

Reconstruction tomographique faible dose en imagerie dentaire : régularisation TV versus UNet

Louise FRIOT--GIROUX, Théo LEULIET, Françoise PEYRIN, Voichita MAXIM

Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm,
CREATIS UMR 5220, U1294, F-69100, LYON, France

`louise.friot--giroux@creatis.insa-lyon.fr`

Résumé – La tomographie 3D avec faisceau conique (CBCT) s’est beaucoup développée dernièrement et en particulier en imagerie dentaire. Toutefois, les images dentaires tomodensitométriques souffrent d’artefacts dus aux projections tronquées, à la géométrie du faisceau conique ou encore à la présence de métal. La reconstruction CBCT dentaire en faible dose est donc un problème difficile et peu traité dans la littérature. Dans ce papier, nous évaluons une méthode itérative permettant d’introduire des informations a priori sur le volume et une méthode d’apprentissage profond sur des acquisitions réelles CBCT dentaires. Les résultats montrent que la méthode par apprentissage profond permet de reconstruire avec davantage de précision un volume avec seulement 20% de la dose normale.

Abstract – Cone-beam computed tomography (CBCT) is increasingly used and in particular in dental imaging. However, dental CT images suffer from artifacts due to truncated projections, cone beam geometry or the presence of metal inserts. Low-dose dental CBCT reconstruction is therefore a difficult problem and not well addressed in the literature. In this paper, we evaluate an iterative method allowing to introduce a priori volume information and a deep learning method on real dental CBCT acquisitions. The results show that the deep learning method is more efficient for low-dose tomography and can accurately reconstruct a volume with only 20% of the normal dose.

1 Introduction

La tomographie par rayons X consiste à reconstruire l’intérieur d’un volume à partir de projections de celui-ci. Elle a révolutionné la médecine en permettant aux spécialistes de visualiser l’intérieur des patients de manière non invasive. Cependant, l’exposition aux rayons X n’étant pas sans danger, un des grands enjeux actuels de l’imagerie par rayons X est la diminution de la dose absorbée par les tissus tout en gardant une qualité d’image suffisante pour effectuer un diagnostic correct.

Les méthodes de reconstruction les plus populaires sont les méthodes dites analytiques. En 2D, l’algorithme de rétro-projection filtrée reconstruit le volume de manière exacte sous des conditions idéales. Il est en revanche sensible au bruit - donc à une réduction de la dose - et présente des artefacts lors d’acquisition à géométrie limitée. Une généralisation de cet algorithme à la géométrie conique 3D (CBCT) a été proposée par Feldkamp (FDK) *et al.* [1]. En pratique, les mesures effectuées sur patient dépendent de nombreux paramètres physiques et contiennent du bruit, particulièrement présent dans les acquisitions à faible dose, ce qui rend la performance de ces solutions analytiques limitée.

Les méthodes de reconstruction itératives offrent plus de flexibilité pour la résolution du problème au prix de temps de calcul plus élevés. Le problème de reconstruction tomographique peut être modélisé par un problème linéaire qui consiste à retrouver le volume f à partir des projections p , avec A l’opérateur de projection : $Af = p$. Il peut être abordé par une approche variationnelle qui a pour objectif de résoudre le problème inverse en minimisant une fonctionnelle constituée d’un terme d’attache aux données et d’un a priori sur la solution recherchée. Parmi ces méthodes itératives, nous pouvons mentionner SIRT (Simultaneous Iterative Reconstruction Technique) [2] dans laquelle

le bruit sur les données est supposé Gaussien, MLEM [3] pour un bruit de type Poisson, et leurs versions régularisées respectives [4] et [5]. Dans [6], une étude sur les performances de ces méthodes itératives a été effectuée dans le cas d’un mélange de bruit Poisson-Gaussien.

L’application des méthodes d’apprentissage profond à l’imagerie médicale tomodensitométrique a permis d’obtenir des performances supérieures à celle des méthodes analytiques et itératives [7, 8] en terme de suppression de bruit notamment pour la reconstruction d’images tomodensitométriques. Ces méthodes ont pour principal avantage de ne nécessiter ni une connaissance précise du bruit présent sur les données, ni de recourir à un a priori explicite sur la solution. En effet, ces informations peuvent être déduites par le réseau à partir de la base de données dans le cas d’un apprentissage supervisé. Le temps de reconstruction est également un argument en faveur de ces méthodes d’apprentissage lorsqu’elles sont comparées aux méthodes itératives.

