Réseau de neurones convolutif et parcours visuel 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 méthode d'estimation de la qualité d'image sans référence basée sur la sélection de patchs saillants et sur l'utilisation d'un réseau de neurones convolutif. L'idée développée ici est de ne considérer que les patchs impactant le plus notre jugement subjectif. Pour ce faire, une méthode de prédiction du parcours visuel a été utilisée. Cette méthode vise à reproduire à partir des informations de la saillance le comportement visuel humain lorsque ce dernier analyse une image. Un réseau de neurones de convolution est ensuite utilisé pour prédire le score de qualité. La méthode proposée a été évaluée à l'aide de trois bases de données (LIVE-P2, TID 2008 et CSIQ). Les résultats obtenus ont montré une nette amélioration des performances par rapport à l'état de l'art.

Abstract - In this paper, we propose an image quality framework without reference based on selection of saliency patches and Convolutional Neural Network. The idea is here to not consider all patches of the distorted image but rather only some of them, which are considered as the more perceptually relevant and thus impact more our subjective judgment. To do so, A scanpath predictor, that aims to reproduce the human visual behavior based on the saliency information, is applied to select the more relevant patches. A Convolutional Neural Network model is then used to predict the quality score. The proposed was evaluated using four well-known datasets (LIVE-P2, TID 2008 and CSIQ) and the results obtained show improvements compared to the state-of-the-art.

1 Introduction

La qualité des images est un des éléments essentiels pour le bon fonctionnement des applications de vision par ordinateur. En effet, les performances de ces méthodes peuvent être impactées par la qualité des données utilisées. Dans [1], une centaine de mesures de qualité ont été répertoriées. 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) sont employées.

Dans [2], un réseau de neurones convolutif (CNN) a été utilisé pour estimer la qualité des images sans référence. Les auteurs proposent de faire un apprentissage avec comme entrée du CNN des patchs de taille 32x32 et la note moyenne subjective de l'image (MOS : Mean Opinion Score) comme sortie désirée. Les auteurs supposent que la dégradation est homogène et considèrent que le MOS de chaque sous-image est égale au MOS de l'image globale. Cependant, cette hypothèse n'est pas en conformité avec notre jugement subjectif. A titre d'exemple, une image dégradée et ses sous-images sont présentées Fig. 1. En comparant les différentes sous-images, on s'aperçoit que chaque sous images a une qualité différente (la sous-image 5 est de moins bonne qualité que la sous-image 7).

Dans cet article, nous proposons un schéma d'estimation de la qualité d'image sans référence basé sur deux étapes principales. La première étape vise à sélectionner les régions de l'image les plus perceptuellement pertinentes tandis que la seconde étape vise à prédire le score de qualité subjectif via un CNN.

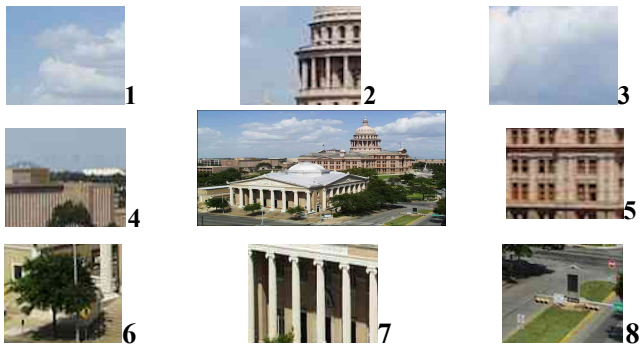

Figure 1: Image dégradée et ses sous-images.

Notre article est organisé comme suit: La section 2 est dédiée à la description de la méthode proposée. Dans la section 3, nous présentons les résultats obtenus en termes de corrélations avec les jugements subjectifs. La dernière section est consacrée à la conclusion et aux perspectives.

