Évaluation sans référence de la qualité de maillages 3D par combinaison de prédictions sur projections 2D

Zaineb Ibork, Anass Nouri, Olivier Lézoray, Christophe Charrier, Raja Touahni

Évaluation sans référence de la qualité de maillages 3D par combinaison de prédictions sur projections 2D

Zaineb Ibork[1,2] Anass Nouri[1,2] Olivier Lézoray[2] Christophe Charrier[2] Raja Touahni[1]

[1]SETIME Laboratory, Information Processing and A.I Team
Faculty of Sciences, Ibn Tofail University, Kénitra, Morocco

[2] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

{zaineb.ibork,anass.nouri, touahni.raja}@uit.ac.ma {olivier.lezoray, christophe.charrier}@unicaen.fr

Résumé

Les maillages 3D ont suscité un grand intérêt dans la communauté de la vision par ordinateur en raison de leur utilisation dans plusieurs applications telles que la réalité virtuelle, les jeux, la préservation du patrimoine, etc. Cependant, ces contenus 3D peuvent être altérés lors d'étapes de pré-traitement telles que la compression ou le débruitage. Dans ce contexte, les algorithmes d'évaluation de la qualité peuvent être utilisés pour quantifier la quantité de distorsions qui affectent un maillage 3D et dégradent son rendu visuel. Nous introduisons une mesure d'évaluation de la qualité des maillages sans référence basée sur des caractéristiques convolutionnelles profondes, appelée DCFQI (Deep Convolutional Features Quality Index). En tirant parti de l'apprentissage par transfert, l'approche proposée permet d'évaluer la qualité visuelle sans avoir besoin d'un contenu de référence, imitant ainsi la vision humaine. Partant d'un rendu d'un maillage 3D en vues et patchs 2D, un réseau neuronal convolutionnel pré-entraîné est utilisé pour extraire automatiquement des caractéristiques profondes. Celles-ci sont utilisées dans un perceptron multicouche (MLP) afin de prédire le score de qualité objectif. Deux stratégies d'apprentissage sont présentées et comparées pour l'estimation de la qualité sans référence. Les résultats obtenus en termes de corrélation avec les scores humains subjectifs de qualité démontrent la supériorité de l'indice proposé par rapport aux méthodes existantes

Mots Clés

Maillage 3D, Evaluation de la qualité visuelle, Réseaux de neurones à convolution, Apprentissage profond, Apprentissage par transfert.

Abstract

3D meshes have gained significant interest in computer vision community due to their use in several applications such as virtual reality, gaming, heritage preservation, etc. However these 3D contents might be altered in the pre-processing steps like compression or denoising. In this context, quality assessment algorithms can be used to quantify the amount of distorsions that affect a 3D mesh and hence degrade its visual rendering. We introduce a no-reference mesh quality assessment index based on deep convolutional features named DCFQI (Deep Convolutional Features Quality Index). Leveraging the power of deep learning, particularly transfer learning, allows the proposed approach to score visual quality without the need of reference content, hence emulating the human vision. By rendering a 3D mesh into 2D views and patches, a pre-trained convolutional neural network is used to automatically extract deep features from the latter. The obtained features are used in a Multi Layer Perceptron (MLP) to predict the objective quality score. Two learning strategies are presented and compared for blind quality estimation. Obtained results in terms of correlation with subjective human scores of quality demonstrate the superiority of the proposed index over existing methods.

Keywords

3D mesh, Visual Quality Assessment, Convolutional Neural Network, Deep learning, Transfer Learning.

1 Introduction

L'évaluation de la qualité perceptuelle des maillages 3D, (Mesh Visual Quality Assessement- (MVQA), a suscité un grand intérêt ces dernières années en raison de l'utilisation des modèles 3D dans diverses applications, allant de la réalité virtuelle à la préservation du patrimoine culturel. Comme les maillages 3D subissent généralement différentes opérations (avec perte) de traitement de la géométrie, des distorsions peuvent se produire. Ceci a un impact sur leur qualité visuelle et éventuellement sur les performances des applications associées. Bien que l'évaluation subjective de la qualité par des observateurs humains soit une méthode fiable, elle est coûteuse, laborieuse et prend du temps [1]. Les méthodes objectives d'évaluation de la qualité offrent une solution viable à ces défis. Elles

peuvent être classées en trois catégories en fonction de la disponibilité d'un objet de référence : les méthodes à référence complète (FR), les méthodes sans référence (NR) ne disposant d'aucune information de référence, et les méthodes à référence réduite (RR) disposant d'informations de référence partielles, telles que des caractéristiques extraites. Les méthodes existantes axées sur la perception se concentrent principalement sur les méthodes avec référence [2, 3, 4, 5] et avec référence réduite [6, 7] afin d'évaluer la qualité perçue. Toutefois, dans la pratique, une référence n'est pas toujours disponible, ce qui nécessite la mise au point de méthodes sans référence. Récemment, les réseaux neuronaux convolutionnels (CNN) ont été largement adoptés pour définir des indices de qualité sans référence des maillages [8, 9, 10]. Afin d'imiter l'évaluation subjective effectuée par le système visuel humain, nous proposons la solution suivante : nous générons des projections 2D de chaque maillage à partir de plusieurs points de vue et divisons celles-ci en patchs avec chevauchements. Ensuite, alors que de nombreuses approches de MVQA sont basées uniquement sur des patchs, nous considérons les informations issues à la fois des vues ou de patchs issus des vues. Nous considérons des architectures pré-entrainées pour extraire des caractéristiques (ce qui correspond à un apprentissage par transfert) qui permettent d'apprendre à prédire la qualité de vues ou de patchs. Le score final de qualité d'un maillage est ensuite obtenu en moyennant les prédictions issues des vues ou des patchs.