Si les méthodes d’apprentissage profond semblent plus performantes que les méthodes analytiques et itératives en terme de suppression de bruit, il est primordial d’évaluer cette performance avec des critères pertinents d’un point de vue de l’application clinique considérée.

Le but de ce travail est d’étudier la capacité d’un réseau de neurones de type UNet à supprimer les artefacts et à débruiter un volume reconstruit à partir d’acquisitions à faible dose dans le cas du CBCT dentaire. Nous comparons les performances de cette méthode avec une reconstruction itérative incluant un a priori de type TV (Total Variation), en se focalisant sur l’utilisation spécifique de ces reconstructions CBCT dentaires à des fins cliniques.

2 Méthodes

2.1 Reconstruction itérative

Les méthodes itératives effectuent la reconstruction en résolvant le problème de minimisation suivant :

$$f^* = \arg \min_f d(Af, p) + \alpha R(f) \quad (1)$$

où $d(Af, p)$ est le terme d'attache aux données et modélise les propriétés statistiques du bruit, $R(f)$ est un terme de régularisation qui force la solution à satisfaire certaines conditions et α est un paramètre faisant varier l'importance donnée à la régularisation. Dans le cadre d'acquisitions à faible dose, on peut supposer que les données suivent une distribution de Poisson. Le terme d'attache aux données est alors la divergence de Kullback-Leibler (KL) donnée par :

$$d_{KL}(Af, p) = \sum_i [Af - p \log(Af)]_i. \quad (2)$$

Pour l'a priori sur le volume, nous utilisons la variation totale (TV) [9]. Elle est définie par $TV(f) = \|\|\nabla f\|\|_1$, où ∇ est l'opérateur gradient, et $\|\|\cdot\|\|_1$ la somme des modules du gradient sur les voxels. La norme TV favorise les régions constantes par morceaux, réduisant ainsi le bruit dans les images, mais donnant un effet "en escalier" à celles-ci et modifiant leur texture.

Pour la résolution numérique de (1), dans cet article nous utilisons l'algorithme primal-dual de Chambolle-Pock [10]. Il s'applique à de nombreux problèmes d'optimisation et a été adapté à la reconstruction tomographique par Sidky et al. [11]. Nous utilisons ici une version pré-conditionnée de cet algorithme [12] que nous appelons $KL-TV_\alpha$. Nous avons cherché un équilibre entre atténuation des artefacts et préservation des détails qui nous conduit à des résultats avec deux valeurs pour le paramètre α , 0.05 et 0.1. Ce sont des valeurs limites, en dessous de 0.05, les reconstructions sont trop bruitées et au-dessus de 0.1, le lissage induit une trop grande perte de détails.

Pour limiter les artefacts dus aux projections tronquées, quasiment inévitables en imagerie dentaire CBCT, nous doublons la taille du volume reconstruit en ajoutant 150 voxels de chaque côté. Après la reconstruction, le volume est tronqué pour ne conserver que la région d'intérêt (ROI).

2.2 Post-traitement avec un UNet

Initialement utilisé pour effectuer des tâches de segmentation [13], le UNet peut également être un moyen d'améliorer la qualité d'une reconstruction dégradée par la présence de bruit et d'artefacts [7]. En effet, sa structure permet de récupérer les informations contextuelles de cette image dégradée afin de retrouver la qualité de l'image cible correspondante. Jin et al. [7] ont utilisé un réseau UNet pour débruiter et réduire les artefacts d'images reconstruites par rétroprojection filtrée à partir de peu de projections.

Nous utilisons un UNet classique de profondeur 5. Afin d'améliorer les performances du réseau, nous utilisons du *residual learning* et nous avons remplacé les activations ReLU (Rectified Linear Unit) par des activations Leaky ReLU définies par $LReLU_\lambda(x) = \max(x, \lambda x)$ qui ne s'annulent pas lorsque x est négatif. Nous avons en effet remarqué que les activations ReLU ont tendance à mettre à 0 des neurones lorsque des zones

dans l'image présentent des densités particulièrement élevées, causées par la présence de métal, créant des aberrations au niveau de ces zones.

