Évaluation de la qualité des images médicales basée sur un apprentissage par adaptation au domaine

Marouane TLIBA[1], Aymen SEKHRI[2], Aladine CHETOUANI[1]

[1]Laboratoire PRISME, Université d'Orléans, Orléans, France

[2]Institut National Des Télécommunications et TIC, Oran, Algérie

marouane.tliba@univ-orleans.fr , asekhri@inttic.dz
aladine.chetouani@univ-orleans.fr

Résumé – La prédiction de la qualité des contenus multimédia est souvent nécessaire dans différents domaines. Dans certaines applications, les métriques de qualité sont cruciales et ont un impact élevé car elles peuvent affecter la prise de décision comme le diagnostic à partir d'images médicales. Dans ce papier, nous nous concentrons sur ces applications en proposant un modèle efficace et peu profond pour prédire la qualité des images médicales sans référence à partir d'une petite quantité de données annotées. Notre modèle est basé sur l'auto-attention par convolution qui vise à modéliser une représentation complexe à partir des caractéristiques locales pertinentes des images. Nous appliquons également un apprentissage par adaptation au domaine de manière non supervisée et semi-supervisée. Le modèle proposé est évalué à travers un jeu de données composé de plusieurs images et de leurs scores subjectifs correspondants. Les résultats obtenus ont montré l'efficacité de la méthode proposée, mais aussi la pertinence de l'application de l'adaptation au domaine pour généraliser sur différents domaines multimédia en ce qui concerne la tâche de la prédiction de la qualité perceptuelle.

Abstract – Predicting the quality of multimedia content is often needed in different fields. In some applications, quality metrics are crucial with a high impact, and can affect decision making such as diagnosis from medical multimedia. In this paper, we focus on such applications by proposing an efficient and shallow model for predicting the quality of medical images without reference from a small amount of annotated data. Our model is based on convolution self-attention that aims to model complex representation from relevant local characteristics of images, which itself slide over the image to interpolate the global quality score. We also apply domain adaptation learning in unsupervised and semi-supervised manner. The proposed model is evaluated through a dataset composed of several images and their corresponding subjective scores. The obtained results showed the efficiency of the proposed method, but also, the relevance of the applying domain adaptation to generalize over different multimedia domains regarding the downstream task of perceptual quality prediction.

1 Introduction

La prédiction de la qualité des contenus multimédia est souvent nécessaire dans plusieurs domaines. Elle permet de quantifier dans quelle mesure les distorsions introduites sur un contenu multimédia peuvent endommager la qualité visuelle perçue. Pour certaines applications spécifiques, les mesures de qualité sont cruciales et ont un impact important. Les images médicales font partie des données les plus sensibles car leur qualité peut conduire à un diagnostic et à un pronostic erronés [1]. Il est donc important de développer des métriques efficaces dédiées à ces images particulières. Ici, nous nous concentrons sur les approches sans référence car elles correspondent davantage au cas réel. Des métriques intéressantes ont déjà été proposées dans la littérature. Dans [2], les auteurs ont proposé d'étendre la méthode NIQE [3] aux images médicales. Cette métrique, appelée NIQE-k, est basée sur une analyse du signal dans le domaine des fréquences. Dans [4], les auteurs ont développé une métrique basée sur une analyse de gradient, tandis que les caractéristiques de texture ont été utilisées dans [5].

Dans cet article, nous proposons un modèle efficace pour prédire la qualité des images médicales sans référence à partir d'une petite quantité de données annotées. Les architectures de réseaux de neurones de convolution profond (CNN) sont conçues pour bien fonctionner sur l'apprentissage de représentations hiérarchiques. Leurs premières couches détectent des motifs simples comme les bords et les gradients, tandis que les couches supérieures détectent des caractéristiques plus abstraites liées à la structure globale [6]. La robustesse de ces modèles peut conduire à ignorer les effets perceptuels introduits. Pour résoudre ce problème dans notre contexte, nous employons des modèles CNNs peu profonds qui incorporent un module d'auto-attention, mais aussi qui glissent eux-mêmes sur l'image afin de modéliser des caractéristiques locales complexes liées à la tâche de prédiction de la qualité. Néanmoins, le manque de bases de données publiques pour la qualité des images médicales empêche le développement de mesures de qualité profondes et personnalisées. Comme solution, nous appliquons un apprentissage par adaptation au domaine de manière non supervisée et semi-supervisée afin de

créer un lien entre les domaines de données.