2 Méthode proposée

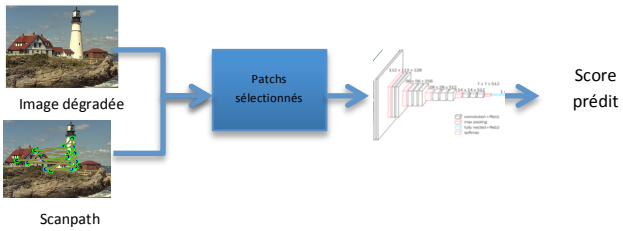

Figure 2: Schèma synoptique de la méthode proposée

Comme le montre la Fig. 2, nous proposons d'estimer la qualité d'une image en deux étapes principales. La première étape consiste à identifier les régions pertinentes en exploitant l'image de la saillance visuelle, tandis que la seconde étape permet d'estimer sa qualité par le biais d'un modèle CNN.

2.1 Sélection des régions pertinentes

L'attention visuelle joue un rôle important dans le domaine de l'analyse d'images et a été exploitée dans plusieurs applications (récupération d'images [3], localisation à l'intérieur [4], etc.). Les régions détectées, appelées régions saillantes, représentent les zones les plus attractives de l'image et jouent donc un rôle important dans la compréhension de l'image. Nous proposons ici d'exploiter ces informations en extrayant les patchs les plus pertinents et de les utiliser comme entrée d'un modèle CNN. L'idée sous-jacente développée ici est que ces régions ont un impact important sur le jugement subjectif et donc sur la qualité globale [5,6]. Cette procédure permet de ne pas prendre en compte les correctifs ayant un impact limité en termes de qualité.

Dans cette étude, nous avons utilisé la méthode proposée dans [7] pour sélectionner les patchs les plus pertinents. Cette dernière prédit le parcours visuel (*scanpath* en anglais) des observateurs sur la base d'un modèle de saillance et de biais (biais d'amplitude de saccade et d'orientation de saccade). La Fig. 2.d présente le parcours visuel de l'image de la Fig. 2.a. Chaque position (point bleu) du parcours visuel représente l'une des régions de saillance les plus pertinentes. Comme nous pouvons le constater, les positions sélectionnées sont localisées sur les régions les plus attractives de l'image (phare et maisons) et sont donc en accord avec notre perception. Ce résultat est obtenu à partir de la carte de saillance de l'image (Fig. 2.b).

Parmi les premières méthodes proposées dans la littérature, la méthode de Itti et al. [8]. Les auteurs ont proposé de combiner différentes cartes en utilisant des attributs de bas niveau (intensité, couleur et orientation). Dans [9], une méthode plus complexe basée sur certaines caractéristiques du Système Visuel Humain (SVH) a été proposée. La carte de saillance est ici obtenue par l'application de différents modèles perceptuels (espace colorimétrique perceptuel → fonction de sensibilité au contraste [10] → transformation en cortex [11] → effet de masquage). D'autres méthodes, plus simples, ont également été proposées. Dans [12], les auteurs proposent de déterminer la saillance d'une image dans le domaine de Fourier. Le spectre résiduel de l'image est d'abord calculé et est soustrait à sa version filtrée. La carte de saillance est ensuite obtenue par la transformée de Fourier inverse. Dans ce travail, la méthode GBVS (Graph-Based Visual Saliency) [13], qui est l'une des méthodes performantes de l'état de l'art [14], a été utilisée.

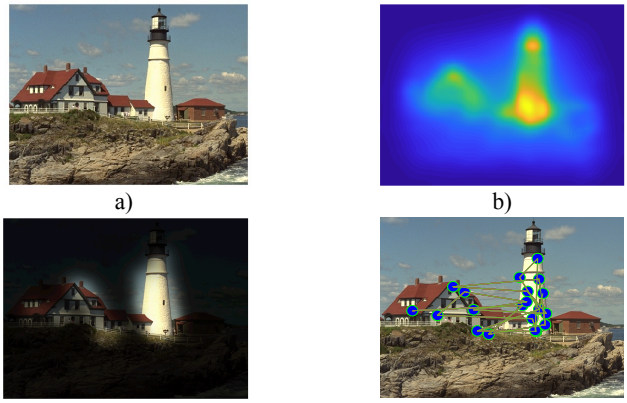

Figure 2: Schèma synoptique de la méthode proposée

Pour chaque position du parcours visuel, un patch de taille 32x32x3 est extrait et est normalisé (i.e. moyenne locale = 0 et écart type local = 1, taille du filtre = 3x3).