L'article est organisé comme suit. La section 2 décrit la préparation des données et la méthode proposée. La section 3 décrit les résultats expérimentaux, y compris les détails de la base de données utilisée, le protocole de validation et une discussion comparative des résultats. Enfin, nous concluons par des remarques et des perspectives sur le sujet.

2 Mesure de qualité d'un maillage 3D à partir de projections 2D

2.1 Principe de l'approche

Étant donné un maillage 3D dont la qualité visuelle doit être évaluée, l'approche proposée génère plusieurs projection 2D (nommées vues) en faisant varier le point de vue autour du maillage. Les vues 2D obtenues sont normalisées afin de minimiser la quantité de fond blanc dans l'image. Les vues pré-traitées sont ensuite divisées en quatre patchs qui se chevauchent. Chaque vue 2D (ou chaque patch extrait des vues 2D) est fournie à un réseau convolutionnel pré-entraîné (VGG 16) [11] afin d'obtenir un vecteur de caractéristiques. Ce dernier est fourni à un MLP entrainé afin d'estimer la qualité d'une vue ou d'un patch, à partir des valeurs de référence (MOS - Mean Opinion Score). Comme chaque maillage 3D M_i est décrit par un ensemble de mesures de qualité associées à des vues 2D ou des patchs, nous faisons la moyenne des notes de qualité objectives obtenues pour estimer la qualité prédite du maillage 3D (PMOS - Predicted Mean Opinion Score). Dans la suite, nous décrivons le processus de préparation de nos bases de données (vues et patchs).

2.2 Vues d'un maillage 3D

Comme nous l'avons déjà mentionné, notre approche considère des vues et patchs 2D d'un maillage 3D afin d'en évaluer la qualité. Étant donné une base de données de N maillages 3D, l'objectif est d'effectuer des projections 2D de chaque maillage 3D $M_i, i \in [1, N]$ afin d'obtenir des vues/patchs 2D selon différents angles de vue. Pour s'assurer que chaque maillage est positionné de manière similaire dans les projections rendues (que nous nommons vues), le centre du maillage est placé à l'origine du système de coordonnées. Cela permet d'obtenir des rendus comparables pour tous les maillages, ce qui est indispensable pour prédire des scores de qualité fiables et reproductibles à partir des caractéristiques extraites représentant les vues/patchs.

Vues 2D Les vues 2D du maillage 3D M_i sont obtenues à partir de 11 points de vue en modifiant les angles d'azimut (θ_a) et d'élévation (θ_e) de $\frac{\pi}{3}$ (60 degrés) pour chaque point de vue. La figure 1 illustre ce processus.

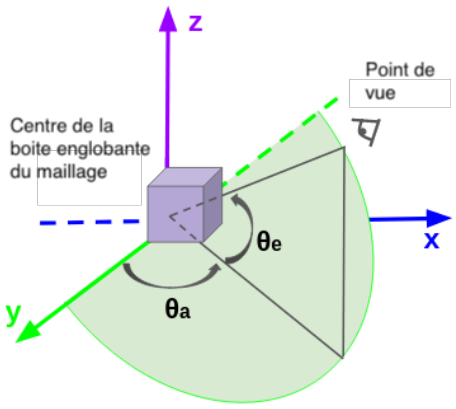

FIGURE 1 – Illustration de la position de la caméra dans le processus de rendu : angle d'azimut (θ_a) dans le plan horizontal avec $z = 0$ et angle d'élévation (θ_e) à partir du plan xz avec $y = 0$.

L'angle d'élévation est fixé à 0 degrés tout en faisant varier l'azimut (et vice versa) pour capturer les vues, assurant ainsi des transitions en douceur. La position de la caméra et la distance à l'objet sont fixées manuellement pour s'assurer que l'objet apparaît proche de la caméra, maximisant ainsi la présence des détails les plus fins. Un seul spot d'éclairage est utilisé et est attaché à la caméra (c-à-d un éclairage frontal). Ce processus nous permet de créer un ensemble complet de vues 2D, présentant différentes perspectives avec les détails importants de l'objet. La figure 2 présente un exemple de vues 2D du maillage 3D Armadillo issu de la base de données Liris/Epfl General Purpose [1].

Pré-traitement et décomposition des vues 2D Les vues 2D qui en résultent ont une définition de 1024×1024

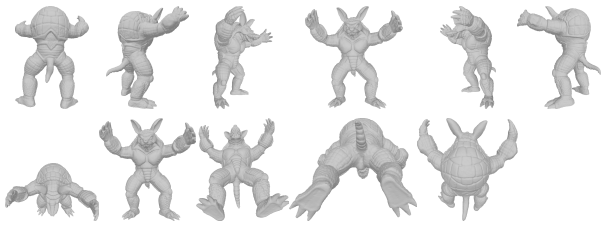

FIGURE 2 – Les 11 vues 2D du maillage Armadillo : les vues de la première ligne sont obtenues en fixant $\theta_e = 0$ et en variant θ_a de 60 degrés. Ce processus est inversé dans la deuxième ligne.

pixels. Cette taille a été fixée pour capturer les détails discriminants qui sont d'une importance capitale pour l'évaluation de la qualité visuelle. Cependant, ces images contiennent également une quantité importante de fond blanc. Pour minimiser l'impact du fond blanc, inutile pour l'évaluation de la qualité, nous recadrons et redimensionnons les images pour n'inclure que la boîte englobante du maillage, en supprimant effectivement la majeure partie du fond blanc environnant. Comme cette opération de recadrage dépend de la taille de la boîte englobante du maillage, les tailles des images résultantes peuvent être différentes. Pour éviter cela, nous redimensionnons toutes les images à 512×512.

