Exploration de l'impact de la normalisation sur la performance de l'évaluation de la qualité basée sur des réseaux de neurones convolutionnels

Abderrezzaq Sendjasni, David Traparic, Mohamed-Chaker Larabi

CNRS, Univ. Poitiers, XLIM, UMR 7252, France
{abderrezzaq.sendjasni,david.traparic,chaker.larabi}@univ-poitiers.fr

Résumé – L'entraînement des réseaux de neurones convolutionnels (CNN) pour l'évaluation de la qualité des images requiert souvent une étape de normalisation des images d'entrées. Même si la normalisation des images améliore l'entraînement du modèle en mettant en évidence des caractéristiques importantes, elle peut engendrer une perte d'informations. Afin de comprendre son intérêt dans ce contexte, nous proposons une étude comparative pour examiner son effet sur les performances du modèle, et de déterminer les méthodes les plus adaptées à l'évaluation de la qualité. Ainsi, les performances de neuf méthodes sont étudiées et comparées statistiquement à trois méthodes basiques. L'application de la normalisation s'avère statistiquement significative sur trois bases de données et une amélioration des performances a pu être observée.

Abstract – Prior to training convolutional neural networks (CNNs) for image quality assessment (IQA), input normalization is sometimes recommended and sometimes not, according to the literature. Although input normalization is known to improve model training and helps in learning important features, it may result in the loss of information. To better explore this issue, we conduct an empirical study to first investigate the effect of normalization on model performance and to find the best fits for IQA. The performances of the selected methods are statistically compared with three basic scaling methods. The application of normalization is found to be statistically significant on three IQA databases. The performance improvement on the overall databases, as well as per-individual degradation, is demonstrated in the experimental results.

1 Introduction

L'évaluation de la qualité des images est considérée comme l'une des tâches les plus difficiles du traitement d'images. Un effort considérable a été réalisé pour prédire avec précision la qualité des images telle qu'elle est perçue par l'observateur. Avec l'introduction des Réseaux de Neurones Convolutionnels (RNC) dans l'évaluation de la qualité, un progrès significatif a été réalisé en termes de performances et de précision. Les RNC sont des architectures entraînables inspirées par la biologie humaine. Ces architectures peuvent apprendre des caractéristiques invariantes à partir des scènes visuelles [1]. Avec leurs capacités d'apprendre des caractéristiques hiérarchiques à plusieurs niveaux et leur efficacité à représenter ces caractéristiques, ils contribuent fortement au développement de modèles performants. Cet objectif peut être garanti grâce à (i) un traitement optimal des images d'entrées, (ii) une architecture adéquate et (iii) une stratégie d'entraînement appropriée.

En général, les images d'entrée des RNC sont pré-traitées pour assurer une meilleure représentation des données. L'objectif principal du pré-traitement, comme la normalisation, est d'améliorer le contenu de l'image en mettant en évidence les composantes visuelles spécifiques qui contribuent à l'apprentissage de la tâche traitée [2]. La normalisation avant la phase d'entraînement est fortement encouragée pour aider le modèle à apprendre les informations utiles. Dans le cas de la qualité, la normalisation consiste principalement à mettre en évidence les informations de haute fréquence [3] au détriment des basses fréquences, car ces dernières sont moins affectées par les dégradations et sont donc moins perceptibles par le Système Visuel Humain (SVH). Dans la littérature, on peut trouver plusieurs méthodes de normalisation adoptées pour l'évaluation de la qualité. En particulier, Kang *et al.* [4] ont utilisé une méthode basée sur la normalisation divisive, développée comme un calcul canonique implémenté dans le néo-cortex [5], et utilisée pour expliquer les réponses des neurones dans le cortex visuel primaire. Cette méthode est appelée normalisation du contraste local et a été adoptée dans plusieurs travaux par la suite [6–8]. Kim *et al.* [3] a utilisé un simple filtrage des basses fréquences pour ne retenir que les composantes hautes fréquences. Bien que la normalisation des images présente des avantages indéniables, la contrepartie réside dans la perte d'une partie des informations comme les changements de luminance et de contraste. Bosse *et al.* [9] ont mis l'accent sur le fait que ces informations perdues peuvent permettre au modèle d'apprendre des informations utiles et complémentaires.

