Réseau de neurones convolutionnel et support vecteur machine pour l'estimation de la qualité d'image sans référence

ALADINE CHETOUANI

Laboratoire PRISME
12 rue de blois, 45067 Orléans, Orléans, France
aladine.chetouani@univ-orleans.fr

Résumé - Dans cet article, nous proposons une méthode d'évaluation de la qualité d'image basée à la fois sur un réseau de neurones convolutionnel (CNN : Convolutionnal Neural Network) et la combinaison de certaines mesures. Composé de 4 couches de convolution et 1 couche entièrement connectée, le modèle CNN est utilisé pour identifier le type de dégradation contenu dans l'image, tandis que l'étape de combinaison de mesures vise à prédire sa qualité, selon le type de dégradation identifié. La fusion des mesures est réalisée à l'aide d'un modèle SVR (Support Vector Regression). Trois bases d'images ont été utilisées pour valider la méthode proposée. Les résultats obtenus montrent sa pertinence aussi bien en termes d'identification et qu'en termes d'estimation de la qualité.

Abstract - In this paper, we propose a framework for image quality assessment based on both Convolutional Neural Network (CNN) and combination of some features. The CNN model is used to identify the degradation type contained in a given distorted image, while the combination step aims to predict its quality according to the identified degradation. Our CNN model is composed of four convolutional layers and one fully connected layer. The fusion step is here done using a Support Vector Regression (SVR) model. The proposed method has been evaluated through three common used datasets. The achieved results show its relevance in terms of degradation identification and quality estimation.

1 Introduction

L'estimation de la qualité des images est un des éléments essentiels pour certaines applications de vision par ordinateur qui peuvent être impactées par la présence de dégradations. Dans [1], une centaine de mesures de qualité ont été répertoriées dont l'utilisation est liée à l'application. Lorsque l'image originale est supposée disponible, les mesures avec référence (FR) sont utilisées. On parle alors plutôt de fidélité d'images. Cependant, lorsque l'application n'a pas accès à cette information, les mesures sans référence (NR) seront utilisées.

Dans cet article, nous proposons un schéma d'estimation de la qualité d'images sans référence basé sur deux étapes principales. La première étape vise à identifier le type de dégradation en utilisant un réseau de neurones convolutionnel (CNN), tandis que la seconde étape vise à prédire le score de qualité subjectif en combinant certaines mesures, sélectionnées en fonction du type de dégradation identifié. Il convient de noter que généralement le type de dégradation n'est pas explicitement considéré dans le processus d'estimation de la qualité. Dans [2], les auteurs proposent de calculer certains indices de qualité (un par type de dégradation) et de les combiner à l'aide d'un modèle SVR (Support Vector Machine). Des approches similaires ont été aussi proposées dans [3,4]. Dans [5], un modèle Perceptron multiple couches (MLP) a été utilisé. Dans [6], un modèle CNN multitâches a été proposé pour estimer à la fois la qualité de l'image et son type de dégradation. Cependant, toutes ces études considèrent implicitement le type de dégradation comme une information primordiale dans le processus d'estimation de la qualité sans véritablement l'exploiter.

Ainsi, ce travail se distingue par deux contributions majeures: la première contribution est l'utilisation d'un modèle CNN pour l'intégration du type de dégradation dans le processus d'estimation de la qualité d'images. La seconde contribution est la sélection et la combinaison de certaines mesures spécifiques selon le type de dégradation identifié.

Notre article est organisé comme suit: La section 1 est dédiée à description de la méthode proposée. Dans la section 2, nous présentons les résultats obtenus en termes d'identification de la dégradation et en termes de corrélation avec les jugements subjectifs. La dernière section est consacrée à la conclusion et aux perspectives.

2 Méthode proposée

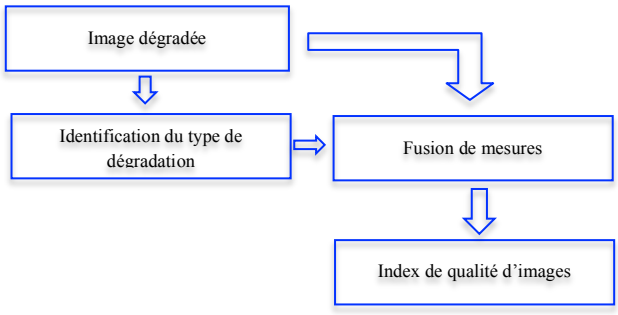

Figure 1: Schèma synoptique de la méthode proposée

Comme le montre la Fig. 1, nous proposons d'estimer la qualité d'une image en deux étapes principales. La première étape consiste à identifier le type de dégradation contenu dans l'image via un modèle CNN, tandis que la seconde étape permet d'estimer sa qualité en combinant plusieurs mesures via un modèle SVR.