Les principales contributions de notre article sont résumées ci-dessous :

— Nous proposons un nouveau modèle CNN peu profond et efficace pour prédire la qualité perceptuelle qui repose sur l'extraction d'informations utiles à partir de caractéristiques locales.
— Nous analysons l'impact de l'adaptation au domaine non supervisée et semi-supervisée pour réduire le décalage dans l'extraction de la distribution des caractéristiques pertinentes entre les images médicales et celles naturelles.

2 Méthode proposée

La principale contribution de notre approche est l'utilisation d'un module d'auto-attention basé sur une convolution peu profonde qui glisse elle-même sur l'image pour extraire des caractéristiques locales informatives et modéliser le score de qualité global. Les principaux composants de notre modèle sont décrits en détail ci-dessous.

2.1 Module d'auto-attention

L'intégration du mécanisme d'attention a récemment montré un saut important dans la performance de diverses tâches de vision par ordinateur [7] [8]. Contrairement aux mécanismes d'attention absolue, il apprend de manière totalement adaptative, conjointe et orientée vers la tâche permettant au réseau de hiérarchiser et d'associer des poids aux vecteurs de caractéristiques [9]. L'auto-attention calcule la réponse à une position dans un vecteur ou une séquence via toutes les positions dans la même séquence. De manière plus détaillée, elle établit la relation entre des caractéristiques distantes en incorporant l'auto-attention à notre réseau peu profond afin de capturer des caractéristiques complexes liées à notre tâche (c'est-à-dire l'évaluation de la qualité). Cela permet de renforcer la capacité de représentation du réseau complet.

Pour un vecteur donné, nous devons en extraire les vecteurs de **requête**, de **clé** et de **valeur** à l'aide de l'architecture CNN peu profonde. Cette dernière mesure l'attention en calculant une similarité entre la requête et les caractéristiques clés les plus pertinentes à l'aide d'une fonction de score. Les scores de sortie passent par l'étape de normalisation pour que la somme des valeurs de probabilité soit égale à un. Le vecteur de valeur ajusté final est une combinaison pondérée des vecteurs de valeur précédents basée sur le résultat du score normalisé [9]. Chaque patch est transformé en trois variables : le couple résultant $(Query, Key) \in R^{C/8\times N}$ de Q(X_i) et K(X_i), respectivement, simplifiant ainsi la dimension de $X_i \in R^{C\times N}$ où N =$(h = 8, w = 8)$ représentant le nombre d'emplacements de caractéristiques et C le nombre de canaux de sortie du module $V(.)$. La carte d'attention est obtenue après normalisation de la sortie du produit scalaire entre les vecteurs $Query$ et key en utilisant une fonction Softmax où S représente la similarité entre les espaces de caractéristiques $Query$ et Key :

$$S_{lj} = Query[l]^T.Key[j] \tag{1}$$

$$A_{j,l} = \frac{\exp(S_{lj})}{\sum_{j=1}^{N}\exp(S_{lj})}, \tag{2}$$

La carte d'attention A$\in R^{N\times N}$ représente la probabilité qu'une caractéristique de position particulière à l'emplacement l_{th} apparaisse à l'emplacement j_{th} dans N emplacements de caractéristiques $(_{j,l})\in R^N$. L'espace de caractéristiques $Value$ est encore amélioré en le multipliant par la carte d'attention. Un paramètre apprenant γ est également utilisé afin d'apprendre dans quelle mesure la prédiction globale du patch doit reposer sur le contexte à partir des caractéristiques locales.

$$Output = \gamma \times Value.A + Value \tag{3}$$

Enfin, un Max-pooling global est appliqué à la sortie afin d'obtenir un vecteur unidimensionnel qui est ensuite passé par un regresseur de type réseau de neurones multi-couche (MLP) peu profond pour interpoler le score de qualité des patchs, le score de qualité global étant obtenu en faisant la moyenne de la qualité de tous les patchs.