2.2 Modèles CNN utilisés

Un modèle CNN a été ensuite utilisé pour prédire les scores de qualité. Le modèle doit avoir une entrée de taille 32x32x3 et une sortie correspondante au score de qualité. Plusieurs modèles ont été proposés dans la littérature avec différentes architectures. Certains auteurs ont proposé leurs propres modèles [15], tandis que d'autres exploitent des modèles pré-entrainés. Dans cette étude, le modèle pré-entrainé VGG a été adapté à notre contexte et comparé à l'état de l'art. Ce modèle, Proposé en 2014, a été développé par le groupe *Oxford Visual Geometry Group* [16]. Pour augmenter la capacité du modèle CNN à discriminer les objets, les auteurs ont intégré plus de fonctions d'activation Relu en utilisant des couches de convolution avec des filtres 3x3 au lieu de filtres 7x7. Cela permet de diminuer le nombre de paramètres et d'augmenter le nombre d'unités d'activation. Plusieurs versions ont été proposées avec 11 (VGG11), 13 (VGG13), 16 (VGG16) et 19 (VGG19) couches. Le modèle VGG16 a été ici choisi.

Initialement entrainé sur la base de données ImageNet, ce modèle a été ajusté (*fine tuning*) pour adapter leurs paramètres à notre contexte. Il est à noter que la taille des patchs du modèle VGG16 est initialement de 224x224x3. La taille d'entrée a donc été adaptée à nos patchs (32x32x3). La partie FC a aussi été remplacée par 2 couches FC de taille 128 et d'une fonction d'activation ReLu. Nous avons également ajouté une couche de régression logistique avec une sortie (MOS prévue). L'erreur quadratique moyenne a été ici utilisée comme fonction de perte.

2.3 Bases de données utilisées

Trois bases d'images ont été utilisées pour évaluer la méthode proposée:

- **LIVE Image Database - Phase 2 (LIVE2) [17]:** Cette base est composée de 5 types de dégradation (JPEG2K, JPEG, White Noise, Gaussian Blur et Fast Fading). Elle est constituée de 779 images dégradées dérivés de 29 images de référence. Pour chaque image dégradée, le DMOS (Differential Mean Opinion Score) est disponible.
- **TID 2008 [18]:** composée de 17 types de dégradation obtenues à partir de 25 images originales, cette base de données est constituée de 1700 images dégradées et des MOS correspondants.
- **CSIQ [19]:** totalisant 866 images dégradées obtenues à partir 30 images originales, cette base de données est constituée de six types de dégradation et fournie pour chaque image dégradée, une note subjective moyenne (DMOS).

3 Evaluation de la méthode proposée

Les corrélations de coefficient de Pearson (PCC) et de Spearman (SROCC) ont été ici utilisées pour évaluer la capacité de notre méthode à prédire des jugements subjectifs. La meilleure performance est représentée en gras sur fond gris.

Les résultats ont été comparés à l'état de l'art en intégrant des métriques FR et NR traditionnelles (PSNR, SSIM [20], FSIM [21], DIIVINE [22], BLIINDS-2 [23], BRISQUE [24]. et CORNIA [25]) ainsi que des métriques basées sur l'utilisation des CNNs (DeepIQA [26], IQA-CNN [2], IQA-CNN + / IQA-CNN ++ [27], SOM [28], CNN-Prewitt [29] et Image-wise CNN [30]).

3.1 Evaluation sur une seule base de données

Dans ce premier test, la base LIVE2 a été utilisée. Elle a été décomposée de 2 jeux de données : apprentissage-validation (60%-20%) et test (20%), sélectionnés aléatoirement sans recouvrement. Pour assurer la généralisation des résultats obtenus, la procédure a été appliquée 10 fois.

