Réseau de neurones à topologie dynamique, comparaison avec des invariants pour la reconnaissance de caractères multi-orientés multi-échelles

Hubert Cecotti, Christophe Choisy, Abdel Belaid

Réseau de Neurones à Topologie Dynamique

Comparaison avec des invariants pour la reconnaissance de caractères multi-orientés multi-échelles

Hubert Cecotti, Christophe Choisy et Abdel Belaid

LORIA/CNRS, Campus Scientifique
BP 239, 54506 Vandoeuvre-les-Nancy cedex France
cecotti@loria.fr, choisy@loria.fr, abelaid@loria.fr

RÉSUMÉ. *Cet article propose l'utilisation d'une dynamique des liens dans un réseau de neurones. L'idée est de changer dynamiquement les liens entre deux couches neuronales de manière à absorber les phénomènes de rotation. Cette méthode est combinée à l'utilisation d'une topologie spécifique qui effectue une transformation polaire. Ce système est appliqué à la reconnaissance de caractères multi-orientés et multi-échelles extraits de plans techniques provenant de l'EDF*[1]*. Nos résultats prouvent l'intérêt de cette approche, mais aussi l'intérêt d'utiliser l'image plutôt que rechercher des invariants pour ce type de problème.*

ABSTRACT. *This paper deals about dynamic topology in neural networks. The idea is to dynamically modify the input neuron of each link in a specific layer, in order to absorb the deformations of the input image. This methodology is combined with a specific topology, that carry out a polar transformation. These proposals are applied to the recognition of multi-oriented and multi-scaled characters extracted from EDF real maps. Results prove the interest of this approach, but also the power of the image compared to invariant extraction for this kind of problem.*

MOTS-CLÉS : *réseau de neurones, topologie dynamique, multi-orienté, recherche d'invariants, transformation polaire, Fourier, Fourier-Mellin*

KEYWORDS: *neural network, dynamic topology, multi-oriented, research of invariants, polar transformation, Fourier, Fourier-Mellin*

1. EDF : Electricité de France

1. Introduction

Il existe deux principales et opposées approches pour la reconnaissance de caractères : la première essaye de trouver un ensemble de caractéristiques représentatives pour chaque classe [DUR 96], la seconde essaye d'apprendre toutes les déformations et variations possibles d'un caractère [SIM 03]. Si la première méthode peut être efficace avec un faible coût pour des données relativement propres, la seconde approche a prouvé sa supériorité pour des caractères très déformés comme les chiffres manuscrits [LEC 01]. Cependant cette seconde méthode nécessite une base d'apprentissage représentative de toutes les distorsions possibles, ce qui conduit nécessairement à une base de taille conséquente.

Considérons désormais le problème des caractères multi-orientés : si pour des caractères relativement droits une telle base peut rester raisonnable bien qu'importante [SIM 03], cela conduira à une énorme base de données pour les caractères multi-orientés, le nombre de déformations prises en compte étant multipliées par le problème d'orientation. De plus, s'il est possible qu'un système obtienne de bons résultats avec une telle base, il est vraisemblable que sa complexité sera énorme, probablement intractable. C'est pourquoi la littérature propose habituellement d'extraire des caractéristiques invariantes à la rotation pour représenter chaque classe [ADA 00, ZHE 00, DER 01] : cette approche présente le risque de perdre des informations, on peut donc s'attendre aux limites des approches par caractéristiques représentatives.

Il existe donc un dilemme entre la perte d'informations et la complexité extrême du système et de la base d'apprentissage. Une solution intermédiaire consiste à redresser les échantillons : ainsi, la base d'apprentissage et la complexité du système restent raisonnables, et l'on peut travailler sans extraction de caractéristiques, donc en réduisant les pertes d'information au minimum. Pour cela, nous proposons l'utilisation d'une topologie dynamique dans un réseau de neurones (RN). L'idée est de modifier les connexions entre deux couches de manière à "redresser" l'échantillon à l'intérieur même du réseau. Le principe du changement de topologie est d'agir de la même façon sur les liens des neurones d'une couche spécifique, de manière à garantir une distribution adaptée de l'erreur durant la phase d'apprentissage. Lors de la reconnaissance, cette méthode permet de propager l'information comme si l'échantillon était droit.

Pour simplifier la dynamique du réseau, nous utilisons une transformation polaire pour passer du problème $2D$ de la rotation à un problème $1D$; cette transformation est intégrée dans le réseau de neurones au travers d'une topologie spécifique.

