neural Networks could be simplified by using a large amount of neurons with random weights. Only the output weights are adapted, with a single linear regression. Supervised learning is very fast and efficient.
We have adapted this approach to image analysis, the novelty being to initialize weights, not as independent random variables, but as Gaussian functions with only a few random parameters. This creates smooth random receptive fields in the image space.
These Image Receptive Fields - Neural Networks (IRF-NN) show remarkable performances for recognition applications, with extremely fast learning, and can be applied directly to images without preprocessing. The almost errorless recognition of 1000 objects in rotation from the 72,000 photographies of the ALOI dataset illustrates the remarkable properties of this approach.

Keywords

artificial neural networks, image processing, classification, extreme learning machine, echo state networks, image receptive fields, supervised learning.

1 Introduction

La notion d'apprentissage supervisé, à partir d'exemples, est attrayante et semble bien adaptée au domaine de l'image. L'expérience commune de tout un chacun suggère qu'il est aisé de reconnaître un objet après l'avoir vu une ou deux fois sous plusieurs angles.

Les réseaux de neurones artificiels (RNA) fournissent un cadre algorithmique pour les techniques d'*apprentissage statistique* [1]. Leur application à l'image constitue un domaine de recherche actif depuis une trentaine d'années.

Certaines applications, comme la reconnaissance de caractères, de chiffres manuscrits [2] ou la localisation de visages [3], confirment l'efficacité de la technique. De nombreuses architectures neuronales, classiques ou plus spécifiques, ont été proposées et testées [4]. Ces réseaux restent cependant difficiles à mettre en œuvre et les réalisations pratiques des RNA pour l'image sont encore rares.

De nouvelles approches, basées sur une large connectivité aléatoire, ont été introduites sous les termes d'*Echo State Network* (ESN) [5] en 2001 pour les systèmes dynamiques, et, plus récemment, d'*Extreme Learning*