Les opinions contradictoires sur l'utilisation de la normalisation et la diversité des méthodes utilisées dans les métriques de qualité créent de la confusion sur leur intérêt. Ce travail propose d'évaluer l'intérêt de ce pré-traitement et de fournir des recommandations relatives à l'évaluation de la qualité sur onze méthodes de la littérature. Pour effectuer ce benchmark, un mo-

dèle VGG-16 [10] est sélectionné et ré-entraîné sur trois bases d'images, à savoir CSIQ [11], LIVE [12], et TID2013 [13]. Afin de réduire la sur-adaptation (over-fitting) du modèle, un schéma d'apprentissage basé sur les patchs est adopté, augmentant ainsi les exemples d'apprentissage.

TABLE 1 – Liste des approches de normalisation utilisées.

Nom de l'approche de normalisation	Code
Mise à l'échelle des valeurs des pixels [14]	N1
Standardisation [14]	N2
Centrage par la moyenne [14]	N3
Normalisation du contraste local - LCN [4]	N4
Soustraction basse fréquence - LFS [3]	N5
Normalisation de réponses locales - LRN [15]	N6
Différence de gaussienne - DoG [16]	N7
Analyse des composants à zéro-phase - ZCA [14]	N8
Blanchiment simplifié [14]	N9
Égalisation des histogrammes - HE [16]	N10
Égalisation adaptative d'histogramme à contraste limité [16]	N11

2 Expérimentations

Dans cette étude, nous avons adopté l'apprentissage par patchs de 64×64 à partir de chaque image d'entrée. En raison de l'indisponibilité du score d'opinion moyen (MOS) pour les patchs individuels, chaque patch P_I extrait à partir de l'image I est annoté avec le MOS associé à son image d'origine. La normalisation est appliquée aux patchs individuels plutôt qu'aux images entières. En faisant ainsi, nous tenons compte de la luminance et du contraste locaux, qui peuvent varier dans différentes parties de l'image. De plus, le modèle n'a accès qu'aux patchs individuels et non à l'ensemble de l'image.

Afin d'évaluer de manière exhaustive l'impact de la normalisation des images d'entrées pour l'entraînement des CNNs pour l'évaluation de la qualité, nous avons sélectionné onze méthodes allant de la simple représentation des valeurs des pixels en termes d'échelle et de distribution à la normalisation du contraste et du contenu structurel. Le tableau 1 donne la liste des méthodes évaluées.

2.1 Architecture du Modèle

Le modèle convolutionnel utilisé dans cette étude est basé sur l'architecture des VGG-Nets [10]. Ce choix est motivé par le succès de cette architecture dans diverses tâches de traitement d'images. Par conséquent, nous avons affiné et utilisé le VGG-16 comme extracteur de caractéristiques visuelles (ECV). Nous avons ajusté la taille d'entrée du modèle à 64×64. De plus, un bloc de régression (BR) est ajouté à l'ECV pour régresser les caractéristiques extraites en un score de qualité. Les sorties de l'ECV $F_{W \times H \times C}$, où H, W et C représentent la hauteur, la largeur et la dimension, passent d'abord par une couche Mutualisation basée moyenne globale (GAP) afin de réduire leurs dimensions spatiales. La couche GAP produit un vecteur F' de taille $1 \times 1 \times C$. Ce dernier est fourni au BR. Ce dernier est composé d'une couche FC avec une dimension de 512, suivie d'une fonction d'activation ReLU (rectified linear unit) et d'une couche dropout. Une dernière couche FC avec un seul nœud et une activation linéaire est ajoutée pour délivrer un score de qualité.