Dans [7], un modèle CNN a été proposé pour estimer la qualité d'une image sans référence. Les auteurs

proposent de faire un apprentissage avec en entrée une sous-image (taille 32x32) et la note moyenne subjective de l'image (MOS : Mean Opinion Score) comme cible de sortie. Les auteurs considèrent que le MOS de chaque sous-image est égale au MOS de l'image globale. Ils justifient cela en posant comme postulat que la dégradation est homogène. Cependant, cette hypothèse forte n'est pas en conformité avec notre jugement subjectif. A titre d'exemple, une image dégradée et un échantillon de sous-images sont présentés Fig. 2. Le MOS de cette image est égale à 33.28. Si nous comparons la qualité des sous-images, nous ne pouvons pas leur donner la même note subjective (la sous-image 5 semble de moins bonne qualité que l'image 7). Dans cette étude, nous adoptons une stratégie différente qui consiste à calculer des mesures directement sur l'image globale et non à partir des sous-images.

Figure 2: Image dégradée et certaines de ses sous-images.

Dans ce qui suit, nous présentons le modèle CNN utilisé, les mesures sélectionnées ainsi que le modèle SVR.

2.1 Architecture du modèle CNN

L'architecture de notre modèle CNN est présentée Fig. 3 (128x128x1→61x61x16→27x27x16 →10x10x16 →2x2x16→200→4). L'image est décomposée en sous-images de taille 128x128. Chaque sous-image est ensuite normalisée par sa moyenne et son l'écart-type. Le modèle CNN est constitué de quatre couches de convolution, une couche entièrement connectée (FC : Fully Connected) et une couche de sortie. Chaque couche de convolution est composée de 16 noyaux de taille 7x7 (i.e. 16 cartes de caractéristiques) et d'une étape de « max-pooling 2x2» sans recouvrement. La couche FC est composée de 200 neurones, suivie d'une couche de régression logistique avec quatre sorties (types de dégradation). ***tanh*** est utilisée comme fonction d'activation avec la méthode d'optimisation « Stochastic Gradient Descent (SGD) ». Dans ce travail, le modèle proposé a été développé à l'aide de Torch [8].

2.2 Descripteurs utilisés pour chaque type de dégradation

Le type de dégradation détecté est ici utilisé pour sélectionner les mesures adéquates permettant une meilleure estimation de la qualité de l'image. Le Tableau 1 liste l'ensemble de ces mesures pour chaque dégradation.

Tableau 1. Selected features for each degradation type

Dégradation	Mesure
NOISE	Block-based [9]
	Psycho visual-based [10]
	Wavelet-based [11]
	Learning-based [2]
JP2K	Taylor decomposition-based [12]
	Natural Scene Statistics-based [13]
	Frequency-based [14]
	Learning-based [3]
JPEG	Psycho visual-based [20]
	Frequency-based [15]
	Frequency-based [16]
	Learning-based [17]
BLUR	Psycho visual-based [10]
	Learning-based [2]
	Learning-based [3]
	Learning-based [17]

Un modèle SVR est ensuite utilisé pour combiner les valeurs obtenues [18] (un modèle par type de dégradation). Les entrées de chacun des modèles dépendent ainsi du type de dégradation identifié. Plusieurs tests ont été effectués pour fixer le type de noyau (Gaussian, Radial Basis Function, Polynomial, etc.). Les meilleures performances ont été obtenues avec le noyau de type RBF Heavy-Tailed (HT-RBF).

3 Evaluation de la méthode proposée

3.1 Base d'images utilisées

Trois bases d'images ont été utilisées:

- **LIVE Image Database - Phase 2 (LIVE2) [19]:** Cette base est composée de 5 types de dégradation (JPEG2K, JPEG, White Noise, Gaussian Blur et Fast Fading) dérivés de 29 images de référence (779 images dégradées). Pour chaque image dégradée, le DMOS (Differential Mean Opinion Score) est disponible.
- **TID 2008 [20]:** composée de 17 types de dégradation obtenues à partir de 25 images originales, cette base est constituée de 1700 images dégradées ainsi que des MOS correspondants.
- **CSIQ [21]:** totalisant 866 images dégradées obtenues à partir 30 images originales, cette base est constituée de six types de dégradation et propose pour chaque image dégradée, une note subjective moyenne (DMOS).