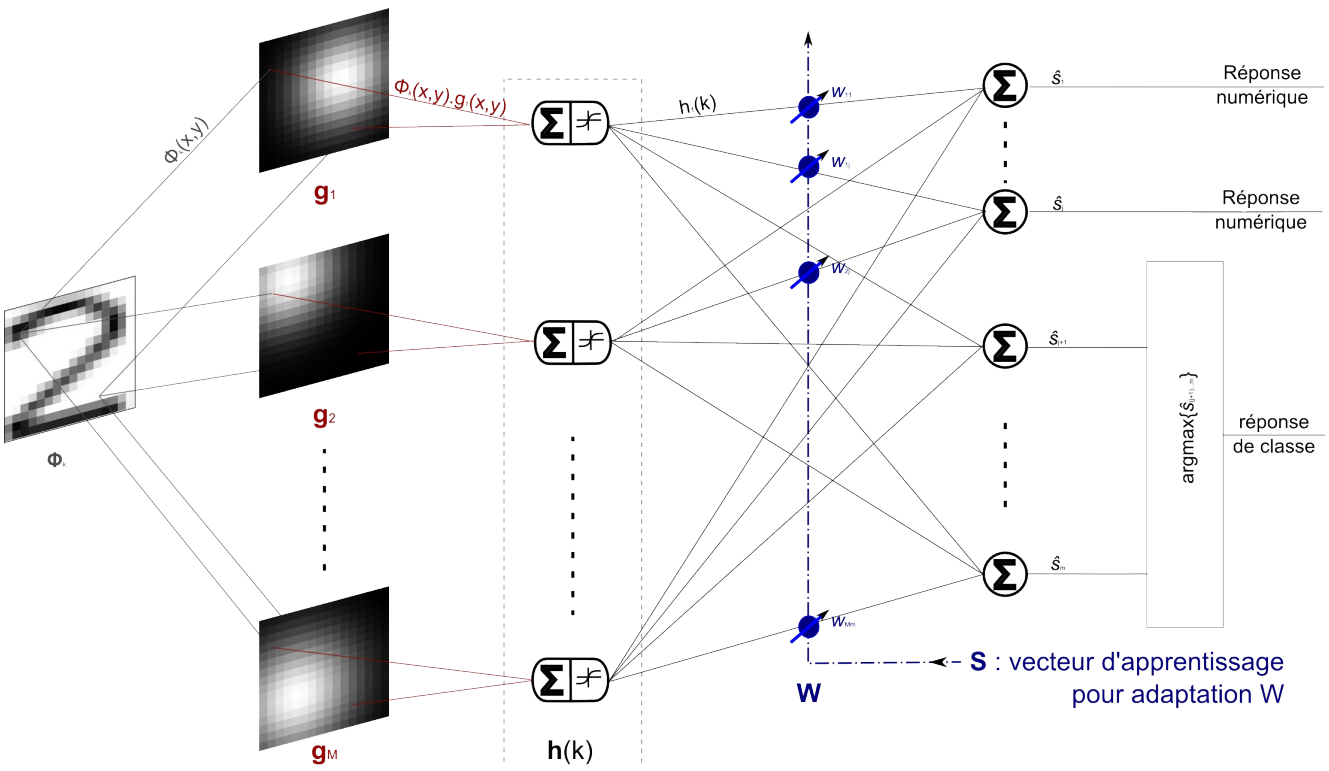

Figure 1 : Réseau de neurones à champs récepteurs aléatoires

Machine (ELM) [6] pour l'apprentissage de fonctions statiques avec des réseaux *feedforward* tels que les perceptrons multicouches. Leur efficacité confirme l'intérêt du concept réseau de neurones et de l'apprentissage supervisé, tout en simplifiant leur mise en œuvre. Nous montrons que ces approches peuvent inspirer des solutions originales dans le domaine de l'image.

Le principe du réseau neuronal n'est pas modifié, nous le détaillerons dans la prochaine section, mais le rôle de l'adaptation est reconsidéré. Plutôt que d'ajuster tous les poids d'un réseau relativement petit pour émuler une fonction, le réseau est constitué d'un grand nombre de neurones dans la couche interne. Les poids d'entrée sont initialisés aléatoirement et restent constants.

L'adaptation du réseau porte donc uniquement sur les poids de la couche de sortie. La phase d'apprentissage est ainsi grandement simplifiée puisque l'ajustement des poids de sortie peut s'exprimer à l'aide de règles linéaires. Malgré la taille des réseaux l'adaptation est rapide et fournit d'excellentes performances.

Pour adapter cette approche au domaine de l'image, nous avons introduit une contrainte sur les poids des neurones de la couche interne. Ces poids ne sont plus considérés comme des variables indépendantes, mais comme les éléments discrets d'une fonction continue et régulière dans l'image. Puisque ces poids modulent la somme pondérée des pixels de l'image, chaque neurone présente une réponse que l'on peut assimiler à un champ récepteur. Ils induisent une propriété de voisinage dans la réponse du réseau à des images qui présentent des similarités.

Le sigle IRF-NN (*Image Receptive Field Neural Networks)* : *Réseau de Neurones à Champs Récepteurs pour l'Image* met en avant cette caractéristique. Reprenant le principe des grands réseaux aléatoires, le IRF-NN comporte une grande couche cachée. Les paramètres libres des champs récepteurs sont déterminés aléatoirement lors de l'initialisation. Les poids de sortie seuls sont ajustés par apprentissage supervisé sur un ensemble d'exemples.

Cette technique apporte d'intéressantes propriétés pour l'utilisation d'images. Nous montrerons dans cette communication qu'elle permet d'utiliser directement une image brute, sans extraction préalable d'attributs. Elle peut certes s'appliquer à des parties d'images préalablement localisées, par exemple pour une reconnaissance de caractères. Mais le IRF-NN en lui-même traite l'ensemble des pixels sans itérations, et sans fenêtres glissantes, même pour des images de taille conséquente.

D'autres auteurs [7] ont très récemment noté l'efficacité d'architectures basées sur des poids aléatoires sans

entraînement pour la reconnaissance d'objet. Nos résultats confirment leurs observations et contribuent à l'évaluation du rôle de l'architecture : les performances sont obtenues ici avec une structure neuronale basique, sans filtres de convolution.

Après une description plus formelle du réseau proposé dans la prochaine section, cette communication illustre les performances du IRF-NN avec deux types d'applications. La section 3 utilise des images de synthèse binaires pour vérifier que le réseau est capable de répondre avec précision à des objets de toute taille et de créer une relative invariance à la position. La section 4 illustre la reconnaissance d'objets tri-dimensionnels dans des photographies avec les bases COIL et surtout ALOI, composée de 72 000 vues d'objets en rotation.