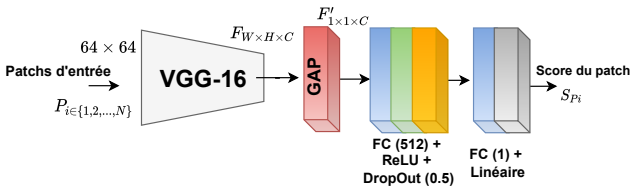


FIGURE 1 – L'architecture du modèle utilisé.

2.2 Protocol expérimental

L'étude proposée est implémentée à l'aide de TensorFlow. Le modèle est entraîné en utilisant une taille de batch de 128 avec l'optimiseur Adam [17]. Les paramètres de ce dernier sont choisis selon [9], à savoir $\beta_1 = 0,9$, $\beta_2 = 0,999$, $\epsilon = 10^{-8}$, et $\alpha = 10^{-4}$. Une validation croisée de cinq itérations est adoptée pour une évaluation complète sur chaque base d'images. Le modèle est entraîné pendant 300 époques et pour éviter un sur-apprentissage, nous avons utilisé l'arrêt conditionné de l'entraînement une fois qu'aucune amélioration n'est observée. Pendant l'apprentissage, les bases d'images sont divisées en deux parties : 80% pour l'apprentissage et 20% pour le test. Pour garantir une séparation optimale, les images dégradées associées à la même image d'origine sont attribuées au même ensemble.

Pour évaluer de manière exhaustive les performances des méthodes de normalisation utilisées, nous avons sélectionné trois bases d'images à savoir CSIQ, LIVE et TID2013.

Les performances sont étudiées en utilisant la corrélation de Pearson (PLCC), le coefficient de Spearman (SRCC) et l'erreur quadratique moyenne (RMSE). De plus, nous avons utilisé la méthode de Krasula *et al.* [18] pour analyser la signification statistique entre les méthodes considérées.

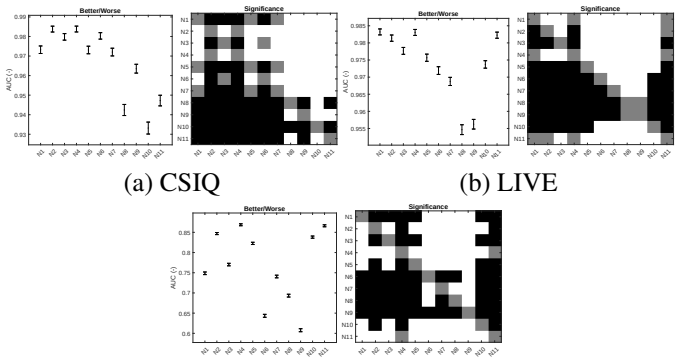
2.3 Performances globales

Le tableau 2 résume l'ensemble des performances des différentes méthodes de normalisation. Les meilleures performances sont obtenues sur LIVE, suivi de CSIQ puis TID2013. Cette dernière est assez complexe, car elle contient 24 dégradations, ce qui rend la capacité de généralisation des modèles de convolutions moins robuste. En ce qui concerne la meilleure normalisation, N4 surpasse nettement les autres méthodes pour toutes les bases d'images. cela démontre l'intérêt de la normalisation des entrées du modèle. Les méthodes N7, N8 et N9, qui ne sont pas spécifiquement conçues pour la qualité, ont obtenu des résultats médiocres. Enfin, malgré leur simplicité, N1, N2 et N3, obtiennent des performances compétitives en comparant avec les méthodes plus élaborées.

TABLE 2 – Performances en fonction des méthodes de normalisation. Meilleure perf. en **gras** et la suivante soulignée.