Dans la suite de l'article, nous présentons tout d'abord le principe général du système proposé. La section suivante discute de la transformation polaire et de son intégration dans un RN. Ensuite, le principe de la topologie dynamique est décrit. Les résultats obtenus sont alors discutés.

2. Description du système

Notre système est basé sur un perceptron multicouches avec plusieurs propriétés spécifiques. Il a été conçu de manière à être le plus possible libre de tout prétraitement, de manière à respecter au maximum les caractéristiques initiales de l'image. Ainsi, la seule opération effectuée avant l'analyse par le RN est une mise à l'échelle, pour ramener l'image à la taille d'entrée du réseau. La figure 1 montre les différentes étapes de notre système.

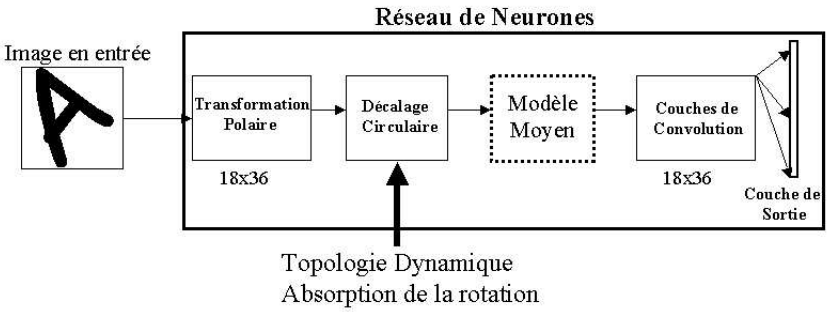

Figure 1. *Vue d'ensemble du système*

Le RN prend en entrée une image en niveaux de gris. Les liens entre la première et la seconde couche sont organisés de manière à effectuer une transformation polaire de l'image : ainsi la seconde couche peut être considérée comme la transformation polaire de l'image d'entrée. Durant la phase d'apprentissage, les poids des liens sont ré-estimés pour prendre en compté l'intérêt de chaque pixel dans la transformation. Cette transformation polaire permet de transformer le problème $2D$ de rotation de l'image en un problème $1D$ de décalage le long de l'axe angulaire.

Les deuxième et troisième couches sont connectées avec des liens dynamiques. La modification dynamique de ces liens permet de suivre le décalage sur l'axe des angles, et donc de transférer l'information aux couches suivantes comme si l'image était initialement droite. Une couche de convolution raffine ensuite l'analyse de l'image ; la couche de sortie, complètement connectée à cette dernière, fournit le résultat de l'analyse.

3. La transformation polaire

Le problème de l'absorption de la rotation entraîne souvent à transformer l'image initiale en coordonnée polaire dans une première étape [GOS 85, BUI 99]. En effet, il est plus facile de traiter la forme polaire car le problème de la rotation est transfor-

mée en problème de translation : en conséquence, une déformation $2D$ devient une déformation $1D$. La figure 2 illustre cet effet.

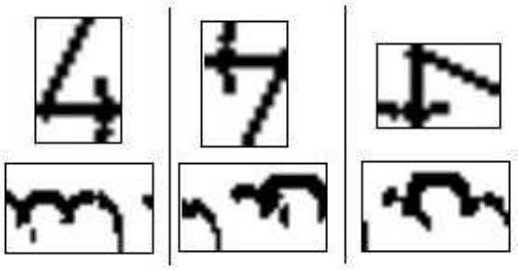

Figure 2. *Transformation polaire : effet du décalage. En haut : image d'origine, En bas : image transformée*

Considérons une image I_c de taille $X * Y$ définie en coordonnée cartésienne $I_c(x;y)$ où $0 \leq x \leq X-1$ et $0 \leq y \leq Y-1$. Si l'image I_c est tournée d'un angle $\alpha \in [0;2\pi]$, la nouvelle image est défini par $I'_c(x';y') = I_c(x * cos(\alpha); y * sin(\alpha))$. Soit G le centre de gravité de l'image. Considérons l'image en coordonnée polaire $I_p = (r,\theta)$: la relation entre I_p et l'image tournée d'un angle α est $I'_p(r';\theta') = I_p(r;\theta+\alpha)$. r est défini par la distance entre G et le point. $\theta \in [0;2\pi]$ correspond à l'angle. La variation d'angle peut donc être traitée comme un décalage le long de l'axe θ. Plusieurs possibilités sont proposées pour effectuer cette transformation.