2 Champs récepteurs aléatoires pour l'analyse d'image

2.1 Structure du réseau

Le réseau de neurone proposé, illustré par la figure 1, utilise la structure classique de réseau feedforward à une couche cachée (MLP – MultiLayer Perceptron) [1,8]. Nous en rappelons brièvement le principe tout en introduisant les spécificités de notre approche.

La couche d'entrée d'un MLP réalise une copie du vecteur d'entrée $\mathbf{u} \in \mathbb{R}^d$. Lorsque l'entrée est une image $\mathbf{I}_k$ en niveau de gris, de taille $n_x \times n_y$, le tableau bidimensionnel est redimensionné comme un vecteur de pixels $\boldsymbol{\phi} = \boldsymbol{\phi}_k \in \mathbb{R}^d$ avec $d = n_x . n_y$.

Un élément ϕ_j de ce vecteur est associé à un pixel de coordonnées $[x(j), y(j)]$. Pour simplifier la notation, bien que $\boldsymbol{\phi}$ soit un vecteur, nous utiliserons $\phi(x, y)$ pour désigner l'élément j en fonction de la position du pixel dans l'image.

Un neurone de la couche interne (ou couche cachée) réalise une somme pondérée des valeurs de tous les neurones de la couche précédente. Son activation est définie comme une fonction non-linéaire, généralement sigmoïde de cette somme.

Nous noterons ici g_{ij} le poids de la connexion du neurone j de la couche d'entrée vers le neurone i de la couche interne. L'activation d'un neurone i peut se représenter par l'équation

$$h_i(\mathbf{u}) = \tanh\left(\sum_{j=1}^{d} g_{ij} u_j + g_{i0}\right) = \tanh\left(\mathbf{g_i} . \mathbf{x} + g_{i0}\right) \qquad (1)$$

ou encore, en indexant les coordonnées des pixels lorsque l'entrée est une image

$$h_i(\boldsymbol{\phi}) = \tanh\left(\alpha_i \sum_{x,y} g_i(x, y)\phi(x, y) + \beta_i\right) = \tanh\left(\alpha_i \mathbf{g_i} . \mathbf{x} + \beta_i\right)$$

où α_i et β_i sont respectivement un coefficient multiplicateur et le biais associé au neurone.

La réponse d'une couche cachée de M neurones forme un vecteur $\mathbf{h} = \mathbf{h}(\boldsymbol{\phi})$. La réponse $\hat{\mathbf{s}} = \hat{\mathbf{s}}(\boldsymbol{\phi})$ du réseau peut être scalaire ou vectorielle. Soit m la taille de ce vecteur. La couche de sortie du réseau comporte alors m neurones dont la réponse est déterminée par un vecteur de poids

$$s_q = \sum_{i=1}^{M} w_{qi} h_i \qquad (2)$$

En notation matricielle $\hat{\mathbf{s}} = \mathbf{W}\mathbf{h}$ lorsque les poids de sortie sont regroupés dans une matrice $\mathbf{W}_{m \times M}$.

La matrice $\mathbf{W}$ est déterminée par un apprentissage supervisé à partir d'un ensemble d'exemples $\{(\mathbf{I}_k, \mathbf{s}_k)\}_{k=1}^{N}$. En notation matricielle, $\mathbf{S}_{m \times N}$ représente les sorties désirées pour les N images, et $\mathbf{H}_{M \times N}$ est l'activation des neurones de ces images. L'apprentissage détermine les poids $\mathbf{W}$, réalisant la régression linéaire

$$\mathbf{W} = \mathbf{S}\mathbf{H}^{\dagger} \qquad (3)$$

en notant $\mathbf{H}^{\dagger}$, la pseudo-inverse de Moore-Penrose de la matrice $\mathbf{H}$.

2.2 Neurones à champs récepteurs image

La section précédente a présenté l'architecture classique d'un MLP et quelques éléments de notations. Les modifications proposées pour exploiter des images conservent cette structure, mais étendent le fonctionnement à un très grand nombre de neurones et introduisent une contrainte sur les poids $\mathbf{G}$.