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11
						CSIQ					
PLCC	0.9009	0.9060	0.9139	**0.9236**	0.8737	0.9163	0.8935	0.7100	0.8498	0.8583	0.8893
SRCC	0.8470	0.8493	0.8686	**0.8932**	0.8050	0.8837	0.8272	0.7379	0.8165	0.7924	0.8158
RMSE	0.1105	0.1081	0.1139	**0.1031**	0.1227	0.1065	0.1147	0.1758	0.1424	0.1306	0.1165
						LIVE					
PLCC	0.9536	0.9400	0.9401	**0.9569**	0.9440	0.9399	0.9235	0.9175	0.9119	0.9413	0.9538
SRCC	0.9463	0.9324	0.9392	**0.9528**	0.9430	0.9390	0.9194	0.9081	0.9060	0.9402	0.9466
RMSE	0.0480	0.0543	0.0557	**0.0477**	0.0548	0.0567	0.0638	0.0661	0.0660	0.0561	0.0478
						TID2013					
PLCC	0.6520	0.6780	0.6841	**0.7406**	0.6686	0.6728	0.6265	0.6354	0.6739	0.6969	0.6798
SRCC	0.5424	0.6150	0.6014	**0.6348**	0.5581	0.5925	0.5407	0.5607	0.5729	0.5531	0.5703
RMSE	0.1043	0.1006	0.0990	**0.0953**	0.1044	0.1013	0.1068	0.1057	0.1006	0.0986	0.1008

2.4 Signification statistique

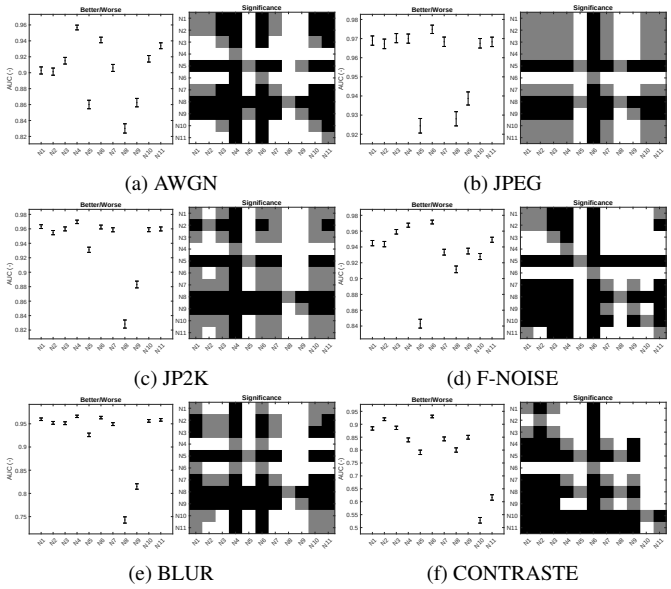


(a) CSIQ (b) LIVE

(c) TID2013

FIGURE 2 – Signification statistique globale. Un carré blanc/noir : ligne meilleure/mauvaise que colonne ; carré gris : statistiquement indiscernable.

Nous présentons dans la Fig. 2 l'analyse de signification statistique globale sur CSIQ, LIVE et TID2013. À gauche, nous indiquons la meilleure/mauvaise méthode en termes de classification de la qualité. Cela permet de savoir combien de fois le modèle reconnaît correctement le stimulus de meilleure qualité. La signification statistique entre les méthodes considérées (à droite) est fournie pour déterminer si la différence de performances est statistiquement significative. Comme on peut le voir, des résultats différents sont obtenus avec chaque normalisation, ce qui montre l'influence de chaque méthode sur les performances finales. Dans l'ensemble, N2 et N4 se démarquent des autres méthodes. Sur CSIQ, on peut remarquer que ces deux méthodes surpassent les autres tout en étant statistiquement indiscernables l'une de l'autre. Sur LIVE, N1, N2 et N11 ont obtenu les meilleures performances. Enfin, sur TID2013, on trouve N4 et N11 qui surpassent les autres méthodes. Comme N2 est essentiellement une simple représentation des valeurs des pixels, on pourrait conclure que l'entraînement des CNNs sans normalisation serait efficace. Cependant, N4 et N11, qui représentent le LCN et le CLAHE, semblent améliorer les performances du modèle en étant statistiquement supérieures aux méthodes de normalisation de base, *i.e.* N1, N2 et N3. N7 à N10 semblent obtenir des performances médiocres par rapport aux autres, tandis que la LCN a obtenu les meilleures performances. Cela démontre son efficacité et explique sa popularité dans la littérature.