Si les vues 2D sont intéressantes pour capturer les détails des maillages 3D, seules $11 \times N$ vues sont obtenues. Un tel nombre de vues pourrait ne pas être suffisant pour exploiter la puissance des architectures d'apprentissage profond. Pour y remédier, quatre patchs sont extraits de chaque vue 2D. Contrairement à l'approche de [10] qui extrait de très petits patchs de taille 32×32, nous considérons des patchs plus grands de taille 288×288. En effet, le fait d'avoir des patchs de très petite taille présente de nombreux inconvénients. Tout d'abord, les petits patchs ne contiennent pas toujours suffisamment d'informations pour l'évaluation de la qualité. Deuxièmement, ces patchs peuvent n'être constitués que d'arrière-plan et des stratégies spécifiques sont alors nécessaires pour les éliminer [12]. Troisièmement, cette méthode ne garantit pas que le nombre de patchs extraits par vue soit toujours le même, ce qui crée un ensemble de données déséquilibré. Pour assurer une meilleure couverture de toutes les informations, en particulier au niveau de la connexion entre des patchs adjacents, les patchs sont extraits avec un chevauchement de 20%. La figure 3 illustre la décomposition d'une vue 2D en quatre patchs superposés. En résumé, étant donné une base de données de N maillages, nous construisons deux bases de données B_k avec $k \in \{\text{view}, \text{patch}\}$. B_{view} contient $N \times 11$ images (les vues), tandis que B_{patch} contient $N \times 11 \times 4$ images (les patchs). Quelle que soit la base de données, chaque image est normalisée entre 0 et 1. Contrairement à [12], où les auteurs ont effectué une normalisation locale du contraste, nous avons préféré une normalisation globale effectuée sur le canal de clarté L^* dans l'espace colorimétrique CIELAB. En effet, la luminosité perçue (L^*) est non linéaire comme la perception visuelle humaine et sa normalisation peut améliorer le processus d'évaluation de la qualité.

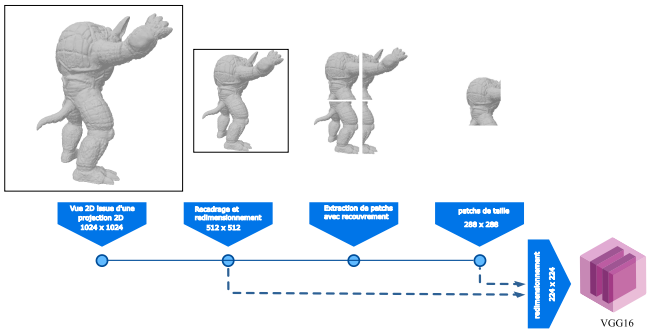

FIGURE 3 – Pré-traitement pour passer des vues aux patchs.

2.3 Apprentissage et régression

Une fois la base de données B_k d'images construite, notre objectif est d'évaluer la qualité de chaque maillage M_i à partir de ses images I_j^i (vues 2D ou patchs issus de vues 2D). Pour ce faire, des caractéristiques sont extraites des images rendues à l'aide d'un réseau convolutionnel VGG16 pré-entraîné (après avoir redimensionné les images à 224×224). Ce réseau a été considéré comme l'extracteur de caractéristiques le plus efficace par rapport à d'autres modèles tels que AlexNet et ResNet [12]. L'organigramme de l'approche proposée est présenté dans la figure 4. Chaque image est introduite dans le modèle VGG16 pré-entraîné qui agit comme un extracteur de caractéristiques ϕ en sauvegardant sa sortie avant ses couches denses. En conséquence, un vecteur de caractéristiques de taille $7 \times 7 \times 512$ est obtenu puis aplati afin d'obtenir un vecteur de taille 25088.

Ce vecteur de caractéristiques $\phi(I_j^i)$ est utilisé comme entrée dans un MLP peu profond pour effectuer la tâche de régression pour l'évaluation de la qualité (nommé MLPR). L'objectif est d'évaluer la qualité $\text{PMOS}_k(I_j^i)$ d'une image I_j^i provenant d'une base de données B_k sur la base du vecteur de caractéristiques correspondant $\phi(I_j^i)$. Pendant l'apprentissage, le MOS de référence pour chaque image I_j^i correspond à celui du maillage 3D M_i. Il est évident que $\text{PMOS}_k(I_j^i)$ doit être proche de $\text{MOS}(M_i)$. Le MLP peu profond contient une seule couche cachée de 512 neurones avec une fonction d'activation ReLu, et une couche de sortie à un seul neurone. Après la couche dense, une couche DropOut avec un taux de $0,5$ est ajoutée pour éviter un sur-apprentissage. Afin d'initialiser les poids de la couche dense, nous avons utilisé la méthode d'initialisation uniforme de Glorot [13].

