

Détection de la coupe L3 par CNN et transfert learning

S. Belharbi, C. Chatelain, R. Hérault, S. Adam, R.
Modzelewski, M. Chastan, S. Thureau

LITIS - Équipe Apprentissage - Centre Henry Becquerel de Rouen

`soufiane.belharbi@insa-rouen.fr`

21 octobre 2016

Plan

Introduction

Problématique

Approche proposée

Résultats

Conclusion

Sommaire

Introduction

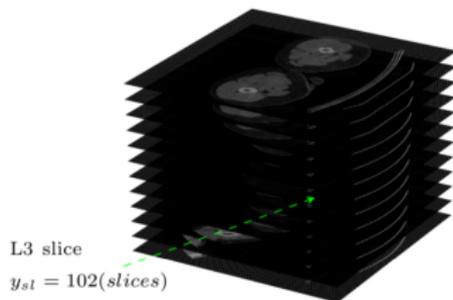
Problématique

Approche proposée

Résultats

Conclusion

Introduction



- Sur une base de **642 CT scans** on a obtenu une **erreur moyenne de localisation < 1.75 coupes (< 5 mm)**.

Sommaire

Introduction

Problématique

Approche proposée

Résultats

Conclusion

Détection de la coupe L3

Étant donné un scan d'une partie du corps, trouver la coupe correspondante à la coupe L3 parmi quelques centaines de coupes

difficultés

- ▶ variabilité inter-patients
- ▶ ressemblances coupe L3 avec d'autres coupes
- ▶ nécessité d'utiliser le contexte pour décider de la L3

⇒ Machine Learning

Approches envisageables

Classification de coupes (prédire une valeur discrète)

Pour chaque coupe décider "L3"/"pas L3"

- ▶ Approche simple, 😊
- ▶ Pas de contexte 😞

Étiquetage de séquences

Chercher à trouver toutes les vertèbres L1, L2, L3, ...

- ▶ Analyse globale : contexte, 😊
- ▶ Travaux existants semblent prometteurs, 😊
- ▶ Nécessite étiquetage de chaque coupe 😞

Regression (prédire une valeur continue)

Étant donnée toute la séquence, estimer la hauteur de la coupe L3

- ▶ Analyse globale : contexte, 😊
- ▶ Pas de travaux existants,
- ▶ Nécessite seulement la position de la L3 en étiquetage 😊

Sommaire

Introduction

Problématique

Approche proposée

Résultats

Conclusion

Regression pour la détection de la coupe L3

Choix du modèle statistique pour la regression

- ▶ Deep learning, Convolutional Neural Network (CNN)
- ▶ Pas de caractéristiques à extraire
- ▶ State-of-the-art sur l'image
- ▶ Modèle statique : requiert un espace d'entrée de taille fixe

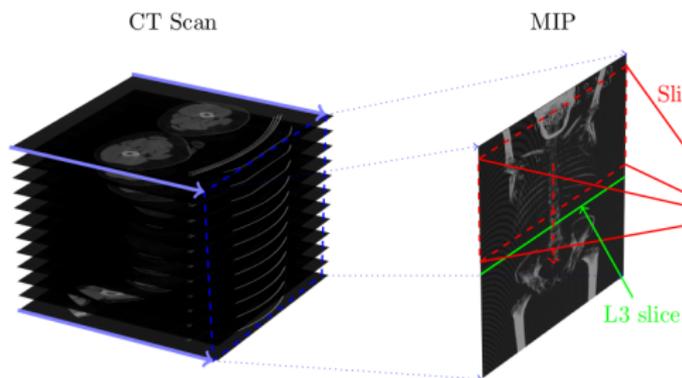
Parlons chiffres

- ▶ Espace d'entrée : 1 scan = $N \times 512 \times 512$, avec $400 < N < 1200$
- ▶ Scans avec L3 annotée et vérifiée : 642 patients
- ▶ La hauteur des scans N est variable d'un scan à un autre

Solution au problème N°1

Problème N°1 : dimension espace d'entrée

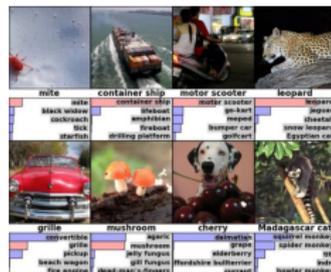
- ▶ 131 M entrées par exemple : ça fait beaucoup même pour un CNN !
- ▶ → Maximum Intensity Projection (MIP) facial ou latéral
- ▶ On passe de $512 \times 512 \times N$ à $512 \times N$
- ▶ Information pertinente *a priori* conservée