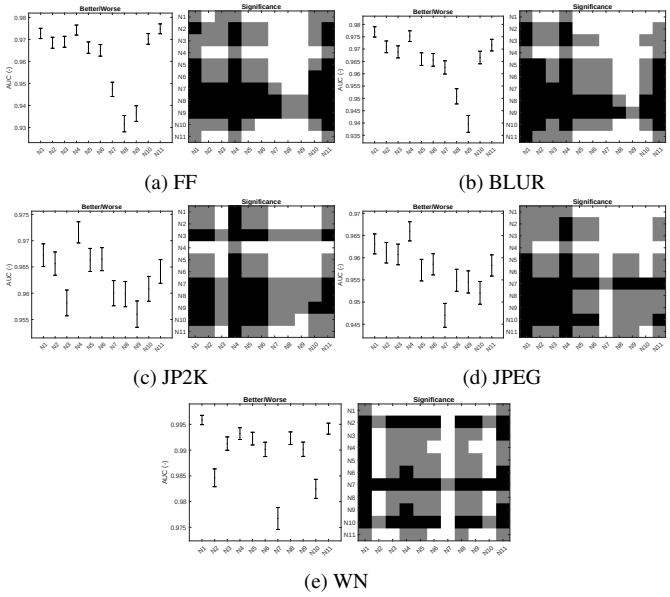
(a) AWGN (b) JPEG

(c) JP2K (d) F-NOISE

(e) BLUR (f) CONTRASTE

FIGURE 3 – Signification statistique par dégradation sur CSIQ.

L'analyse de la signification statistique globale a mis en évidence les performances de LCN, de la standardisation et la CLAHE par rapport aux autres méthodes. Cependant, si l'on considère chaque dégradation séparément, on obtient des résultats intéressants. La signification statistique en termes de meilleure/mauvaise et la signification par dégradation sont fournies dans la Fig. 3 pour CSIQ et 4 pour LIVE. Malheureusement, nous n'avons pas pu inclure TID2013 dans cette analyse en raison de la limitation des pages. Sur CSIQ, on peut remarquer que N6 semble être meilleure que toutes les autres méthodes avec JPEG, F-Noise, et Contraste. Cette méthode donne de bons résultats localement pour certaines dégradations et des résultats médiocres globalement. N4 a obtenu les meilleurs résultats avec AWGN, JP2K, et BLUR, et des résultats compétitifs avec N6 sur F-NOISE. Cela montre l'efficacité de LCN par dégradation. Cependant, sur la dégradation du contraste, la LCN a obtenu des performances médiocres. Ceci est principalement dû au fait que les changements de contraste ne sont pas retenus lors de la normalisation des images à l'aide de LCN, ce qui conduit à de mauvaises performances avec cette dégradation. On peut également observer que des performances satisfaisantes sont obtenues avec les méthodes standard N1, N2, et N3 sur les dégradations JP2K et BLUR. Ces performances n'ont pas été suffisantes pour surpasser N4 et N6, soutenant l'idée d'effectuer une normalisation appropriée avant l'entraînement du RNC. En ce qui concerne les pires performances, N8 et N9 ont obtenu des résultats médiocres parmi les normalisations sélectionnées, sauf pour le contraste. Sur ce dernier, N10 et N11 ont donné les pires performances.



(a) FF (b) BLUR

(c) JP2K (d) JPEG

(e) WN

FIGURE 4 – Signification statistique par dégradation sur LIVE.

Comme sur LIVE (Fig. 4), N4 a surpassé les autres méthodes en obtenant les meilleures performances, sauf sur WN où elle a obtenu de moins bonnes performances par rapport à N1, ce qui est également reflété par l'analyse meilleure/mauvaise. En ce qui concerne les pires performances, N7-10 ne semble pas bien fonctionner sur LIVE, ce qui est également le cas avec CSIQ.