Pour prédire le score de qualité d'une image d'entrée, le MLPR est entraîné sur B_{view} ou B_{patch} (la section suivante présentera le protocole d'entraînement et d'évalua-

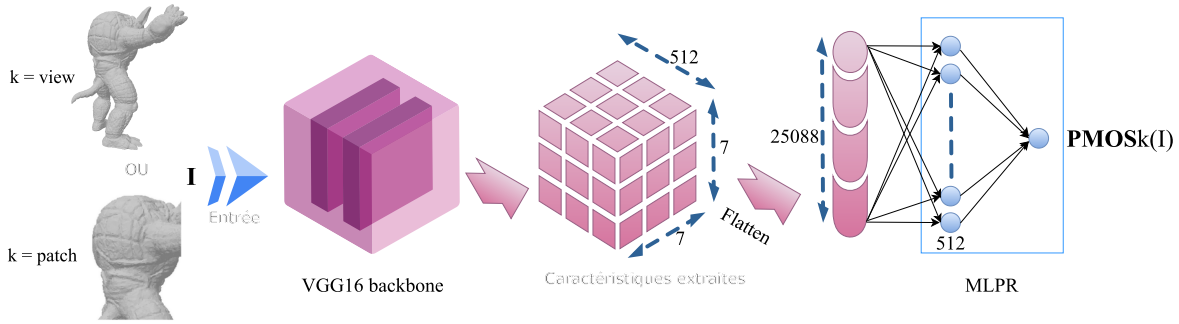

FIGURE 4 – Le pipeline de l'indice d'évaluation de la qualité proposé qui estime la qualité d'une image rendue (une vue 2D ou un patch issu d'une vue 2D).

tion). Une fois l'entraînement effectué, l'estimation de la qualité d'un maillage M_i est calculée comme suit :

$$\text{PMOS}(M_i) = \frac{1}{n_k} \sum_{j=1}^{n_k} \text{PMOS}_k(I_j^i) \qquad (1)$$

avec $k \in \{\text{view}, \text{patch}\}$, et I_j^i est le j^{me} patch ou vue parmi les n_k images du maillage M_i ($n_{\text{view}} = 11$ et $n_{\text{patch}} = 44$). La qualité d'un maillage est donc la moyenne des prédictions $\text{PMOS}_k(I_j^i)$ obtenues pour toutes ses images. Nous investiguerons également une agrégation des scores effectuée par un MLP à une couche cachée contenant n_k poids :

$$\text{PMOS}(M_i) = MLP\left(\text{PMOS}_k\right) \qquad (2)$$

où PMOS_k est le vecteur de toutes les prédictions. Le MLP a une fonction d'activation RELu, est initialisé avec Glorot, et un early stopping est utilisé avec un optimiseur RMSProp.

3 Résultats expérimentaux

3.1 Base de données

Dans ce travail, nous avons considéré la base de données LIRIS/EPFL General-Purpose database [1], qui est un ensemble de données largement utilisé dans le domaine de l'évaluation de la qualité des maillages 3D. Cette base de données contient 88 modèles de maillages divisés en 4 maillages de référence (nommés Dinosaur, Armadillo, RockerArm et Venus) et 84 versions dégradées. Avec cette base de données, nous disposons donc de $N = 88$ maillages. Il y a 21 distorsions pour chaque modèle de référence. Ces distorsions sont générées par l'application de techniques de lissage et d'ajout de bruit sur des zones lisses, des zones rugueuses ou des zones intermédiaires (entre les régions rugueuses et lisses). Un score d'opinion moyen (MOS) a été obtenu pour chaque modèle en faisant la moyenne des scores subjectifs obtenus par 12 observateurs humains. La note est comprise entre 0 (pour une bonne qualité) et 10 (pour une mauvaise qualité). La figure 5 présente le maillage de référence Dinosaur (à gauche) et une version dégradée après ajout de bruit sur les zones rugueuses et intermédiaires.

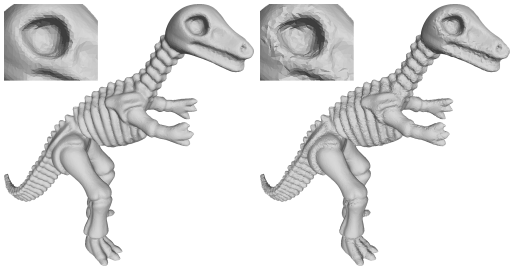

FIGURE 5 – Maillage 3D de référence Dinosaur (à gauche) et sa version déformée (à droite) avec ajout de bruit sur les zones rugueuses et intermédiaires. La sous-figure en haut à gauche présente une partie zoomée.

3.2 Protocole d'évaluation

Comme présenté dans la section précédente, nous pouvons apprendre à prédire la qualité d'un maillage M_i à partir de ses images rendues I_j^i qui proviennent soit de B_{view} soit de B_{patch}. Cela signifie que nous pouvons prédire la qualité d'un maillage soit à partir de 11 vues 2D, soit à partir de 44 patchs issus de vues 2D. Nous comparerons ces deux méthodologies avec l'état de l'art. Les sous-sections suivantes décrivent le protocole d'évaluation que nous avons utilisé.

Pour évaluer la performance de l'évaluation de la qualité, nous considérons deux mesures standard : le coefficient de corrélation de rang de Spearman $SROOC$ et le coefficient de corrélation linéaire de Pearson $PLCC$. Ces mesures sont couramment utilisées dans le domaine de l'évaluation de la qualité visuelle pour mesurer la similarité entre les scores prédits et les valeurs de vérité terrain. Elles serviront de base pour comparer les performances de la méthode proposée avec les mesures existantes d'évaluation de la qualité des maillages sans référence. Le coefficient $SROOC$ r_s est une mesure statistique utilisée pour évaluer la force et la direction de la relation monotone entre les scores de qualité prédits $\text{PMOS}(M_i)$ et les scores d'opinion moyens de référence $\text{MOS}(M_i)$. La métrique $PLCC$ r_p évalue la relation linéaire ou la corrélation entre les scores prédits et les valeurs de référence. Les deux mesures sont comprises entre -1 et 1. Une valeur de 1 indique une corrélation po-

sitive parfaite, -1 indique une corrélation négative parfaite et 0 indique une absence de corrélation. En utilisant ces mesures, nous serons en mesure de comparer les performances des méthodologies basées sur les patchs et sur les vues, ainsi que de les comparer aux approches de l'état de l'art les plus récentes.