Le fonctionnement classique d'un MLP repose sur l'adaptation de tous les poids, donc $\mathbf{W}$ et $\mathbf{G}$, avec des techniques dites de rétropropagation du gradient de l'erreur, algorithmiquement très coûteuses. Des études récentes, avec les approches ELM et ESN [5,6] ont montré que l'adaptation des poids $\mathbf{G}$ n'était pas nécessaire lorsque la couche cachée est composée d'un grand nombre d'unités. Une initialisation aléatoire de ces poids est suffisante. Le réseau apprend très efficacement des fonctions complexes en adaptant uniquement la couche de sortie $\mathbf{W}$.

Est-il possible de reprendre cette idée dans le contexte du traitement d'image ? L'application directe de cette technique à des vecteurs $\mathbf{x}$ de grande dimension est décevante, et l'utilisation d'images ne fait pas exception. Une extraction préalable de caractéristiques, ou une forte compression, semblent alors incontournables.

Une solution élégante est cependant possible en reconsidérant le rôle des poids $\mathbf{G}$ dans l'équation (1). L'initialisation aléatoire et indépendante de leur valeur est adaptée lorsque les composantes du vecteur $\mathbf{x}$ sont

considérées comme indépendantes (ou relativement indépendantes). Dans une image, les notions de voisinage, de corrélation et d'échelle, sont essentielles pour l'extraction d'informations.

La méthode que nous présentons engendre un voisinage multi-échelle, en conservant la simplicité d'une initialisation aléatoire, totalement indépendante des images à traiter.

Les poids d'un neurone peuvent être fixés de façon à réaliser une fonction de filtre, par exemple un filtre gaussien. La réponse d'un neurone est alors le produit scalaire de ce filtre avec l'image conformément à l'équation (1), et non pas un produit de convolution. En fixant stochastiquement les paramètres du filtre, tels que centre, rayon et amplitude, chaque neurone est associé à un champ récepteur aléatoire dans l'image.

Nous avons retenu une forme gaussienne elliptique pour les champs récepteurs. L'équation de détermination des poids s'écrit alors

$$g_{ij} = g_i(x, y) = \gamma_i + \exp\left[-\frac{(x-\mu_{1i})^2}{(n_x \sigma_{1i})^2} - \frac{(y-\mu_{2i})^2}{(n_y \sigma_{2i})^2}\right] \qquad (4)$$

où σ_i , γ_i et μ_i sont des constantes définies aléatoirement, et n_x et n_y représentent la largeur et la hauteur des images.

La figure 2 représente quelques exemples de poids sous contrainte gaussienne. Bien que $\mathbf{g}_i$ soit un vecteur, la figure restitue son organisation dans le plan de l'image pour en illustrer l'organisation spatiale.

Le nombre de degrés de liberté lors de l'initialisation des poids est ainsi réduit à quelques variables par neurone. Mais le fonctionnement du réseau n'est pas modifié.

Le vecteur **h** correspond à l'activation de tous les neurones pour une image. L'importance de la non-linéarité, ici de type sigmoïde, a été étudiée dans de nombreux travaux sur l'approche neuronale, de fonctions noyaux ou Support Vector Machine [8, 9]. Elle contribue largement aux propriétés remarquables que nous mettrons en évidence dans la discussion sur les résultats. En particulier, il permet d'exprimer, dans l'équation (2), la réponse du réseau comme une combinaison linéaire du vecteur **b**. Les coefficients de cette combinaison linéaire sont ainsi déterminées par une simple régression linéaire sur les exemples d'apprentissage. Malgré la simplicité de cette règle, nous pouvons vérifier à la fois une grande précision de la réponse sur les exemples présentés, et une bonne généralisation.

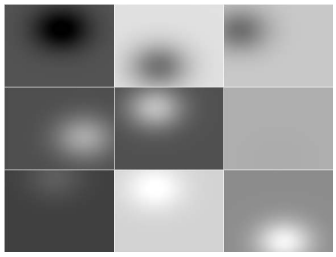

Figure 2 : Champs récepteurs gaussiens $\mathbf{g}_i$ représentés sous forme d'images.

Le réseau ainsi défini est très aisé à implémenter, les seuls coefficients à fixer par l'expérimentateur étant les plages d'initialisation pour les paramètres de la fonction filtre.

2.3 Généralisation à une image couleur

La méthode peut être directement étendue à des images couleur, par exemple à partir d'une représentation RGB. Le vecteur $\phi \in \mathbb{R}^{3d}$ collecte alors successivement les pixels dans les 3 plans de couleurs ; le vecteur $\boldsymbol{g} \in \mathbb{R}^{3d}$ est initialisé comme précédemment dans $\mathbb{R}^d$, puis affecté par une couleur aléatoire dans l'espace RGB.