Solution au problème N°2

Problème N°2 : Peu de données annotées (642 patients)

- ▶ On a essayé quand même ! → résultats mitigés
- ▶ Solution : utilisation de modèles pré-appris sur des grandes quantités de données
- ▶ AlexNet, GoogleNet, VGG16, VGG19, etc.
- ▶ Modèles appris sur ImageNet : 14 millions d'images naturelles annotées [Fei-Fei and Russakovsky 2013]



Solution au problème N°2 (suite)

Idée générale : Transfer Learning

Bénéficier des filtres convolutionnels préappris sur des images naturelles (!!!), qu'on affine ensuite sur nos données.

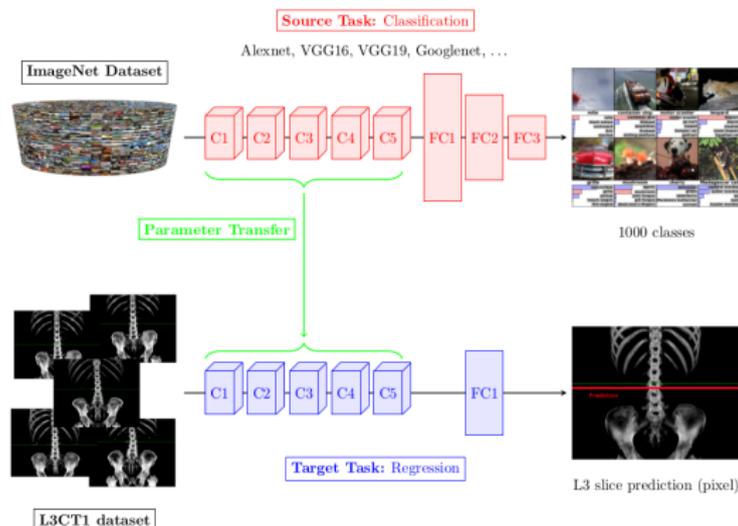
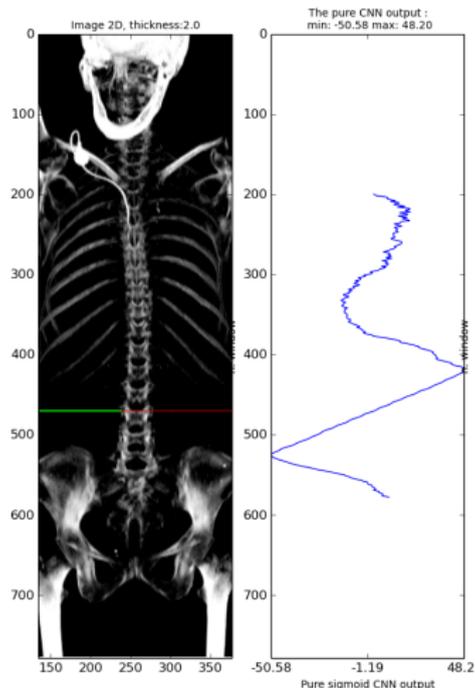


FIGURE 1 : Transfert des filtres

Solution au problème N°3

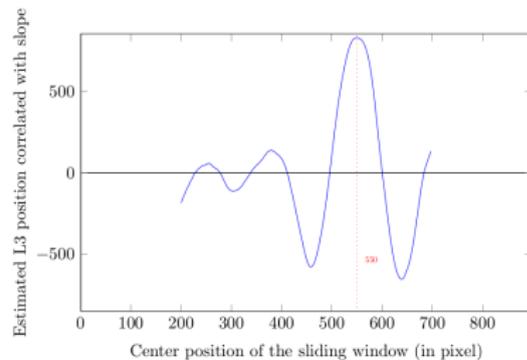
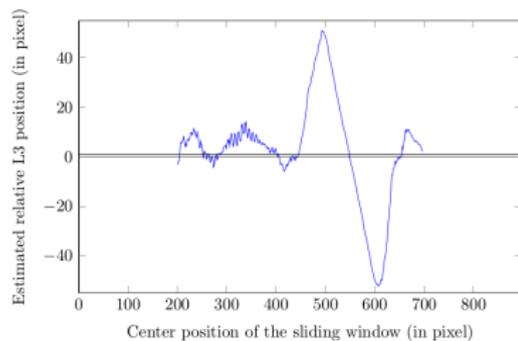
Problème N°3 : Espace d'entrée de taille variable

- ▶ Problème assez classique
- ▶ Utilisation d'une fenêtre glissante
- ▶ Post-traitement pour analyser la sortie du CNN aux différentes positions



Post-traitement

- Corrélation.