3.3 Modèle de base

Nous avons commencé nos expériences avec le modèle de régression MLP (MLPR) présenté dans la section précédente et illustré dans la figure 4. Le modèle a été entrainé pour un nombre fixe de 20 époques en utilisant l'optimiseur RMSprop avec un taux d'apprentissage fixé à 0.001. La fonction de perte utilisée est l'erreur absolue moyenne (MAE). Sur la base de tests approfondis, nous avons constaté que les meilleurs scores de corrélation sont obtenus avec une taille de batch d'un tiers de la taille de l'ensemble d'apprentissage. Ce paramètre sera fixé de cette manière pour toutes nos expériences. Afin d'évaluer la précision du modèle, nous appliquons la procédure de validation croisée Leave-One-Mesh-Out (LOMO-CV). Lors de l'entraînement, tous les maillages sont pris en compte à l'exception d'un maillage et de ses versions dégradées.

La figure 6 présente le processus d'apprentissage et de test pour chaque MLPR entrainé en LOMO-CV. Dans cette approche, les images associées à un maillage spécifique sont toutes exclues du processus d'apprentissage. Le réseau neuronal entraîné est ensuite testé sur ces images exclues afin d'évaluer ses performances car elles représentent des données non vues. Ce processus LOMO-CV est répété pour chacun des 22 maillages de l'ensemble de données (un maillage de référence et ses versions dégradées). En recourant à la validation croisée, nous garantissons une évaluation objective du modèle MLPR, puisqu'il est évalué sur des données strictement indépendantes sur lesquelles il n'a pas été entrainé. Pour un fold de la LOMO-CV, nous obtenons des prédictions de qualité $\text{PMOS}_k(I_j^i)$ pour toutes les images de ce fold. Par conséquent, nous pouvons évaluer les résultats directement au niveau de l'image (vue 2D ou patch issu d'une vue 2D) ou au niveau du maillage. Nous différencierons ces deux modes d'évaluation par les noms **concat** et **average**. En effet, comme le montre l'équation 2, pour obtenir la qualité d'un maillage donné M_i, nous devons faire la moyenne de toutes les prédictions de ses vues ou patchs rendus. Pour ce faire, nous regroupons les scores de prédiction par maillage et calculons le score moyen pour chacun. Ce processus d'agrégation nous permet de consolider les informations provenant de plusieurs prédictions en un score unique qui reflète la qualité globale du maillage. En passant des scores de concaténation aux scores moyens par maillage, nous avons obtenu une évaluation plus ciblée et plus facile à interpréter de la qualité de chaque maillage de notre ensemble de données. Le tableau 1 présente les résultats du modèle de base (avec 20 époques) en termes de valeurs SROOC.

Nous pouvons remarquer que les valeurs de corrélation de

	Sur $\mathbf{B}_{\text{view}}$		Sur $\mathbf{B}_{\text{patch}}$	
Configuration	$\mathbf{r_s}$ **concat**	$\mathbf{r_s}$ **average**	$\mathbf{r_s}$ **concat**	$\mathbf{r_s}$ **average**
Armadillo out	**0.704**	0.832	0.657	**0.949**
Dinosaur out	0.01	0.292	**0.253**	**0.779**
Venus out	0.895	**0.953**	**0.896**	0.942
RockerArm out	0.583	0.929	**0.727**	**0.949**
Moyenne	0.548	0.751	**0.633**	**0.904**

TABLE 1 – Valeurs $SROOC$ pour le modèle de base entraîné pendant 20 époques sur les bases de données B_{view} ou B_{patch}.

Spearman pour $\mathbf{r_s}$ concat sont relativement faibles pour les deux bases de données $\mathbf{B}_{\text{view}}$ et $\mathbf{B}_{\text{patch}}$, en particulier pour le maillage Dinosaur et ses versions dégradées, où la prédiction peut être grandement améliorée en faisant la moyenne de la prédiction de tous les patchs. Il est logique que le protocole exploitant des patchs donne de meilleurs résultats puisque l'ensemble de données est plus important. La seule exception est le modèle Venus, qui présente une corrélation légèrement meilleure avec le protocole basé sur les vues.

Afin d'améliorer les résultats, nous utilisons dans notre mise en œuvre la technique d'early stopping, qui est une forme de régularisation permettant d'éviter le surapprentissage et d'améliorer la généralisation du modèle entrainé. Cela nous permet également d'identifier le nombre optimal d'époques qui peut conduire à de meilleurs résultats. Pour chaque méthodologie (basée sur des vues ou des patchs), le nombre optimal d'époques est fixé après un délai de patience qui détermine le nombre d'époques à attendre avant d'arrêter le processus d'apprentissage si la métrique surveillée ne s'améliore pas. Les résultats sont présentés dans le tableau 2. Cela permet d'améliorer considérablement les résultats. Si nous comparons nos résultats avec ceux de [12] qui considère également VGG16 pour l'extraction des caractéristiques mais avec des patchs de très petite taille (32×32) et un nombre fixe d'époques de 40, ils ont obtenu un $\mathbf{r_s}$ moyen de 0.925 alors que notre approche atteint 0.945. Cela montre que le processus de minimisation peut être sujet à des minima locaux et qu'un arrêt précoce peut aider à y faire face. Des patchs plus grands permettent également d'améliorer les résultats car ils capturent plus de détails.