En résumé, la méthode de normalisation la plus performante sur l'ensemble des bases d'images demeure la LCN. Une méthode spécifique à l'évaluation de la qualité, ce qui a été confirmé par les performances obtenues. En analysant les résultats par dégradation, la meilleure performance est obtenue par LCN et LRN sur CSIQ et LCN sur LIVE. Cela démontre que l'utilisation d'une méthode appropriée peut améliorer les performances du modèle convolutionnel.

3 Conclusion

Dans ce papier, nous avons mené une analyse statistiques sur les performances des méthodes de normalisation des images avant l'entraînement des RNC pour l'évaluation de la qualité. Plusieurs méthodes sont étudiées et les résultats globaux sur trois bases d'images montrent que la normalisation améliore les performances du modèle. L'analyse de la signification statistique a révélé les mêmes résultats. La méthode LCN a surpassé les autres méthodes sur toutes les bases et les dégradations, excepté pour le contraste. Dans ce dernier cas, la perte d'information a affecté les performances. Selon les résultats, l'utilisation d'une normalisation adéquate pour améliorer les performances des modèles RCN est favorable. La prise en compte de la perte d'information due à la normalisation pendant l'entraînement peut améliorer la robustesse du modèle.

Références

[1] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *IEEE ISCS*, Paris, France, 2010, IEEE, pp. 253–256.

[2] M. Sonka, V. Hlavac, and R. Boyle, *Image pre-processing*, pp. 56–111, Springer US, Boston, MA, 1993.

[3] K. Jongyoo, N. Anh-Duc, and L. Sanghoon, "Deep cnn-based blind image quality predictor," *IEEE Trans. on neural networks and learning systems*, vol. 30, no. 1, pp. 11–24, 2018.

[4] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *IEEE CVPR*, Columbus, OH, USA, 2014, pp. 1733–1740.

[5] DJ. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neuroscience*, vol. 9, no. 2, pp. 181–197, 1992.

[6] R. Li, H. Yang, T. Yu, and Z. Pan, "Cnn model for screen content image quality assessment based on region difference," in *IEEE ICSIP*, Wuxi, China, 2019, pp. 1010–1014.

[7] J. Kim and S. Lee, "Deep blind image quality assessment by employing fr-iqa," in *IEEE International Conference on Image Processing*, Beijing, China, 2017, pp. 3180–3184.

[8] C. Pan, Y. Xu, Y. Yan, K. Gu, and X. Yang, "Exploiting neural models for no-reference image quality assessment," in *VCIP*, Chengdu, China, 2016, pp. 1–4.

[9] S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE TIP*, vol. 27, pp. 206–219, 2018.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv :1409.1556*, 2014.

[11] E. Larson and D. Chandler, "Most apparent distortion : full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, pp. 011006, 2010.

[12] H. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "Live image quality assessment database release 2," `https://live.ece.utexas.edu/`, 2005.

[13] N. Ponomarenko, O. Ieremeiev, V. Lukin, and K. Egiazarian et al., "Color image database TID2013 : Peculiarities and preliminary results," in *IEEE EUVIP*, Paris, 2013, pp. 106–111.

[14] J. Brownlee, "Deep learning for computer vision : image classification, object detection, and face recognition in python," 2019.

[15] M. Rad, PM. Roth, and V. Lepetit, "Alcn : Adaptive local contrast normalization," *Computer Vision and Image Understanding*, vol. 194, pp. 102947, 2020.

[16] S. Pizer, E. Philip, J. Austin, and et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[17] D. Kingma and J. Ba, "Adam : A method for stochastic optimization," *arXiv preprint arXiv :1412.6980*, 2014.

[18] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma, "On the accuracy of objective image and video quality models : New methodology for performance eval.," in *QoMEX*, Portugal, 2016.