3.1. *Amélioration de la transformation de Goshtasby*

Soit R le rayon maximum de l'image. R est défini comme la distance euclidienne maximum entre le centre de gravité et le point gris le plus éloigné ; pour éviter de prendre en référence un point de bruit, celui-ci ne doit pas être isolé. Toutes les transformations polaires se basent sur le centre de gravité, référence relativement stable pour une classe donnée.

La méthode de Goshtasby [GOS 85] consiste à dessiner n cercles concentriques autour de G de rayon $R * i/n\,; i \in \{1,..,n\}$. Chaque cercle est découpé en m parts. Dans cette solution, seuls les pixels à l'intersection des cercles et des partitions angulaires sont choisis, pour construire la $n * m$ matrice polaire.

Avec ce type de description, l'image originale n'est pas complètement représentée par l'image transformée ; nous proposons donc d'utiliser la valeur des sections plutôt que des intersections. Cela permet d'améliorer nettement la qualité de la transformation. Cependant les surfaces grandissent avec le rayon : les parties près du centre contiennent peu d'information alors que celles éloignées recouvrent beaucoup d'informations.

Nous proposons donc une autre approche basée sur des sections à surfaces égales. Soit r le pas du rayon, πr^2 est la surface du cercle intérieur. Considérons l'anneau n, la surface S_n de cet anneau est définie par $\pi((n+1)r))^2 - \pi(nr)^2$, $S_n = (2n+1) * \pi r^2$. En divisant le n^{ht} anneau en $2n+1$ parts, les surfaces de sections sont égales pour chaque anneau de l'image. Le nombre de section de l'image polaire pour n anneaux est $\sum_{i=0}^{i=n-1}(2i+1)$. Cette approche permet une meilleure répartition des pixels dans l'image transformée comme le montre la figure 3, permettant d'améliorer de près de 5% les résultats.

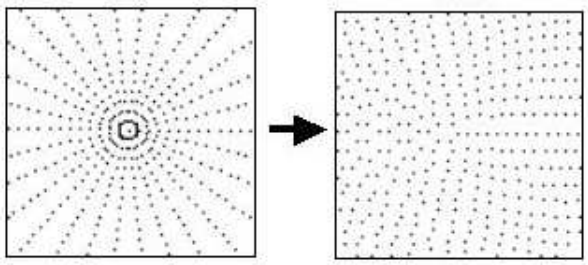

Figure 3. *Répartition des centres des sections considérons la transformation classique (à gauche) et les surfaces égales (à droite)*

3.2. *Intégration de la transformation polaire dans le réseau de neurones*

Le fait de donner une topologie à un réseau de neurones est classique : elle permet de trouver des formes élémentaires et de réduire la complexité du réseau. L'avantage est de laisser au réseau d'apprendre lui-même quelle est la contribution de chaque pixel dans ces images. Nous proposons une autre utilisation : l'intégration de la transformation polaire dans le RN à travers une topologie spécifique.

La méthode classique et celle avec des surfaces équivalentes ont été testées. La transformation à surfaces égales offre des résultats supérieurs de près de 5% que les sections dépendant du rayon, mais pour cet article nous utilisons uniquement la transformation classique, car elle possède une propriété intéressante : chaque colonne de l'image transformée correspond à une section angulaire (voir figure 4). Pour la transformation à surfaces égales, le nombre de zones angulaires varie en fonction du rayon, rendant la gestion de la partie dynamique du réseau plus complexe.

4. Topologie Dynamique pour le processus

L'utilisation d'une topologie dynamique est l'étape suivante dans le réseau de neurones à convolutions. Comme nous l'avons vu dans l'introduction, un apprentissage optimal a besoin d'avoir vu toutes les variations possibles. Pour les caractères

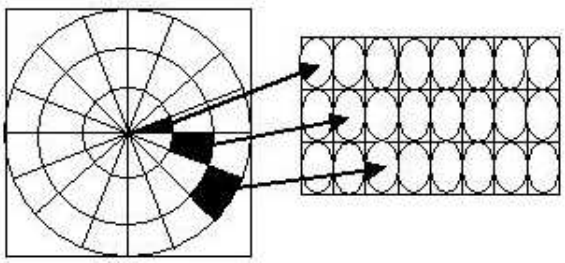

Figure 4. *A gauche : Couche 0, Image d'entrée ; A droite : Couche 1, image en coordonnée polaire. Les fléches représentent des ensembles de liens entre neurones*

multi-orientés, nous devons les avoir pour tous les angles possibles. Cela entraîne une énorme base de données et un système de très grande complexité. Cela pourrait être possible de réaliser un tel système mais ce ne serait pas réaliste de l'utiliser.