2.2 Adaptation du domaine

Pour exploiter pleinement la puissance de nos modèles, nous devons les adapter à un nouveau domaine source. En premier lieu, les modèles doivent être capables de percevoir les deux domaines (c'est-à-dire les scènes naturelles originales et les images médicales) comme faisant partie de la même distribution de données $\mathcal{D}$ [10]. Pour ce faire, il faut minimiser la distance perçue entre les deux distributions d'images $\mathcal{D}_n$ $\mathcal{D}_p$.

Lorsque nous entraînons nos modèles sur des images provenant des deux distributions, nous ajoutons une petite branche au réseau qui classe les images comme provenant de $\mathcal{D}_n$ ou de $\mathcal{D}_p$. En outre, nous ajoutons une couche d'inversion de gradient (GRL) au-dessus de cette branche qui inverse le signe du flux de gradient pendant la rétro-propagation. Eq. 4 définit la propagation vers l'avant, tandis que Eq. 5 concerne la propagation arrière.

$$GRL(x) = x \tag{4}$$

$$GRL(\frac{\partial L_d}{\partial x}) = -\frac{\partial L_d}{\partial x} \tag{5}$$

où x est l'entrée de la couche, et $\frac{\partial L_d}{\partial x}$ représente le gradient de la fonction de perte de domaine L_d lors de la rétro-propagation à travers le réseau.

L'inversion du gradient aide l'extracteur de caractéristiques du réseau à minimiser la distance entre les distributions de domaines. Cela oblige l'extracteur de caractéristiques à ne pas tenir compte des caractéristiques et des bruits spécifiques au domaine, et à mettre l'accent sur les caractéristiques mutuelles des deux domaines. Elle peut également être modélisée comme l'union des 2 distributions considérées moins la distribution du

bruit de chacun des domaines :

$$\mathcal{D} = \mathcal{D}_p + \mathcal{D}_n - (\mathcal{N}_p + \mathcal{N}_n) \quad (6)$$

où $\mathcal{D}_n, \mathcal{D}_p$ et $\mathcal{D}$ sont définis comme précédemment, et $\mathcal{N}_p$ et $\mathcal{N}_n$ sont les distributions spécifiques du bruit du domaine source.

2.3 Détails techniques

Nous avons implémenté nos modèles en utilisant PyTorch et les avons entraînés sur les deux domaines de données avec et sans adaptation au domaine, chaque fois pendant 100 époques. Les modules $K(.)$, $Q(.)$ et $V(.)$ ont été initialisés de manière aléatoire ainsi que le regresseur et le classificateur de domaine. Nous normalisons les scores de qualité de la vérité terrain (c'est-à-dire MOS : Mean Opinion Score) des deux domaines de données pour obtenir des distributions de probabilité. Nous avons utilisé l'entropie croisée binaire (BCE) afin de minimiser le risque global et l'optimiseur Adam pour entraîner le modèle. Nous avons fixé le taux d'apprentissage à $5 * 10^{-4}$ et le paramètre γ a été initialisé à zéro afin de se concentrer sur l'apprentissage de la tâche principale.

3 Résultats expérimentaux

3.1 Bases de données

Comme mentionné ci-dessus, l'adaptation au domaine a été appliqué dans cette étude afin d'optimiser pleinement l'utilisation de nos modèles. A cette fin, deux jeux de données ont été utilisés : l'un composé d'images naturelles considérées ici comme les données sources et le second composé d'images médicales considérées comme les données cibles. Les deux jeux de données sont décrits ci-dessous :

— **Sous-ensemble de la base TID13 :** Un sous-ensemble de la base de données TID13 a été utilisé dans cette étude. Plus précisément, nous avons considéré uniquement les 125 images dé-bruitées (c'est-à-dire la distorsion numéro 9 du jeu de données) dérivées de 25 images de référence.