Tableau 1. Corrélations moyennes obtenues pour la base LIVE2 après 10 tirages aléatoires.

LIVE-P2			
		PCC	**SROCC**
FR-IQM	**PSNR**	0.856	0.866
	SSIM [24]	0.906	0.913
	FSIM [25]	0.960	0.964
	DeepIQA [26]	0.981	0.982
NR-IQM	**DIIVINE [27]**	0.917	0.916
	BLIINDS-II [28]	0.930	0.931
	BRISQUE [29]	0.942	0.940
	CORNIA [30]	0.935	0.942
	IQA-CNN [3]	0.953	0.956
	IQA-CNN+ [4]	0.953	0.953
	IQA-CNN++ [4]	0.950	0.950
	SOM [31]	0.962	0.964
	CNN-Prewitt [32]	0.966	0.958
	Image-wise CNN [6]	0.963	0.964
	Our method	**0.986**	**0.983**

Le tableau 1 présente les corrélations moyennes obtenues pour la base LIVE2. La méthode proposée surpasse l'ensemble des méthodes de l'état de l'art quelque soit l'approche (avec et sans référence). La méthode FR DeepIQA obtient des corrélations proches. Cependant, cette dernière nécessite l'image originale pour estimer la qualité. On s'aperçoit aussi que les métriques les plus performantes sont les méthodes basées sur l'utilisation de modèles CNNs. De plus, la méthode nommée IQA-CNN, qui utilise l'ensemble des patchs de l'image pour estimer la qualité, obtient des corrélations (0.953) très en-dessous de celles obtenues par notre méthode (0.987). Outre les modèles utilisés, la différence majeure entre ces deux méthodes est la sélection ou non des patchs. Ainsi, l'étape de sélection de patchs est une étape essentielle qui permet une nette amélioration des performances.

3.2 Validation croisée sur les bases de données

Afin d'évaluer la généralisation de la méthode proposée, la validation croisée a été appliquée. Pour ce faire, nous avons utilisé la base LIVE2 pour l'apprentissage et les bases TID 2008 et CSIQ pour le test. Les résultats obtenus sont présentés dans les tableaux 2 et 3.

De même que pour la base LIVE2, la méthode proposée a obtenu les meilleures performances pour la base TID08 (voir Tableau 2). La métrique FSIM qui est une mesure traditionnelle avec référence a obtenu des corrélations élevées proche de ceux de notre méthode. Comparé aux mesures basées sur les CNNs, les corrélations obtenues sont très supérieures à ceux de l'état de l'art.

Tableau 2. Corrélations obtenues pour la base TID08. La base LIVE2 a été utilisée pour l'apprentissage-validation et la base TID08 pour le test.

TID08			
		PCC	**SROCC**
FR-IQM	**PSNR**	0.776	0.901
	SSIM	0.817	0.903
	FSIM	0.952	0.954
NR-IQM	**CORNIA**	0.890	0.880
	IQA-CNN	0.903	0.920
	IQA-CNN+	0.893	0.912
	IQA-CNN++	0.895	0.906
	SOM	0.899	0.923
	Our method	**0.954**	**0.972**

Pour ce qui est de la base CSIQ (voir Tableau 3), les mesures avec référence FSIM et DeepIQA ont obtenu les meilleures corrélations. Comparé aux mesures sans référence, nous avons obtenu la meilleure corrélation PCC (0.931) et la seconde meilleure corrélation SROCC (0.933). De même que pour les deux bases d'images précédentes, l'étape de sélection de patchs a permis d'améliorer considérablement les performances.

Tableau 3. Corrélations obtenues pour la base CSIQ. La base LIVE2 a été utilisée pour l'apprentissage-validation et la base CSIQ pour le test.