Après la transformation polaire, nous proposons de corriger l'erreur de décalage selon l'axe θ pour recaler l'image à une position droite. Nous n'avons donc pas besoin d'apprendre toutes les déformations possibles mais seulement les images droites. Ceci permet d'avoir un réseau de neurones à la complexité similaire aux applications classiques.

Le principe des liens dynamiques est de changer la position des liens en entrées d'un neurone, en fonction de l'erreur de décalage. L'idée intuitive est montrée dans la figure 5. L'objectif est que le premier lien de chaque neurone va toujours être connecté à la véritable première colonne de l'image polaire. Si l'image est décalée, les liens sont dynamiquement décalés à la position de la première colonne. Ce mouvement est synchronisé pour tous les liens dynamiques. Sur une ligne de neurones de sortie, tous les neurones observent la même ligne, et les liens sont déplacés de façon synchrone. Cela permet d'avoir classiquement plusieurs neurones de sortie pour le même groupe de neurones d'entrée.

La notion d'ordre dans les liens est introduite : si la phase d'apprentissage et de test donnent le même rôle à chaque lien, la notion de dynamique les différencie selon la position où ils analysent l'image.

5. Implémentation et résultats expérimentaux

Notre système a été testé sur une base réelle de caractères multi-orientés et multi-échelles. Ces caractères proviennent de plans techniques de l'EDF. La distribution et l'orientation des caractères ne sont pas homogènes. La base de données est composée de chiffres, de lettres latines majuscules et minuscules. La base contient 16575 échantillons en apprentissage et 4169 en reconnaissance (répartition 80%/20%), répartis

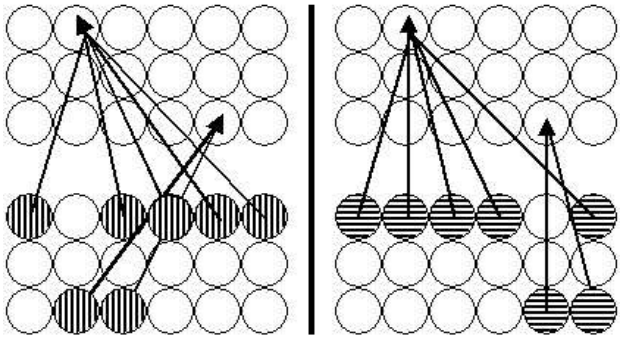

Figure 5. *Topologie dynamique. A gauche : Image droite en entrée, A droite : Image tournée en entrée*

dans 62 classes. EDF fournit une information de l'orientation des caractères ; celle-ci n'est pas entièrement fiable, car elle se base sur la ligne de base des chaînes dont sont extrait les caractères : les chaînes sont souvent courtes, réduisant la fiabilité de l'information d'orientation. Bien qu'approximative, on remarque que la moitié des caractères a une orientation estimée à moins de 5^o : cette quantité est déjà faible compte tenu du nombre de classes. Si l'on veut travailler véritablement en multi-orienté, on pourrait estimer pour un pas de 1° qu'il faudrait multiplier cette base par 360. Notre hypothèse initiale d'une base nécessairement énorme pour traiter du multi-orienté directement s'en trouve donc confirmée.

Nous avons également effectué des tests sur la base MNIST de chiffres manuscrits, comportant 60000 échantillons en apprentissage et 10000 en reconnaissance.

Nous avons testé notre méthode de liens dynamiques d'après les informations d'EDF, un redressement a-priori au degré près basé sur les informations d'EDF, la transformée de Fourier-Mellin déjà utilisée pour cette problématique [ADA 00], une analyse par transformée de Fourier, et les caractères multi-orientés tels-quels.

Les regroupements effectués tiennent compte d'une part des similitudes visuelles existantes en rotation ('6','9' ou 'E','3') ou en taille ('U','u'), d'autre part des confusions notables révélées par l'analyse des résultats. Pour effectuer ces regroupements, nous nous sommes appuyé sur les résultats de Fourier-Mellin, car c'est d'après la littérature l'invariant le plus performant pour ce problème.