Système complet : Prédiction

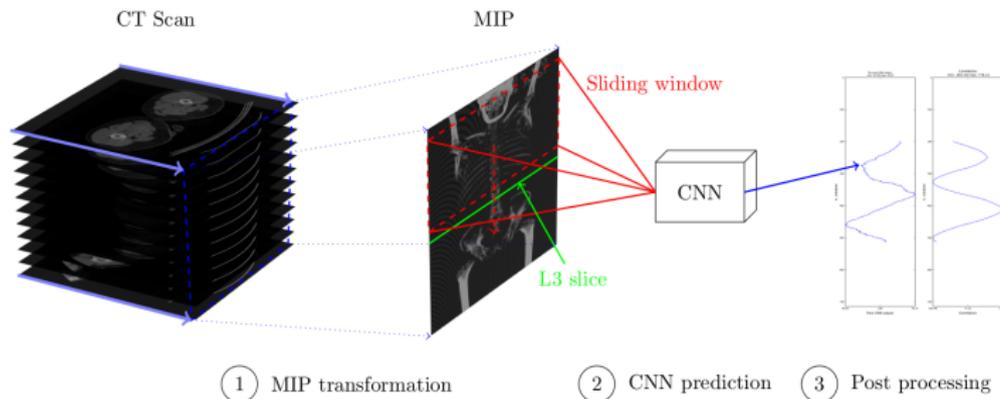


FIGURE 2 : Système complet : Prédiction

Sommaire

Introduction

Problématique

Approche proposée

Résultats

Conclusion

Résultats quantitatifs

Erreur en cross validation : comparaison d'un CNN "maison" et des CNN pré-appris + transfer learning

	CNN4	Alexnet	VGG16	VGG19	Googlenet
fold 0	4.64 ± 14.66	2.24 ± 2.28	1.71 ± 1.59	1.83 ± 1.81	1.78 ± 1.70
fold 1	4.94 ± 16.29	2.43 ± 2.46	1.77 ± 2.14	1.96 ± 2.10	2.19 ± 2.13
fold 2	3.10 ± 3.22	2.42 ± 2.43	1.57 ± 1.62	1.75 ± 1.80	2.73 ± 2.48
fold 3	3.03 ± 2.45	2.46 ± 2.33	1.85 ± 1.60	1.80 ± 1.64	2.16 ± 1.85
fold 4	3.07 ± 8.97	2.69 ± 2.37	1.83 ± 2.08	1.83 ± 1.75	2.12 ± 2.25
Average	3.75 ± 9.19	2.45 ± 2.37	1.75 ± 1.80	1.83 ± 1.82	2.20 ± 2.08

TABLE 1 : L'erreur de test en cross validation (avec post-traitement) (coupes).