CSIQ			
		PCC	**SROCC**
FR-IQM	**PSNR**	0.800	0.806
	SSIM	0.861	0.876
	FSIM	0.961	**0.962**
	DeepIQA	**0.964**	0.960
NR-IQM	**BRISQUE**	0.797	0.756
	CORNIA	0.914	0.899
	IQA-CNN	0.903	0.923
	IQA-CNN+	0.910	0.918
	IQA-CNN++	0.928	0.936
	Our method	0.931	0.933

4 Conclusion

Dans cet article, nous avons proposé une méthode permettant d'estimer la qualité des images en sélectionnant les patchs les plus pertinents perceptuellement. La sélection a ici été réalisée à l'aide d'une méthode de prédiction du parcours visuel qui exploite les informations de saillance. Le modèle VGG16 avec *fine tuning* a été utilisé. Les résultats obtenus ont été comparés à l'état de l'art et montrent la pertinence de notre approche.

Comme perspective, nous allons essayer d'utiliser différentes méthodes de saillance et comparer les variations de performance en fonction de cette entrée.

5 Références

[1] M. Pedersen and J.Y. Hardeberg, " Full-reference image quality metrics: Classification and evaluation", Foundations and Trendsin Computer Graphics and Vision, pp. 1—80, 2012

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

[3] C. A. Hussain, D. V. Rao and S. A. Masthani, "Robust Pre-processing Technique Based on Saliency Detection for Content Based Image Retrieval Systems", In Procedia Computer Science, Volume 85, pp. 571-580, 2016

[4] W. Elloumi, K. Guissous, A. Chetouani and S. Treuillet, "Improving a vision indoor localization system by a saliency-guided detection", IEEE VCIP, pp. 149-152, 2014

[5] D. V. Rao, N. Sudhakar, I. R. Babu, and L. P. Reddy, "Image quality assessment complemented with visual region of interest," in Proc. Int. Conf. Comput.: Theory Applicat., pp. 681–687, 2007

[6] Q. Ma and L. Zhang, "Image quality assessment with visual attention," IEEE ICPR, pp. 1–4, 2008

[7] 0. Le Meur and Liu Z., "Saccadic model of eye movements for free-viewing condition", Vision Research, 2015

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11): p. 1254-1259, 1998

[9] O. Le Meur, P. Le Callet, D. Barba and D. Thoreau, "A coherent computational approach to model the bottom-up visual attention", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 28, N°5, 2006

[10] A.B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models", Optics Express, pp.12–33, 2000

[11] A.B. Watson. The Cortex transform: rapid computation of simulated neural images. Computer Vision Graphics and Image Processing, pp. 311–327, 1987

[12] X. Hou, L. Zhang, "Saliency Detection: A Spectral Residual Approach", Computer Vision and Pattern Recognition, 2007

[13] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency", Neural Information Processing Systems, 2006

[14] http://saliency.mit.edu/

[15] S. Jia and Y. Zhang, " Saliency-based deep convolutional neural network for no-reference image quality assessment", Multimedia Tools and Applications, 10.1007/s11042-017-5070-6, 2017

[16] K. Simonyan and A. Zisserman, " Very deep convolutional networks for large-scale image recognition", CoRR, abs/1409.1556, 2014

[17] 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.

[18] 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

[19] 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

[20] 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

[21] 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

[22] A. K. Moorthy and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality", IEEE Transactions on Image Processing, 2011

[23] 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

[24] 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

[25] 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

[26] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang and A. C. Bovik, "Deep Convolutional Neural Models for Picture-Quality Prediction: Challenges and Solutions to Data-Driven Image Quality Assessment," in *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 130-141, 2017

[27] 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

[28] Peng Zhang, Wengang Zhou, Lei Wu and Houqiang Li, "SOM: Semantic obviousness metric for image quality assessment," IEEE Conference on Computer Vision and Pattern Recognition (CVPR*)*, pp. 2394-2402, 2015

[29] J. Li, L. Zou, J. Yan, D. Deng, T. Qu, and G. Xie, "No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks", Signal, Image and Video Processing. Vol. 10, 10.1007/s11760-015-0784-2, 2015

[30] S. Bianco, L. Celona, P. Napoletano and R. Schettini, "On the Use of Deep Learning for Blind Image Quality Assessment", Computer Vision and Pattern Recognition, 2016