Pour obtenir les vérités-terrain, nous avons reconstruit des volumes à partir d'acquisitions à dose normale avec une méthode analytique (FDK). L'entrée du réseau correspond à la reconstruction analytique obtenue en considérant une projection sur cinq, ce qui équivaut à 20% de la dose. Nous travaillons avec un UNet 2D entraîné dans un premier temps sur les coupes axiales. Nous testons également une configuration où trois UNets sont entraînés selon les différentes directions - un UNet pour chacune des coupes axiales, sagittales et coronales. Nous considérons ensuite une moyenne pondérée de ces trois sorties comme reconstruction finale, à laquelle nous nous référons par UNet-MP (Multi-Plans). Dans cette étude, nous ne considérons pas de UNet-3D qui nécessiterait une plus grande base de données.

Les réseaux ont été entraînés sur 50 époques avec une taille de batch égale à 10. L'optimiseur choisi est Adam et après l'étude de la valeur optimale du taux d'apprentissage sur le jeu de données de validation, nous l'avons fixé à 10^{-3} pour les coupes axiales et coronales, et à 10^{-2} pour les coupes sagittales. La moyenne de l'erreur absolue (MAE) a été utilisée comme fonction coût et nous avons choisi $\lambda = 0.2$ comme paramètre pour les activations.

2.3 Base de données

Nous disposons d'une base de données anonymisée d'acquisitions sur 32 patients des deux sexes, âgés de 18 à 70 ans. Les projections coniques ont été acquises en dose normale, avec un shortscan par un scanner CS 8100SC 3D. Chaque volume contient un nombre moyen de $280 \times 280 \times 320$ voxels, selon la taille de champ de vue choisie, de 300 μm de côté.

Nous avons séparé la base de données en trois parties. Les données de 25 patients ont été utilisées pour l'entraînement, 6 volumes ont été utilisés pour la validation et l'ajustement des hyper-paramètres. Le dernier volume a servi de test, c'est celui qui est présenté dans cet article. La figure 1 montre trois coupes de patients tirés des données d'entraînement ainsi qu'une coupe de celui sur lequel nous effectuons les tests.

3 Résultats

Les reconstructions itératives et les entraînements ont été réalisés avec un processeur Intel Xeon Gold 6226R et une GPU Nvidia Tesla V100. L'algorithme de FDK ainsi que les opérations de projection et de rétro-projection ont été effectués à l'aide de la bibliothèque ASTRA [14].

La figure 2 présente les résultats des différentes méthodes sur le volume d'évaluation. La première ligne illustre une coupe axiale où l'on voit que la seconde molaire maxillaire gauche en métal génère des artefacts. Avec FDK et $KL-TV_{0.05}$, les artefacts métalliques sont encore très visibles, alors que les trois autres méthodes les suppriment correctement. Les canaux de la première molaire gauche (indiqué par une flèche bleue) sont mieux résolus avec les UNet.

La deuxième ligne est un agrandissement sur une coupe axiale ne contenant pas de métal. On voit que le ligament alvéolo-dentaire (flèche orange) n'est pas visible sur les reconstructions

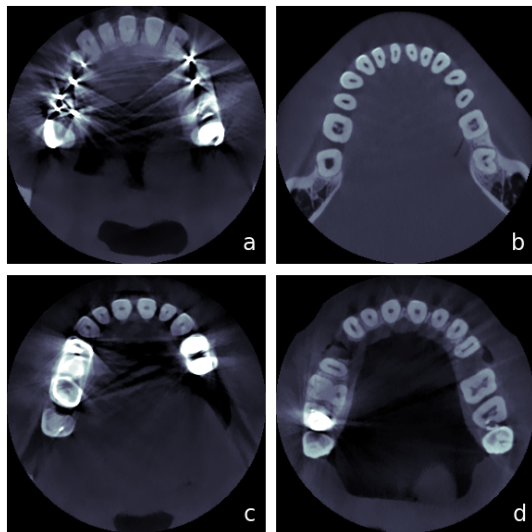


FIGURE 1 – a, b, et c : Coupes de patients utilisés pour l’entraînement. d : Patient utilisé pour les tests.

itératives, et apparaît plus net sur la reconstruction du UNet moyenné que sur celle du UNet axial.

Sur les deux dernières lignes, il est intéressant d’observer la longueur du canal de la dent sur la coupe sagittale (flèche jaune) et la qualité de la reconstruction du canal mandibulaire (flèche rose). La base du canal n’est pas visible sur les images KL-TV et la structure de l’os alvéolaire est supprimée. Les contours du canal mandibulaire restent très flous pour ces deux reconstructions également. Le traitement coupes par coupes du UNet axial laisse apparaître de légères rayures horizontales qui sont atténuées par le UNet-MP.