Pour les tests concernant les liens dynamiques, les caractères redressés et les caractères multi-orientés, la transformation polaire a été faite par partition angulaire de 10°, valeur déterminée empiriquement : beaucoup d'images ayant un faible diamètre maximum, une discrétisation angulaire plus fine n'apportait rien. Le rayon moyen des images, de 18 pixels, a cependant été conservé pour la transformation polaire : les images transformées sont donc de taille $18x36$.

Dans le cas de la transformée de Fourier, celle-ci s'effectue pour chacun des 18 rayons. Nous avons opté pour 32 paramètres par rayon, cela pour pouvoir utiliser plus simplement la FFT[1], qui s'applique bien sur les séries dont la longueur est une puissance de 2.

Pour la transformée de Fourier-Mellin, nous avons utilisé la version travaillant directement en coordonnées cartésiennes [ADA 00], afin de conserver au maximum la qualité de l'image d'origine ; nous avons évalué les invariants obtenus à l'aide d'un perceptron multicouches complètement connecté, la couche d'entrée et la couche cachée comportant 33 ou 119 invariants selon le test, la couche de sortie ayant 62 neurones comme pour tous les autres tests.

Le RN à topologie dynamique nécessite une estimation du changement de topologie à effectuer. Actuellement, nous ne disposons pas encore de système d'évaluation : nous nous sommes donc appuyé sur les informations données par EDF. L'objectif est en effet de montrer le potentiel de notre approche. Notons qu'il existe une différence avec les tests sur les caractères redressés : pour ces derniers, le redressement se fait au degré près lors de la transformation polaire, ce qui permet d'être au plus juste pour la discrétisation. Pour la topologie dynamique, la transformation polaire se fait comme si l'image était droite, ce qui introduit un premier biais ; ensuite, l'information d'EDF est retranscrite en terme de décalage, introduisant un second biais par sa discrétisation.

Le tableau 1 montre les résultats obtenus, considérant les 62 classes, et pour des regroupements optimisés pour Fourier-Mellin.

Méthode	Entrée	62 classes	44 classe	36 classes
multi-orientés	18x36=648	66.01%	76.30%	77.52%
redressés EDF	18x36=648	77.72%	87.41%	87.26%
topologie dynamique	18x36=648	77.60%	87.65%	87.29%
Fourier-Mellin	33	58.53%	72.70%	75.68%
Fourier-Mellin	119	62.37%	73.93%	75.87%
Fourier	18x32=576	62.92%	72.32%	75.27%

Tableau 1. *Taux de reconnaissance pour différentes méthodes sur la base de EDF*

Nous avons également effectué quelques tests sur une base synthétique de 62 classes également. Elle est constituée de caractères multi-orientés par pas de 15°. Les bases d'apprentissage et de reconnaissance ont été générées en prenant une image sur deux, ce qui provoque un décalage minimum de 15° entre les images en apprentissage et celles en test. La table 2 montre les scores obtenus.

Enfin, nous avons également évalué Fourier-Mellin sur la base MNIST, obtenant un score de 88.75%, résultat proche de la littérature [S. 01]. La référence actuelle sur cette base obtient 99.96% en travaillant directement sur l'image [SIM 03].

1. FFT : Fast Fourier Transformation

Méthode	Apprentissage	Tests
Multi-orientés	76.48%	45.83%
Redressés	99.44%	93.33%
Topologie dynamique	99.44%	90.28%
Fourier-Mellin (33 invariants)	91.94%	83.20%
Fourier (18x32)	90.46%	81.89%

Tableau 2. *Taux de reconnaissance pour différentes méthodes sur une base synthétique, pour 62 classes*

La base synthétique montre clairement qu'il est nécessaire de bien couvrir le spectre des orientations pour obtenir des scores corrects, ce qui signifie avoir une base très conséquente. Sur ce point, on voit également l'intérêt des invariants de Fourier et Fourier-Mellin, qui offrent des scores respectables. Ce résultat ne se retrouve pas pour la base EDF : la nature des plans ainsi que diverses expériences sur cette base nous amène à penser qu'il existe un certain nombre de classes d'orientations, permettant au RN de s'en sortir mieux qu'avec les invariants. Cela confirme à nouveau nos hypothèses de départ : s'il n'y a pas trop de variations, il est possible d'avoir un résultat correct avec suffisamment d'échantillons, même s'il y a plusieurs orientations relativement bien définies.