— **MD72:** La base de données médicales proposée dans [2] a été ici utilisée. Cette dernière, désigné ici par MD72, est composée de 72 images d'échographie du foie avec les MOS correspondants. Plus précisément, 60 images dé-bruitées ont été obtenues à partir de 12 images de référence grâce à 5 algorithmes de dé-bruitage.

Il convient de noter que seul un sous-ensemble spécifique de l'ensemble de données TID13 a été utilisé comme données source pour l'apprentissage de l'adaptation au domaine. Ce choix a été motivé par le fait que le jeu de données d'images médicales considéré n'est composé que d'images dé-bruitées de leur contenu et que ce sous-ensemble est donc le plus lié à notre tâche. Dans cette section, nous analysons plus en détail l'impact des schémas d'adaptation au domaine (AD) proposés sur les performances. L'objectif est de montrer la pertinence de l'utilisation de l'AD dans un tel contexte. À cette fin, nous avons calculé les corrélations pour trois configurations :

1. Approche entièrement supervisée **sans DA** en entraînant le modèle sur la base MD72 et en le testant sur les deux ensembles de données.
2. Approche entièrement supervisée **sans DA** en entraînant le modèle sur la base TID13 et en le testant sur les deux ensembles de données.
3. **AD non supervisée** en utilisant la basse TID13 comme donnée source et MD72 comme donnée cible.
4. **AD semi-supervisée** en utilisant TID13 comme donnée source et MD72 comme donnée cible.

Config.	MD72		TID13	
	PLCC ↑	SROCC ↑	PLCC ↑	SROCC ↑
1	0.557	0.670	0.324	0.414
2	0.685	0.560	0.906	**0.794**
3	0.756	0.769	0.683	0.685
4	**0.810**	**0.812**	**0.907**	0.784

Table 1 – Impact de l'AD sur les performances.

A partir des résultats présentés dans le Tableau 1, plusieurs observations peuvent être faites. D'après les résultats obtenus pour la configuration 1, nous pouvons voir que l'entraînement de notre modèle directement sur le jeu de données médicales n'était pas suffisant pour bien prédire la qualité même avec un modèle peu profond. Les corrélations obtenues étaient donc très faibles pour les deux jeux de données. Les résultats de la configuration 2 montrent qu'il semble plus facile d'atteindre un certain niveau de performance pour les tâches liées aux images naturelles (c'est-à-dire le sous-ensemble du jeu de données TID13) que pour les images médicales. A travers les résultats des configurations 3 et 4, nous pouvons clairement constater l'impact de l'AD sur les performances. En effet, les performances obtenues sur le jeu de données cible (i.e. MD72) ont augmentées de manière significatives. Cependant, nous avons également remarqué que les corrélations chutent sur les données sources (i.e. le sous-ensemble de TID13) lorsque l'apprentissage non supervisé de l'AD est appliqué. Les meilleures corrélations globales ont été obtenues par l'apprentissage semi-supervisé de l'AD pour les deux ensembles de données avec des améliorations considérables.

3.2 Comparaison avec les méthodes de l'état de l'art

Notre méthode est évaluée en termes de corrélations avec les scores subjectifs uniquement sur les données cibles (c'est-à-dire les images médicales). Nous avons calculé les corrélations de Pearson et de Spearman entre les scores prédits et ceux subjectifs. Nous avons comparé les performances de notre modèle à certaines mesures de qualité sans référence de l'état de l'art : BRISQUE [11], NIQE [3], bliinds [12] et BIQAA [13]. Nous avons également considéré une métrique, appelée NIQEK [2], qui a été développée spécifiquement pour ce type d'images médicales. En plus de ces méthodes, nous avons enfin considéré 2 métriques dédiées au flou (i.e. MARZILIANO [14] et RadialIndex [15]) et 1 autre pour l'estimation du bruit (i.e. [16]).