Notez que dans cette étude, les quatre types de dégradation communs aux bases d'images utilisées ont été considérés: Bruit blanc (WN), jpeg (JPEG), jpeg2000 (JP2K) et Flou Gaussien (BLUR).

Notre méthode a été évaluée aussi bien en termes d'identification de dégradation, qu'en termes de corrélation avec le jugement subjectif. La méthode proposée a également été comparée à l'état de l'art: PSNR, SSIM [22], FSIM [23], CORNIA [24], IQA-CNN [7], IQA-CNN + \ IQA-CNN ++ [6], BRISQUE [3] et BLIINDS-II [4].

3.2 Identification du type de dégradation

Dans ce premier test, la base LIVE2 est utilisée. Elle est décomposée de 2 jeux de données : apprentissage (80%) et test (20%), sélectionnés aléatoirement sans recouvrement. Pour assurer la généralisation des résultats obtenus, la procédure a été appliquée 100 fois.

Le tableau 2 montre les taux moyens de bonne classification obtenus. La méthode proposée obtient un taux de bonne classification supérieur à toutes les autres méthodes avec peu d'erreurs.

Tableau 2. Pourcentage moyen de bonne classification: LIVE2

	Pourcentage moyen de bonne classification
BLIINDS-II	0.838
BRISQUE	0.886
CORNIA	0.875
IQA-CNN+	0.921
IQA-CNN++	0.951
Méthode proposée	**0.993**

Tableau 3. Pourcentage moyen de bonne classification: TID 2008 et CSIQ

Méthode	**Taux moyen de bonne classification**	
	TID 2008	**CSIQ**
CORNIA	0.920	0.768
IQA-CNN+	0.890	0.730
IQA-CNN++	0.933	0.783
Méthode proposée	**0.998**	**0.917**

Nous avons ensuite testé la méthode en utilisant l'ensemble de la base LIVE2 pour l'apprentissage et les bases TID 2008 et CSIQ pour le test. Les résultats obtenus sont présentés tableau 3. Le taux de classification obtenu pour la base TID 2008 est égal à 99,8% et est égal à 91,7% pour la base CSIQ. La méthode proposée atteint également la meilleure performance avec une nette amélioration pour la base CSIQ.

Tableau 4. Matrice de confusion obtenue pour la base TID 2008 (%)

	JP2K	**JPEG**	**WN**	**BLUR**
JP2K	**100**	0	0	0
JPEG	0	**100**	0	0
WN	0	0	**100**	0
BLUR	1	0	0	**99**

Tableau 5. Matrice de confusion obtenue pour la base CSIQ (%)

	JP2K	**JPEG**	**WN**	**BLUR**
JP2K	**93.33**	0	0.67	6
JPEG	9.33	**86**	2	2.67
WN	0	0	**100**	0
BLUR	6	0	0	**94**

Les tableaux 4 et 5 montrent les matrices de confusion obtenues. Certaines confusions sont observées, en particulier pour la base CSIQ. Les confusions entre les types de dégradation JP2K et BLUR sont dues au fait que pour une forte compression, le flou est une des dégradations inhérentes aux images compressées jpeg2000.

3.3 Corrélations obtenues

La procédure d'apprentissage/test décrite à la section 3.2 a été appliquée pour évaluer la prédiction des scores subjectifs. Les coefficients de corrélation de Pearson (PCC) et de Spearman (SROCC) ont ici été utilisés pour évaluer la capacité de notre méthode à prédire les scores subjectifs.

Tableau 6. LCC et SROCC obtenues pour la base LIVE2.