Nous constatons également que Fourier-Mellin s'avère plus performant que Fourier simple, car il parvient à l'égaler avec près de 5 fois moins de caractéristiques. Notons que les tests avec 33 caractéristiques nous ont été suggérés par la littérature, avec des résultats notablement différents.

Cependant, aucune de ces approches ne parvient à égaler ni la topologie dynamique, ni le redressement au degré près. Cela montre bien que l'extraction d'invariants ne permet pas de conserver toute l'information pertinente, et qu'il est donc préférable de travailler directement sur l'image.

On peut noter que la topologie dynamique s'avère similaire au redressement au degré près sur la base EDF, mais moins bonne sur la base synthétique. Nous pensons que cela s'explique par l'incertitude de la base EDF : l'orientation proposée n'étant pas tout à fait fiable, elle introduit une variabilité que le RN sait gérer correctement, soit par les liens dynamiques, soit par le redressement a priori. Par contre, la base synthétique est complètement fiable et d'une grande "netteté" : les biais introduits par la double discrétisation (transformation polaire et interpolation de l'angle) pénalisent donc un peu cette approche.

6. Conclusion

Nous avons proposé un réseau de neurones à topologie dynamique, pour l'absorption de la rotation d'images. Notre approche travaille directement sur l'image, la rendant nettement plus efficace (+15%) que l'extraction d'invariants à la rotation, prouvant ainsi qu'ils ne permettent pas de récupérer toute l'information pertinente. Notre solution permet de pallier les limite inhérentes de la multi-orientation traitée de manière brute : une base de donnée énorme et un système très complexe. En conservant une complexité raisonnable tant pour le modèle que pour la base d'apprentissage, elle permet d'atteindre des scores similaires à ceux obtenus sur des caractères redressés.

Nous étudions l'adaptation de la topologie dynamique à une transformation polaire plus évoluée, améliorant de près de 5% les scores sur des caractères redressés, mais qui nécessite une gestion de la topologie plus complexe. Nous travaillons également sur un processus décisionnel apte à piloter la dynamique des liens ; en particulier, certains résultats obtenus par des modèles de Markov cachés semblent prometteurs pour ce travail.

7. Bibliographie

[ADA 00] ADAM S., OGIER J., CARIOU C., MULLOT R., J. G., LECOURTIER Y., « Fourier-Mellin based Invariants for the recognition of multi-oriented and multi-scaled shapes : application to engineering drawings analysis », *Machine perception and artificial intelligence*, vol. 42, 2000, p. 132-147.

[S. 01] ADAM S., « Interprétation de documents techniques : des outils à leur intégration dans un système à base de connaissances », PhD thesis, Université de Rouen, 2001.

[BUI 99] BUI T., CHEN G., ROY Y., « Translation-Invariant Multiwavelets for Image Denoising », *Proc. 3rd World Multi-Conference on Circuits, Systems, Communications and Computers.*, , 1999.

[DER 01] DERRODE S., GHORBEL F., « Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description », *Proc. 3rd World Multi-Conference on Circuits, Systems, Communications and Computers.*, vol. 83, n° 1, 2001.

[DUR 96] DUR TRIER O., JAIN A. K., TAXT T., « Feature Extraction Methods for Character recognition – A Survey », *Pattern Recognition*, vol. 4, n° 29, 1996, p. 641-662.

[GOS 85] GOSHTASBY A., « Description and discrimination of planar shapes using shapes matrices », *IEEE Transactions of Pattern Recognition and Machine Intelligence*, vol. PAMI-7, n° 6, 1985, p. 738-743.

[LEC 01] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P., « Gradient-Based Learning Applied to Document Recognitionntion », *Intelligent Signal Processing*, , 2001, p. 306-351.

[SIM 03] SIMARD P., STEINKRAUS D., PLATT J., « Best Practice for Convolutional Neural Networks Applied to Visual Document Analysis International Conference on Document Analysis and Recogntion », *ICDAR, IEEE Computer Society*, , 2003, p. 958-962.

[ZHE 00] ZHENJIANG M., « Zernike moment-based image shape analysis and its application », *Pattern Recognition Letters*, vol. 2, n° 21, 2000, p. 169-177.