3 Classification de figures géométriques

Le réseau RN-CRI présente d'intéressantes propriétés pour l'apprentissage supervisé à partir d'un ensemble d'exemples images :

- le réseau utilise les images sans étape préalable d'extraction de caractéristiques ;
- l'apprentissage est rapide et ne nécessite aucune itération ;
- les paramètres à configurer sont très peu nombreux et simples à déterminer.

Dans cette section, ces propriétés sont tout d'abord illustrées avec des objets géométriques et des images binaires. La section 4 présentera des résultats pour la classification de photographies.

3.1 Présentation de l'apprentissage

La première étude est consacrée à l'apprentissage et la généralisation lorsque la taille et la position des objets varient. Nous utilisons des images binaires et des figures géométriques simples. Le jeu de données GEO-1 est constitué d'images de 30x30 pixels. Chaque type de figure prend successivement toutes les tailles de 10 à 28 pixels et toutes les positions possibles sans déborder du cadre de l'image. Au total, GEO-1 (figure 3), comporte 9915 images pour 3 classes d'objets géométriques (carré, disque et triangle).

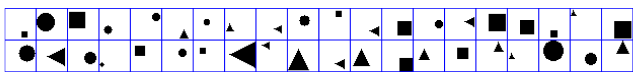

Figure 3 : Quelques exemples de la base GEO-1, constituée de 3 types d'objets synthétiques : des disques, des carrés et des triangles

Table 1 : Paramètres de configuration GEO

	M	σ	α
GEO – 1	200	[0.5 2]	[-1 1]/11
GEO – 2	400	[0.5 2]	[-1 1]/32

La moitié des images, tirée aléatoirement, forme la base d'apprentissage. Le réseau RN-IRF est simultanément entraîné pour identifier la classe de l'objet et les paramètres de position et d'aire avec le vecteur de données $\boldsymbol{S}=[s_h, s_x, s_y, B]^T$.

Les valeurs numériques $s_h(k), s_x(k)$ et $s_y(k)$ représentent respectivement la hauteur et les coordonnées de l'objet dans l'image k. Sa classe $s_c(k)$ est représentée par le vecteur binaire $\boldsymbol{b}(k)\in\{-1,1\}^{n_c}$ de dimension n_c, le nombre de classes, en utilisant un code 1 parmi n_c. Le seul élément mis à 1 indique la classe : $b_{s_c}(k)=1$. La réponse de classification du réseau $\hat{s}_c(k)$ correspond à l'indice de la plus grande valeur de sortie de $\hat{\boldsymbol{b}}(k)$.

Les résultats sur le jeu de test, constitué des images non présentées lors de l'entraînement, sont exprimés comme la racine de l'erreur quadratique moyenne pour les valeurs numériques. Le taux de classification T correspond au pourcentage de bonnes réponses.

Le tableau 1 résume les principaux paramètres de configuration du réseau, avec notamment la plage de valeurs pour le tirage (loi aléatoire uniforme) des variables aléatoires des champs récepteurs. Les centres des gaussiennes μ_{xi} et μ_{yi} sont tirés aléatoirement dans les dimensions de l'image. Les variables β_i et γ_i sont fixées à zéro dans toutes les applications présentées ici.

3.2 Expériences

Une première expérience est menée sur GEO-1. Le jeu d'apprentissage utilisé définit des fonctions relativement complexes, puisqu'à une forme donnée correspondent des images très différentes où varient la taille et la position de l'objet.

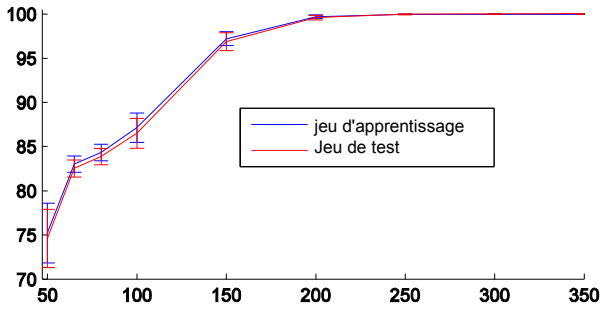

Figure 4 : Taux de bonne classification (en %) en fonction de la taille du réseau (nombre de neurones).