PCC				
	JP2K	**JPEG**	**WN**	**BLUR**
PSNR	*0.873*	*0.876*	*0.926*	*0.779*
SSIM	***0.921***	*0.955*	***0.982***	*0.893*
FSIM	*0.910*	***0.985***	*0.976*	***0.978***
DIIVINE	0.922	0.921	0.988	0.923
BLIINDS-II	0.935	0.968	0.980	0.938
BRISQUE	0.922	0.973	0.985	0.951
CORNIA	0.951	0.965	0.987	0.968
IQA-CNN	0.953	**0.981**	0.984	0.953
Our method	**0.974**	0.979	**0.991**	**0.979**
SROCC				
	JP2K	**JPEG**	**WN**	**BLUR**
PSNR	*0.870*	*0.885*	*0.942*	*0.763*
SSIM	*0.939*	*0.946*	*0.964*	*0.907*
FSIM	***0.970***	***0.981***	***0.967***	***0.972***
DIIVINE	0.913	0.910	0.984	0.921
BLIINDS-II	0.929	0.942	0.969	0.923
BRISQUE	0.914	0.965	0.979	0.951
CORNIA	0.943	0.955	0.976	0.969
IQA-CNN	0.952	**0.977**	**0.978**	0.962
Our method	**0.953**	0.952	0.976	**0.975**

Dans le tableau 6, les coefficients de corrélation obtenus sont présentés pour la base LIVE2. Les parties grisées et non-grisées correspondent respectivement aux mesures avec et sans référence. La mesure ayant obtenue le meilleur score pour chaque approche est représentée en gras. Comparée aux mesures sans référence, la méthode proposée obtient les meilleurs résultats pour la plupart des dégradations (en particulier pour les dégradations JP2K, WN et BLUR) avec une légère différence pour JPEG. Comparée aux mesures avec référence, les performances sont aussi compétitives.

Tableau 8. Performance globale: TID 2008 et CSIQ.

TID 2008			
	PCC	**SROCC**	**Class. (%)**
FSIM	**0.952**	**0.954**	-
CORNIA	0.890	0.880	0.920
IQA-CNN	0.903	**0.920**	-
IQA-CNN+	0.893	0.912	0.890
IQA-CNN++	0.895	0.906	0.933
Our method	**0.905**	0.899	**0.997**
CSIQ			
	PCC	**SROCC**	**Class. (%)**
FSIM	**0.961**	**0.962**	-
CORNIA	0.914	0.899	0.768
IQA-CNN	0.913	0.923	-
IQA-CNN+	0.910	0.918	0.730
IQA-CNN++	**0.928**	**0.936**	0.783
Our method	0.921	0.901	**0.917**

Le tableau 7 présente les corrélations obtenues pour les bases TID 2008 et CSIQ. Pour cette partie, la base LIVE2 a été utilisée pour l'apprentissage. Comme on

peut le constater, les corrélations obtenues sont élevées pour tous les types de dégradation. Les performances globales sont présentées dans tableau 8. La méthode IQA-CNN++ a obtenu une valeur PCC légèrement supérieure pour la base CSIQ, alors que son taux de bonne classification est beaucoup plus faible.

4 Conclusion

Un nouveau schéma d'estimation de la qualité d'image basé sur l'identification de dégradations (modèle CNN) et la fusion de caractéristiques (modèle SVR) a été proposé dans cet article. Les résultats obtenus montrent la pertinence de l'approche proposée avec une évaluation sur 3 bases d'images. Comme perspectives, nous allons essayer d'améliorer l'étape identification en ajoutant des attributs globaux et les utiliser directement lors de l'étape de combinaison.

5 Références

[1] M. Pedersen and J. Y. Hardeberg, " Survey of full-reference image quality metrics," Høgskolen i Gjøviks rapportserie ISSN: 1890-520X, 2009

[2] A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices", IEEE Signal Processing Letters, pp. 513-516, vol. 17, no. 5, 2010

[3] A. Mittal, A. K. Moorthy and A. C. Bovik, "Referenceless Image Spatial Quality Evaluation Engine," 45th Asilomar Conference on Signals, Systems and Computers, November 2011

[4] M.A Saad and A. C. Bovik, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain, " IEEE Transactions on Image Processing, pp. 1, 2012

[5] A. Chetouani, A. Beghdadi, S. Chen and G. Mostafaoui, "A free reference image quality measure using neural networks", International Workshop on Video Processing and Quality Metrics, 2010

[6] L. Kang, P. Ye, Y. Li and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," 2015 IEEE International Conference on Image Processing, pp. 2791-2795, 2015

[7] L. Kang, P. Ye, Y. Li and D. Doermann, "Convolutional Neural Networks for No-Reference Image Quality Assessment," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 1733-1740, 2014