Afin de comparer quantitativement les reconstructions, nous avons utilisé le PSNR (Peak Signal to Noise Ratio), NRMSE (Erreur quadratique moyenne normalisée) et le SSIM (Structural Similarity).

La table 1 résume les différentes valeurs des métriques utilisées pour comparer les méthodes. Nous les avons appliquées sur le volume total et sur trois ROIs correspondant respectivement aux trois agrandissements présentés dans la figure 2. Les métriques des deux reconstructions par méthode itérative sont meilleures que celles de la reconstruction avec FDK et ont des résultats très similaires entre elles. La reconstruction avec $\alpha = 0.1$ a de meilleurs PSNR et NRMSE, mais le SSIM avec $\alpha = 0.05$ qui lisse moins est supérieur. Les méthodes d’apprentissage profond surpassent largement les résultats des autres méthodes, notamment pour le UNet moyenné. Les artefacts produits par les projections tronquées pourraient être à l’origine de la différence observée entre les valeurs NRMSE de la méthode itérative et des UNets. Malgré l’augmentation de la taille du volume, les projections tronquées engendrent une augmentation de l’intensité des voxels.

Pour les temps de calcul, il a fallu 35 minutes pour reconstruire itérativement le volume avec 500 itérations, 5 heures d’entraînement pour chaque réseau, et une quinzaine de secondes pour l’inférence.

4 Discussion et conclusion

La reconstruction itérative permet d’améliorer la qualité visuelle par rapport à FDK, mais la perte de détails cruciaux, tels que la longueur des canaux dentaires, engendrée par la régularisation TV sont un frein pour leur utilisation clinique. De plus, les résultats dépendent du choix du paramètre de régularisation qui doit être ajusté en fonction du niveau de bruit présent dans les projections. Les temps de calcul assez longs et les ressources numériques nécessaires présentent également des inconvénients dans l’utilisation de ces méthodes.

Nous observons une réduction des artefacts métalliques bien qu’aucun algorithme de réduction de ces artefacts n’ait été utilisé. La qualité pourrait être améliorée avec de tels algorithmes.

Le post-traitement avec les UNet donne de très bons résultats en des temps suffisamment courts pour une utilisation médicale. Les comparaisons et l’évaluation effectuées montrent que la reconstruction FDK + UNet-MP conserve bien les détails nécessaires à un diagnostic correct, tout en ayant un aspect visuel agréable.

De plus, l’entraînement repose sur une reconstruction FDK, qui n’est exacte que dans le plan de la source. Par conséquent, il serait intéressant d’inclure le processus de reconstruction dans le réseau de neurones.

Dans des travaux futurs, nous pourrions utiliser un réseau de neurones un peu plus complexe qu’un UNet-2D. En effet, nos résultats montrent que la cohérence spatiale est améliorée par le UNet-MP par rapport au UNet axial. En disposant d’un plus grand nombre de données pour éviter le sur-apprentissage, un UNet-3D pourrait davantage améliorer cette cohérence spatiale.

Références

- [1] L.A. Feldkamp et al. Practical cone-beam algorithm. *Josa a*, 1(6) :612–619, 1984.
- [2] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of theoretical biology*, 36(1) :105–117, 1972.
- [3] K. Lange et al. EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr*, 8(2) :306–16, 1984.
- [4] H. Banjak et al. Evaluation of noise and blur effects with SIRT-FISTA-TV reconstruction algorithm : Application to fast environmental transmission electron tomography. *Ultramicroscopy*, 189 :109–123, 2018.
- [5] V. Maxim et al. Tomographic reconstruction from Poisson distributed data : a fast and convergent EM-TV dual approach. working paper or preprint, October 2018.
- [6] T. Leuliet et al. Efficiency of TV-regularized algorithms in computed tomography with Poisson-Gaussian noise. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1294–1298. IEEE, 2021.
- [7] K.H. Jin et al. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9) :4509–4522, 2017.

Ces travaux ont été réalisés dans le cadre du LABEX PRIMES (ANR-11-LABX-0063) de l’Université de Lyon, au sein du programme "Investissements d’Avenir" (ANR-11-IDEX-0007). Ces travaux ont bénéficié d’un accès aux moyens de calcul de l’IDRIS au travers de l’allocation de ressources 2021-AD011012734 attribuée par GENCI. Nous remercions Carestream Dental de nous avoir fourni les données réelles.