Métrique	PLCC ↑	SROCC ↑
BRISQUE	0.6551	0.3829
NIQE	0.4972	0.2768
NIQEK	0.5682	0.4429
bliinds	0.4373	0.4131
BIQAA	**0.6643**	0.5297
Noise	0.5978	0.5286
MARZILIANO	0.5303	0.3028
RadialIndex	0.5905	**0.5338**
Our	**0.810**	**0.812**

Table 2 – Résultats obtenus sur le jeu de données MD72

Le tableau 2 montre les performances obtenues pour chacune des méthodes comparées. Les 2 meilleurs résultats sont mis en évidence en gras. Comme on peut le voir, toutes les métriques comparées ont obtenu des corrélations inférieures à 0.7 sauf notre méthode qui les a surpassées en atteignant des corrélations supérieures à 0,8. Les seconds meilleurs PLCC et SROCC ont été obtenus respectivement par BIQAA et RadialIndex, loin des résultats obtenus par notre méthode.

4 Conclusion

Dans cette article, nous avons proposé une nouvelle méthode d'évaluation de la qualité des images médicales. En particulier, nous avons cherché à aborder la tâche de prédiction de la qualité des images échographiques pour appliquer un diagnostic précis à partir d'une petite quantité de données annotées. À cette fin, nous avons utilisé un CNN peu profond comme module d'auto-attention afin de modéliser efficacement une représentation complexe à partir de caractéristiques locales qui affectent la perception visuelle. Nous avons également utilisé l'apprentissage non-supervisé et semi-supervisé de l'adaptation au domaine pour réduire le décalage entre les distributions des domaines de données et bien généraliser sur de petits ensembles de données médicales annotées. Les résultats obtenus montrent l'efficacité de notre méthode.
Nous espérons créer un lien plus fort entre les domaines de données en incitant notre modèle peu profond à apprendre d'abord les structures communes pertinentes de qualité de différents domaines à partir d'images uniquement de manière auto-supervisée.

References

[1] Lévêque Lucie a et, al., "On the subjective assessment of the perceived quality of medical images and videos," *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2018.

[2] M Outtas et, al., "Subjective and objective evaluations of feature selected multi output filter for speckle reduction on ultrasound images," *Physics in Medicine and Biology*, vol. 63, no. 18, pp. 185014, Sept. 2018.

[3] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[4] Köhler Thomas et, al., "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proceedings of the 26th IEEE international symposium on computer-based medical systems*. IEEE, 2013, pp. 95–100.

[5] Beatriz Remeseiro, Ana Maria Mendonça, and Aurélio Campilho, "Objective quality assessment of retinal images based on texture features," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 4520–4527.

[6] Jason Yosinski, Jeff Clune, Anh M Nguyen, Thomas J. Fuchs, and Hod Lipson, "Understanding neural networks through deep visualization," *ArXiv*, vol. abs/1506.06579, 2015.

[7] Yasser A. Dahou Djilali et, al., "Atsal: An attention based architecture for saliency prediction in 360 videos," *ArXiv*, vol. abs/2011.10600, 2020.

[8] Mohamed A. Kerkouri et, al., "Salypath: A deep-based architecture for visual attention prediction," in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 1464–1468.

[9] Tliba Marouane et, al., "Satsal: A multi-level self-attention based architecture for visual saliency prediction," *IEEE Access*, pp. 1–1, 2022.

[10] Yaroslav Ganin et, al., "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[11] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[12] Michele A. Saad, Alan C. Bovik, and Christophe Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.

[13] Salvador Gabarda and Gabriel Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A*, vol. 24, no. 12, pp. B42–B51, Dec 2007.

[14] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings. International Conference on Image Processing*, 2002, vol. 3, pp. III–III.

[15] Aladine. Chetouani et, al., "A new reference-free image quality index for blur estimation in the frequency domain," in *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2009, pp. 155–159.

[16] Dimitri Van De Ville and Michel Kocher, "Sure-based non-local means," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 973–976, 2009.