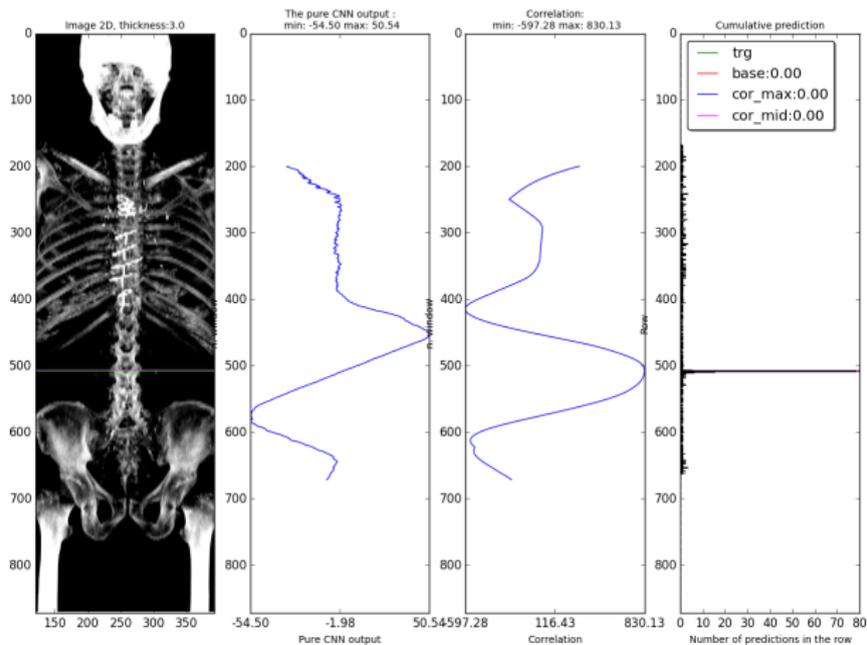
Temps d'évaluation

	Nombre de paramètres	Temps moyen de traitement (seconds/CT scan)
CNN4	55 K	04.46
Alexnet	2 M	06.37
VGG16	14 M	13.28
VGG19	20 M	16.02
Googlenet	6 M	17.75

TABLE 2 : Nombres de paramètres et temps d'évaluation d'un scan sur un GPU (K40).

Résultats Qualitatifs

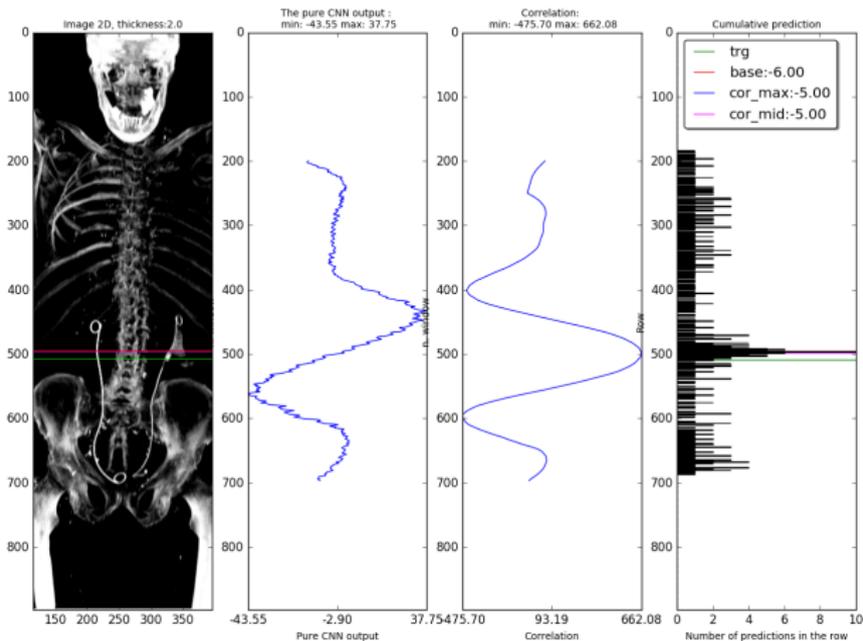
Prediction for pt_id: 165_5112614581.



Erreur : 0 coupes.

Résultats Qualitatifs

Prediction for pt_id: 1_9352086790.



Erreur : 6 coupes.

CNN vs. radiologistes

- Nouvelle base : 43 CT scans annotés par le radiologiste de référence.
- Faire prononcer 3 autres radiologistes sur la base.

Erreurs (coupes) / operateur	CNN4	Alexnet	VGG16	VGG19	Googlenet	Radiologiste #1	Radiologiste #2	Radiologiste #3
Erreur	2.60 ± 2.88	3.51 ± 4.52	1.72 ± 1.55	2.19 ± 1.73	2.16 ± 2.20	0.81 ± 0.97	0.72 ± 1.51	0.51 ± 0.62

TABLE 3 : Comparaison : CNN et radiologistes.

- 1.75 coupes ~ 5 mm.
- Nous avons perdu le combat avec les radiologistes mais avec les honneurs 😊.

Sommaire

Introduction

Problématique

Approche proposée

Résultats

Conclusion

Conclusion

Conclusion

- ▶ Résultats intéressants
- ▶ Pipeline adapté : prétraitement / CNN / post traitement
- ▶ Généricité : utilisation du transfer learning requiert assez peu de données et de temps d'apprentissage. → test sur d'autres problèmes ?

Questions ?

soufiane.belharbi@insa-rouen.fr