Les figures 4 et 5 présentent le taux T de bonne classification ainsi que l'erreur de prédiction de la taille et de la position de l'objet sur des séries de 50 essais avec 3 classes d'objets géométriques. A chaque essai, toutes les variables aléatoires et poids du réseau sont réinitialisés, ainsi que la distribution entre le jeu d'apprentissage et le jeu de généralisation.

Les figures montrent que :

- après adaptation des poids de sortie du réseau, celui-ci est capable de déterminer sans erreur le type d'objet, et de déterminer sa position avec une erreur moyenne inférieure $< 1{,}6e^{-2}$ pixels ;
- les performances augmentent avec le nombre de neurones. Ainsi des réseaux de plus grande taille garantissent un taux de bonne classification de 100% à chaque essai.

La généralisation s'effectue bien même avec de grands réseaux. En effet, la technique de Moore-Penrose, basée sur la décomposition en valeurs singulières, permet de réaliser l'inversion de la matrice **H** avec une stabilité suffisante.

Une seconde expérience avec les figures binaires illustre la précision du codage interne. Une nouvelle classe d'objets est définie en modifiant un seul pixel dans les images précédentes : tous les carrés de la première série sont tronqués (biseautés) en supprimant le pixel du coin supérieur gauche. Les réseaux entraînés précédemment reconnaissent tous, à 100 %, ces objets comme des carrés, ce qui est conforme à la propriété de généralisation souhaitée. Mais il est également possible d'entraîner de nouveaux réseaux à différentier les 2 classes, c'est-à-dire les carrés complets des carrés biseautés. Ainsi GEO-2 comporte 4 classes: carré, disque, triangle et carré biseauté. GEO-2 est utilisé en apprentissage puis en généralisation, de la même façon que GEO-1. Bien que les 2 classes de carrés ne diffèrent que d'un pixel, la taux de bonne classification est de 99,9%. Des résultats similaires sont obtenus avec des triangles ou des disques : la modification d'un seul pixel est suffisante pour discriminer les formes (voir tableau 2).

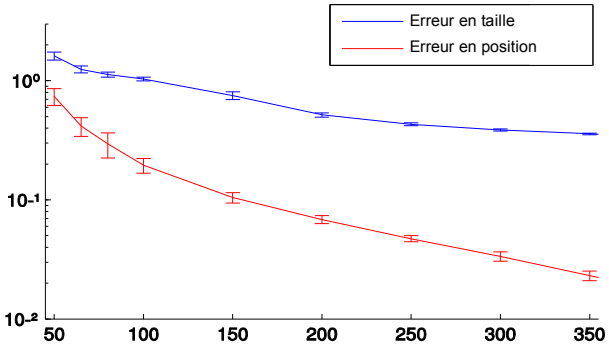

Figure 5 : Erreurs d'estimation (en pixels) en fonction de la taille du réseau (en nombre de neurones)

Table 2 : Résultats GEO : taux de bonne classification et erreurs sur l'estimation de position et de taille

	T (%)	ϵ_x	ϵ_y	ϵ_h
GEO – 1	100	$1{,}6e^{-2}$	$1{,}6e^{-2}$	$4e^{-1}$
GEO – 2	99,9	$4e^{-3}$	$5e^{-3}$	$3e^{-1}$

Le coût algorithmique des réseaux IRF-NN est également intéressant: la durée de l'adaptation des poids pour 200 neurones sur 4959 exemples (image de taille 30x30 pixels) est de 500 ms, et le temps de réponse pour 1 image est < 4 ms[1].

3.3 Principaux résultats

Les résultats confirment expérimentalement quelques caractéristiques intéressantes des réseaux IRF-NN :

- les champs récepteurs gaussiens et la fonction sigmoïde transforment l'image en une représentation interne remarquable qui permet d'approximer des fonctions arbitraires de l'image avec une bonne précision, par simple combinaison linéaire ;
- l'adaptation des poids à partir d'exemples est précise et numériquement stable même avec de grands jeux de données ou un grand nombre de neurones ;
- la généralisation à des images n'appartenant pas au jeu d'apprentissage est efficace.