	Sur $\mathbf{B}_{\text{view}}$			Sur $\mathbf{B}_{\text{patch}}$		
	Early stopping (patience=30)			Early stopping (patience=60)		
Configuration	**époques**	$\mathbf{r_s}$ **concat**	$\mathbf{r_s}$ **average**	**époques**	$\mathbf{r_s}$ **concat**	$\mathbf{r_s}$ **average**
Armadillo out	77	0.815	0.963	307	**0.939**	**0.989**
Dyno out	211	0.142	0.789	680	**0.189**	**0.814**
Venus out	71	0.933	0.984	311	**0.971**	**0.995**
RockerArm out	110	**0.957**	0.962	400	0.836	**0.981**
Moyenne	–	0.712	0.924	–	**0.734**	**0.945**

TABLE 2 – Valeurs $SROOC$ pour le modèle de base entrainé avec un arrêt précoce sur les bases de données B_{view} et B_{patch}.

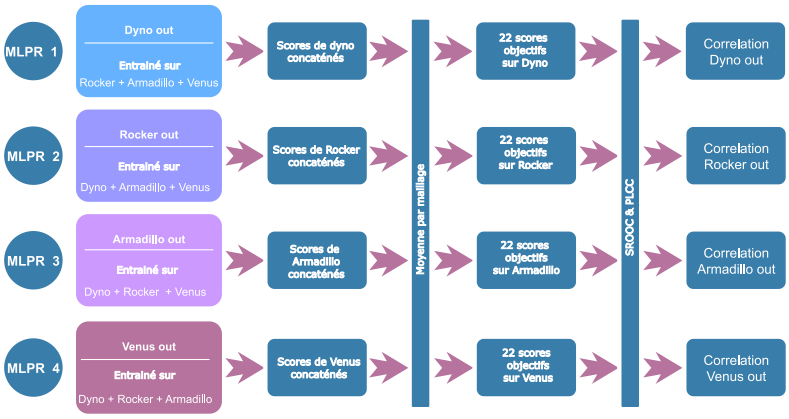

FIGURE 6 – Réseaux neuronaux résultant de la validation croisée leave-one-mesh-out (LOMO-CV).

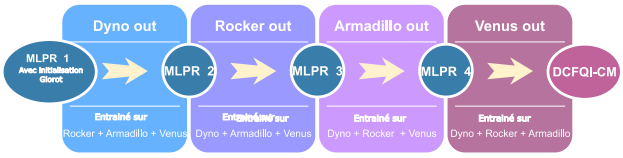

FIGURE 7 – Entrainement du modèle cumulatif.

	Sur $\mathbf{B}_{\text{view}}$				Sur $\mathbf{B}_{\text{patch}}$			
	CM		RCM		CM		RCM	
Configuration	$\mathbf{r_S}$	$\mathbf{r_P}$	$\mathbf{r_S}$	$\mathbf{r_P}$	$\mathbf{r_S}$	$\mathbf{r_P}$	$\mathbf{r_S}$	$\mathbf{r_P}$
Armadillo out	0.993	0.999	0.992	0.998	0.998	0.999	**0.997**	**0.999**
Dyno out	0.999	0.998	0.995	**0.998**	0.998	0.999	**0.998**	**0.998**
Venus out	0.999	0.999	0.997	0.998	1	1	**0.999**	**0.999**
RockerArm out	0.986	0.995	0.986	0.992	0.994	0.997	**0.991**	**0.996**
Moyenne	0.994	0.997	0.992	0.996	0.997	0.998	**0.996**	**0.998**

TABLE 3 – $SROOC$ et $PLCC$ pour les modèles cumulatif (CM) et cumulatif ré-entraîné (RCM) entraînés sur les bases de données B_{view} ou B_{patch}.

3.4 Modèle cumulatif

Jusqu'à présent, nous avons montré dans la section précédente que notre approche est compétitive par rapport à des approches similaires de l'état de l'art (en particulier avec [12], mais d'autres comparaisons seront présentées dans la section suivante). Comme nous avons utilisé un apprentissage en LOMO-CV, nous sommes en mesure d'estimer la qualité de l'approche que nous proposons. Cependant, nous ne sommes pas en mesure d'utiliser un réseau résultant pour évaluer la qualité de nouveaux maillages 3D non vus, puisque quatre MLPRs ont été entrainés pour prédire les scores de qualité objectifs pour chaque fold. Dans cette section, nous proposons une nouvelle stratégie d'apprentissage qui permet d'obtenir un réseau neuronal unique pouvant être utilisé pour évaluer la qualité visuelle de maillages 3D n'appartenant pas à la base de données polyvalente LIRIS/EPFL et pouvant dépasser les performances des modèles obtenus par LOMO-CV. Pour ce faire, nous considérons un apprentissage cumulatif dont le principe est décrit dans la Figure 7. Cette stratégie d'apprentissage commence par entraîner et tester le modèle MLPR avec initialisation Glorot sur le premier fold pendant 1000 époques. Ensuite, le même MLPR est utilisé pour l'entrainement sur le fold suivant et ainsi de suite. Par conséquent, le MLPR cumulatif final peut être utilisé pour effectuer des prédictions futures sur des données non vues, ce qui contribue également à améliorer sa précision. Toutefois, si nous mesurons les performances de ce modèle final sur la base LIRIS/EPFL, les résultats sont manifestement surestimés, car cet apprentissage cumulatif a été progressivement entraîné sur l'ensemble de la base de données. Pour atténuer cet effet et mieux évaluer les performances du modèle cumulatif (CM), nous le ré-entraînons à l'aide de LOMO-CV afin d'obtenir un modèle cumulatif ré-entraîné (RCM) final. Pour ce faire, sur chaque fold, un nouveau MLPR est initialisé avec les poids du CM et est entraîné pour un nombre fixe d'époques. Ce dernier est déterminé en trouvant le minimum global de la moyenne de la fonction de coût sur tous les folds pendant l'entraînement cumulatif. Le tableau 3 présente les résultats du RCM en termes de corrélation de Spearman. Comme prévu, le modèle RCM est plus performant que le modèle de base que nous avons développé dans la section précédente. Le réajustement de ce modèle permet d'obtenir une meilleure évaluation de ses capacités de généralisation.