[8] http://torch.ch/

[9] D. Van de Ville and M. Kocher, "SURE-Based Non-Local Means", IEEE Signal Processing Letters, Vol. 16, pp. 973-976, 2009

[10] N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," accepted and to appear in the IEEE Transactions on Image Processing, 2011

[11] S. Singh, V. Kumar and H.K. Verma, "Reduction of blocking artefacts in JPEG compressed images", Digital Signal Processing, Vol.17, Issue 1, pp. 225-243, 2007

[12] http://www.mathworks.com/matlabcentral/fileexchange/16683-estimatenoise, John D'Errico

[13] H.R. Sheikh, A.C. Bovik and L.K. Cormack, ''No-Reference Quality Assessment Using Natural Scene Statistics: JPEG2000,'' IEEE Transactions on Image Processing, Vol. 14, No. 12, pp. 1918-1927, 2005

[14] A. V. Murthy and L. J. Karam, "MATLAB Based Framework For Image and Video Quality Evaluation," International Workshop on Quality of Multimedia Experience (QoMEX), pages 242-247, 2010

[15] A. Chetouani, A. Beghdadi, and M. Deriche, ''A new free reference image quality index for blur estimation in the frequency domain,'' IEEE International Symposium on Signal Processing and Information Technology, pp. 155-159, 2009

[16] Z. Wang, Hamid R. Sheikh and Alan C. 'Bovik, ''No-Reference Perceptual Quality Assessment of JPEG Compressed Images,'' IEEE International Conference on Image Processing, Vol. 1, pp. 477-480, 2002

[17] L. Zhang, L. Zhang and A.C. Bovik, "A feature-enriched completely blind local image quality analyzer," IEEE Transactions on Image Processing, vol. 24, no. 8, pp. 2579-2591, 2015

[18] http://asi.insa-rouen.fr/enseignants/~arakoto/toolbox/

[19] H.R. Sheikh, Z.Wang, L. Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2", http://live.ece.utexas.edu/research/quality.

[20] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics", Advances of Modern Radioelectronics, Vol. 10, pp. 30-45, 2009

[21] E. C. Larson and D. M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," Journal of Electronic Imaging, 19 (1), March 2010.

[22] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol.13, no.4pp. 600- 612, 2004

[23] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: a feature similarity index for image quality assessment", IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, 2011

[24] P. Ye, J. Kumar, L. Kang and D. Doermann, "Unsupervised Feature Learning Framework for No-reference Image Quality Assessment", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1098-1105, 2012

Table 7. PCC et SROCC obtenues pour les bases TID 2008 et SCIQ.

Database	JP2K		JPEG		WN		BLUR	
	PCC	SROCC	PCC	SROCC	PCC	SROCC	PCC	SROCC
TID 2008	0.920	0.9190	0.950	0.927	0.903	0.902	0.848	0.846
CSIQ	0.887	0.883	0.933	0.922	0.924	0.925	0.904	0.875

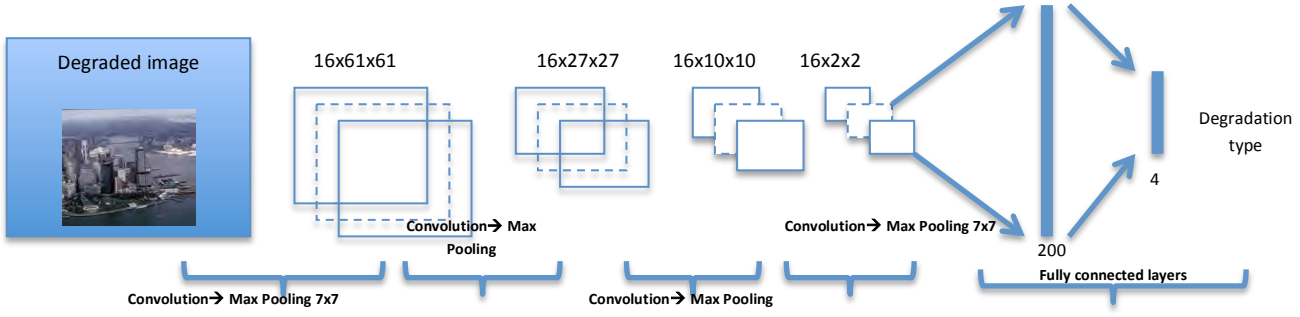

Figure 3: Architecture du modèle CNN proposée