4 Classification de photographies - la base ALOI

Le réseau IRF présente également de bonnes performances avec des images plus grandes, en niveau de gris ou en couleur, et sur des objets réels. Les tests présentés dans cette section sont réalisés avec la base d'images ALOI [10], une collection d'images d'un millier d'objets. Pour chaque objet, la base contient plus de 100 images enregistrées avec des variations systématiques de l'angle de vue, de l'angle d'illumination et de la couleur de l'illumination. Nous avons utilisé divers sous-ensembles de la base ALOI pour entraîner et tester le réseau. Toutes les images ont toujours été utilisées directement sans prétraitement ni redimensionnement.

Dans cet article nous présentons des résultats obtenus à partir du sous-ensemble *ALOI-Viewpoint*, où les images ont été prises sous différents angles de rotation. Chaque objet est représenté par 72 images couleurs enregistrées en faisant tourner l'objet autour de son axe vertical avec une résolution de 5° pour des images de taille 192x144 pixels.

[1] Implémentation du réseau avec Matlab v.7.9 dans l'environnement Linux OpenSuse 11 (64bits), sur une machine processeur Intel Core 2 CPU 2.4GHz et 2Go de RAM.

Table 3 : Paramètres de configuration COIL et ALOI

	σ	α
COIL	[0,01 0,5]	[-1 1]/0,3
ALOI	[0,01 0,5]	[-1 1]/7,5

La figure 6 présente quelques exemples d'objets pris sous différents angles de vue, illustrant la grande variété de vues de certains objets en rotation. La figure 7 montre que d'autres objets, très similaires, voire identiques, représentent cependant des classes différentes. La base ALOI par sa richesse et sa taille représente un vrai challenge ; les résultats publiés sont encore rares [11, 12].

Les séries de tests sont réalisées successivement sur trois série : les 100 premiers objets, les 250 premiers, puis l'ensemble des 1000 objets. A titre de comparaison, les tests ont aussi été menés sur COIL-100 [13], une base plus ancienne et bien connue, de même type.

Pour chaque objet, 1 vue sur 4 est sélectionnée pour le jeu d'apprentissage. L'adaptation du réseau se fait donc avec un pas de rotation de 20°, le test de validation est réalisé sur les images restantes. La fonction évaluée est la reconnaissance des objets, chaque objet correspond donc à une classe avec un encodage de type 1-parmi-n. Les paramètres d'initialisation du réseau sont indiqués dans le tableau 3.

Les résultats présentés dans le tableau 4 montrent que IRF-NN est capable de reconnaître un objet parmi 1000 autres quel que soit l'angle de vue sur 360°. Les erreurs sont rares, bien que certaines vues de la base ALOI soient difficiles à discriminer.

Dans chaque série d'essais, la base d'apprentissage est identique, l'écart type est donc uniquement induit par l'initialisation aléatoire du réseau, on peut vérifier qu'il est suffisamment faible pour valider ce type d'initialisation.

Table 4 : Résultats de classification COIL et ALOI

Base et N° des objets	COIL 1 à 100	ALOI 1 à 100	ALOI 1 à 250	ALOI 1 à 1000
Nombre de vues apprentissage	1 800	1 800	4 500	18 000
Nombre de neurones	700	700	1 750	9 000
Nombre de vues test	5 400	5 400	13 500	54 000
Nombre de vues mal classées (40 essais)	32,8 ± 2 *moyenne et écart-type*	0,6 ± 0,7	5,4 ± 2	99,4 ± 5
Taux de reconnaissance	99,4 %	99.98 %	99.96 %	99,8 %

Figure 6 : Quelques objets sous différents angles de vues.

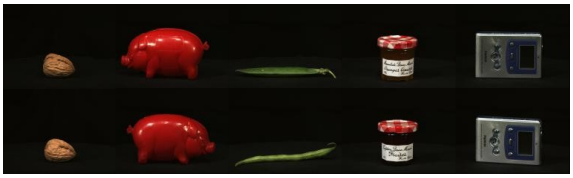

Figure 7 : Quelques couples d'objets très similaires (identifiés par leur numéro dans la base ALOI) constituants cependant des classes différentes.

À titre de comparaison, Elazary et Itti [12] comparent plusieurs techniques de reconnaissance d'objets sur la base ALOI avec rotation des angles de vue. Le taux de reconnaissance atteint 90 % pour un jeu d'apprentissage composé de 25 % des exemples, avec des temps de calculs de plusieurs heures.