3.5 Comparaison avec l'état de l'art

Plusieurs approches ont récemment été proposées pour l'évaluation de la qualité des maillages 3D sans référence. Dans cette section, nous comparons notre approche à plusieurs méthodes existantes de l'état de l'art [10, 14, 8, 15, 16, 17, 18, 9, 12]. Les valeurs de corrélation $\mathbf{r_S}$ et $\mathbf{r_P}$ de nos méthodes (modèles de base ou cumulatifs), ainsi que celles des méthodes existantes, sont présentées dans le tableau 4. Nous mentionnons que les corrélations dans les colonnes "Tous les maillages" ont été calculées entre les scores prédits de tous les objets et leur MOS correspondants. Les méthodes que nous proposons présentent des scores de corrélation élevés. Nos méthodes de base basées sur les patchs (DCFQI-PBM) et sur les vues (DCFQI-

VBM) sont toutes deux plus performantes que CNNs-CMP [12] pour les modèles Armadillo, Venus et RockerArm. En outre, ils présentent des corrélations compétitives sur l'ensemble de la base de données, avec des valeurs de corrélation de $rs = 92,4\%$ et $rp = 92,1\%$. CNNs-CMP [12] utilise une approche complexe qui repose sur la combinaison de caractéristiques provenant de trois modèles pré-entraînés (VGG/AlexNet/ResNet) combinée à une sélection des patchs basée sur leur saillance. Notre approche montre que l'utilisation de patchs plus grands (ou même de vues) avec une optimisation minutieuse peut être aussi efficace. Enfin, les modèles cumulatifs basés sur les patchs (DCFQI-PCM) et sur les vues (DCFQI-VCM) que nous proposons surpassent l'état de l'art pour tous les maillages, à l'exception du modèle RockerArm, pour lequel nous sommes légèrement en deçà de la méthode BMQA-GSES [16]. Néanmoins, la performance globale de nos méthodes est supérieure à cette dernière. Si nous remplaçons l'agrégation moyenne des scores par une agrégation apprise par un MLP, on constate un gain significatif pour le modèle de base que ce soit pour les modèles basés vues ou patchs. Cette différence disparait lorsque l'on considère un modèle cumulatif avec des patchs et une moyenne est suffisante.

4 Conclusion

Dans cet article, nous avons présenté une approche d'évaluation de la qualité des maillages sans référence. Elle effectue un rendu du maillage en plusieurs projections 2D (vues) qui peuvent ensuite être subdivisées en patchs. À partir de ces images, des caractéristiques profondes sont extraites par le CNN VGG16 pré-entraîné et introduites dans un MLP qui effectue la prédiction de la qualité. Les prédictions des images de vues ou patchs sont ensuite agrégées pour obtenir la qualité d'un maillage. Ce modèle de base est compétitif par rapport à l'état de l'art, même en exploitant uniquement des vues. Enfin, un entraînement cumulatif a été proposé pour obtenir un modèle final unique de prédiction qui dépasse l'état de l'art. Les travaux futurs envisageront de combiner les prédictions au niveau des vues et des patchs, y compris pour des maillages colorés [19].

Remerciements

Ce travail de recherche a bénéficié d'un financement du PHC TOUBKAL TBK/22/142-CAMPUS N°47259YH.

Références

[1] G. Lavoué, E. D. Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3d distance metrics with application to watermarking," in *Applications of Digital Image Processing XXIX*, vol. 6312. SPIE, 2006, p. 63120L.

[2] N. Aspert, D. S. Cruz, and T. Ebrahimi, "MESH : measuring errors between surfaces using the hausdorff distance," in *ICME*, 2002, pp. 705–708. [Online]. Available : https ://doi.org/10.1109/ICME.2002.1035879

[3] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro : Measuring error on simplified surfaces," *Comput. Graph. Forum*, vol. 17, no. 2, pp. 167–174, 1998. [Online]. Available : https ://doi.org/10.1111/1467-8659.00236

[4] G. Lavoué, "A multiscale metric for 3d mesh visual quality assessment," *Comput. Graph. Forum*, vol. 30, no. 5, pp. 1427–1437, 2011. [Online]. Available : https ://doi.org/10.1111/j.1467-8659.2011.02017.x

[5] L. Vása and J. Rus, "Dihedral angle mesh error : a fast perception correlated distortion measure for fixed connectivity triangle meshes," *Comput. Graph. Forum*, vol. 31, no. 5, pp. 1715–1724, 2012. [Online]. Available : https ://doi.org/10.1111/j.1467-8659.2012.03176.x

[6] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3-d mesh quality assessment," *IEEE Trans. Multim.*, vol. 9, no. 2, pp. 247–256, 2007. [Online]. Available : https ://doi.org/10.1109/TMM.2006.886261

[7] K. Wang, F. Torkhani, and A. Montanvert, "A fast roughness-based approach to the assessment of 3d mesh visual quality," *Comput. Graph.*, vol. 36, no. 7, pp. 808–818, 2012. [Online]. Available : https ://doi.org/10.1016/j.cag.2012.06.004

[8] I. Abouelaziz, M. E. Hassouni, and H. Cherifi, "A convolutional neural network framework for blind mesh visual quality assessment," in *ICIP*, 2017, pp. 755–759. [Online]. Available : https ://doi.org/10.1109/ICIP.2017.8296382

[9] I. Abouelaziz, A. Chetouani, M. E. Hassouni, L. J. Latecki, and H. Cherifi, "3d visual saliency and convolutional neural network for blind mesh quality assessment," *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16 589–16 603, 2020. [Online]. Available : https ://doi.org/10.1007/s00521-019-04521-1

[10] I. Abouelaziz, A. Chetouani, M. E. Hassouni, and H. Cherifi, "A blind mesh visual quality assessment method based on convolutional neural network," in *3DIPM*. Ingenta, 2018. [Online]. Available : https ://doi.org/10.2352/ISSN.2470-1173.2018.18.3DIPM-423

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.