Le coût algorithmique de l'approche IRF-NN se compare très favorablement. Le temps d'apprentissage, pour 4500 images et 1750 neurones, est inférieur à 60 secondes, dont 8 secondes pour l'adaptation des poids (3) et 50 secondes pour la lecture et l'encodage des images (1).[2] Le temps d'évaluation pour 1000 images avec ce réseau entraîné est de 12 s, temps de lecture des images comprises.
Avec un réseau de 9000 neurones, les temps augmentent proportionnellement tant que la mémoire vive est suffisante, soit par exemple, 1583 s pour l'apprentissage avec 18 000 images et 69 s pour l'évaluation de 1000 images avec le réseau entraîné.

[2] Implémentation du réseau avec Matlab v.7.12 dans l'environnement Windows 7 (64bits), sur une machine processeur Intel Core i7 3,2 GHz et 16 Go de RAM.

5 Conclusions et perspectives

Cette communication présente un outil pour l'apprentissage adapté aux images. Le réseau IRF introduit la notion de champs récepteurs aléatoires. Il utilise un grand nombre de neurones mais organise les poids de la couche d'entrée uniquement lors de l'initialisation, avec des règles stochastiques indépendantes des exemples. Il adapte uniquement les poids de sortie lors d'une phase d'apprentissage rapide et sans itérations.

Le réseau utilise directement les images pour la reconnaissance d'objets, sans extraction préalable de caractéristiques. Les expérimentations montrent une excellente capacité de généralisation lorsque la vue de l'objet subit un décalage, un changement de taille ou de point de vue. La représentation est suffisamment précise pour différencier des silhouettes qui diffèrent par un unique pixel, bien que l'objet se déplace dans l'image.

Malgré le nombre élevé de neurones, les temps d'apprentissage ou de réponse restent très performants. L'invariance indispensable pour la reconnaissance d'objets dans les images est induite à la fois par le jeu d'apprentissage et par les propriétés des champs récepteurs.

Les résultats décrits ici sont confortés par de nombreuses expérimentations. Plusieurs améliorations sont à l'étude, par exemple pour déterminer les paramètres optimaux, pour intégrer un apprentissage de type récursif sur des lots d'exemples, et pour étendre le champs d'application à des images plus complexes.

Bibliographie

[1] G. Dreyfus, J.M. Martinez, M. Samuelides, M. B. Gordon, F. Badran, S. Thiria, L. Hérault, *Apprentissage statistique*, Eyrolles, 2008.
[2] Simard, P.Y., Steinkraus, D., Platt, J.C., Best Practice for Convolutional Neural Networks Applied to Visual Document Analysis. *Int. Conf. on Document Analysis and Recognition*. IEEE Computer Society, 2003.
[3] Yang, F., Paindavoine, M., Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity Verification. *IEEE Trans. on Neural Networks,* vol. 14(5), pp. 1162-1175, 2003.
[4] Egmont-Peterson, M., de Ridder D., Handels, H., Image Processing with Neural Networks - A Review. *Pattern Recognition,* **35**(10), pp. 2279-2301, 2002.

[5] Jaeger, H., The Echo State Approach to Analysing and Training Recurrent Neural. *Technical Report* (GMD148), German National Research Center for Information Technology, 2001.
[6] Huang, G.B., Extreme Learning Machine: Theory and Applications. *Neurocomputing*, vol. 70(1-3), pp. 489-501 (2006).
[7] Saxe, A., Koh, P.W., Chen, Z., Bhand, M., Suresh, B., & Ng, A. On random weights and unsupervised feature learning. In *ICML 2011*.
[8] Haykin, S., *Neural Networks: A Comprehensive Foundation.* Macmillan College Publishing Compagny Inc., New York (1994).
[9] Cristianini N., Shawe-Taylor J., *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
[10] Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M., The Amsterdam Library of Object Images. *International Journal of Computer Vision*, vol. 16(1), pp. 103-112, 2005.
[11] Song, X., Muselet, D., Trémeau, A., Local Color Descriptor for Object Recognition across Illumination Changes. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.). *Advanced Concepts for Intelligent Vision Systems.* LCNS, vol. 5807, pp. 598-605, Springer, Heidelberg, 2009.
[12] Elazary, L., Itti, L., A Bayesian model for efficient visual search and recognition. *Vision Research*, vol. 50(14), Visual Search and Selective Attention, pp. 1338-1352, 2010.
[13] Nene, S. A., Nayar, S. K., Murase, H. Columbia object image library (coil-100), 1996.