[12] I. Abouelaziz, A. Chetouani, M. E. Hassouni, L. J. Latecki, and H. Cherifi, "No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling," *Pattern Recognit.*, vol. 100, p. 107174, 2020. [Online]. Available : https ://doi.org/10.1016/j.patcog.2019.107174

		Armadillo		Dyno		Venus		RockerArm		Tous les maillages	
Type de méthode	**Métriques**	$\mathbf{r_s}$	$\mathbf{r_p}$	$\mathbf{r_s}$	$\mathbf{r_p}$	$\mathbf{r_s}$	$\mathbf{r_p}$	$\mathbf{r_s}$	$\mathbf{r_p}$	$\mathbf{r_s}$	$\mathbf{r_p}$
Basé caractéristiques	BMQI [14]	20.1	-	83.5	-	88.9	-	92.7	-	78.1	-
	BMQA-GSES [18]	98.7	80	99.2	80.4	98.8	80.1	**99.5**	**99.9**	90.5	87.9
	NR-GRNN [16]	87.1	97.3	91.2	94.1	86.3	85	78.6	74.8	86.2	88.7
	MVQ-GCN [17]	91.8	92.5	87.7	84.5	93.7	91.9	89.6	88.4	89.3	88.6
	NR-CNN 1 [8]	87.2	84.3	86.4	86.2	92.2	85.6	91.3	85.2	83.6	82.7
	NR-SVR [15]	76.8	91.5	78.6	84.1	85.7	88.6	86.2	86.6	81.5 8	7.8
Basé Vue	DCFQI-VBM	96.3	98.2	78.9	89	98.4	99.5	96.2	95.7	90.4	89.9
	DCFQI-VCM	99.2	99.8	99.5	**99,8**	99.7	99.8	98.6	99.2	96.5	96.6
	DCFQI-VBM-MLP	95.1	98.4	95.1	95.7	99.3	99.6	97	96.4	96.4	96.6
	DCFQI-VCM-MLP	98.8	99.8	99.8	99.9	99.3	99.8	98.8	99.4	98.7	99
Basé Patch	CNN-BMQA [9]	89.8	91.4	91.6	92.2	94.6	93.8	91.9	93.9	90	92
	NR-CNN 2 [10]	93.4	95.6	86.2	84.3	94.1	90.3	80.4	82.2	81.7	82.5
	CNNs-CMP [12]	95.8	95.6	93.6	92.9	93.4	91.3	94.5	95.2	92.6	91.3
	DCFQI-PBM	98.9	98	81.4	98.1	99.5	99.7	98.1	97	92.4	92.1
	DCFQI-PCM	**99.7**	**99.9**	**99.8**	**99.8**	**99.9**	**99.9**	99.1	99.6	**99.1**	**99.4**
	DCFQI-PBM-MLP	95,7	98,6	91,8	93,1	99,4	99,6	98,8	98,6	95,9	96,7
	DCFQI-PCM-MLP	**99.7**	**99.9**	99.7	**99.8**	99.8	99.8	**99.4**	**99.8**	99	99.2

TABLE 4 – Comparaison de notre approche DCFQI avec des modèles de base et cumulatifs basés sur les vues et les patchs (DCFQI-VBM & DCFQI-VCM / DCFQI-PBM & DCFQI-PCM respectivement) avec les métriques sans référence de l'état de l'art. Lorsque -MLP est précisé pour notre approche l'agrégation des scores est effectuée par un MLP.

[13] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *AISTATS*, vol. 9, 2010, pp. 249–256.

[14] A. Nouri, C. Charrier, and O. Lézoray, "3d blind mesh quality assessment index," in *3DIPM*, 2017, pp. 9–16. [Online]. Available : https ://doi.org/10.2352/ISSN.2470-1173.2017.20.3DIPM-002

[15] I. Abouelaziz, M. E. Hassimosouni, and H. Cherifi, "No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression," in *ICISP*, vol. LNCS 9680, 2016, pp. 369–377. [Online]. Available : https ://doi.org/10.1007/978-3-319-33618-3_37

[16] I. Abouelaziz, M. E. Hassouni, and H. Cherifi, "A curvature based method for blind mesh visual quality assessment using a general regression neural network," in *SITIS*, 2016, pp. 793–797. [Online]. Available : https ://doi.org/10.1109/SITIS.2016.130

[17] I. Abouelaziz, A. Chetouani, M. E. Hassouni, H. Cherifi, and L. J. Latecki, "Learning graph convolutional network for blind mesh visual quality assessment," *IEEE Access*, vol. 9, pp. 108 200–108 211, 2021. [Online]. Available : https ://doi.org/10.1109/ACCESS.2021.3094663

[18] Y. Lin, M. Yu, K. Chen, G. Jiang, F. Chen, and Z. Peng, "Blind mesh assessment based on graph spectral entropy and spatial features," *Entropy*, vol. 22, no. 2, p. 190, 2020. [Online]. Available : https ://doi.org/10.3390/e22020190

[19] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3d colored point cloud and mesh models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7618–7631, 2022. [Online]. Available : https ://doi.org/10.1109/TCSVT.2022.3186894