

Orthogonalité et diversité dans les filtres de ConvNet pour la segmentation

8 Avril 2021

Thomas DARGENT

Directrice de thèse : Pr. Caroline PETITJEAN (LITIS)

Co-encadrant : Florian YGER (LAMSADE)



Contexte (et rappels)

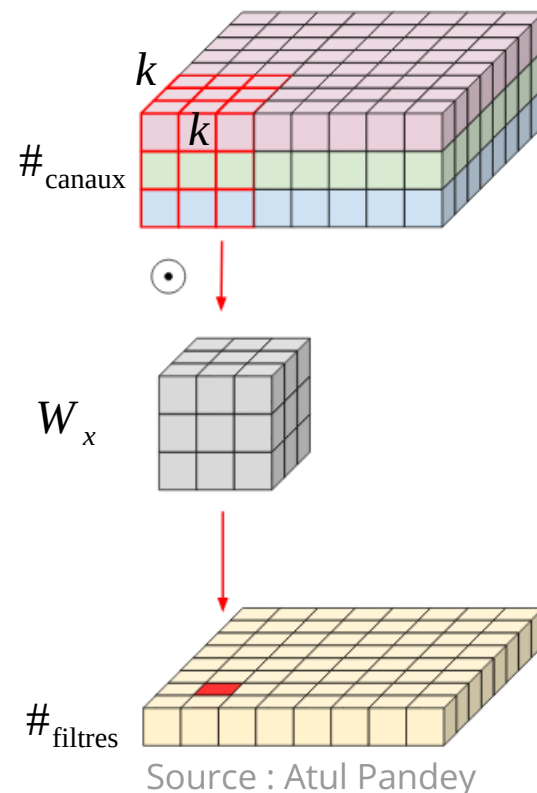
- Matrice $A \in \mathbb{R}^{n \times n}$ est orthogonale si $A^T A = I$
- Matrice $U \in \mathbb{R}^{m \times n}$ est semi-orthogonale si $U^T U = I$
- Propriétés importantes de la semi-orthogonalité :

Si $m \geq n$, U préserve la norme $\|Ux\|_2 = \|x\|_2$

Si $m < n$, U est simplement non expansif (contraction) $\|Ux\|_2 \leq \|x\|_2$

Contexte (et rappels)

- Convolution orthogonale si :
 - $\|conv(X)\| = \|X\|$
 - $\#_{\text{canaux}} = \#_{\text{filtres}}$ (semi orthogonale sinon)
- Filtres orthogonaux («kernel orthogonality») si :
 - Pour $W \in \mathbb{R}^{\#_{\text{filtres}} \times (\#_{\text{canaux}} k^2)}$ $W W^T = I_{\#_{\text{filtres}}}$
 - Nécessaire mais pas suffisant pour atteindre une convolution orthogonale (Wang et al.)



Motivation

- Pour les RNN et autres réseaux non-convolutionnels :
- Possibles intérêts de **contraindre des poids** à l'orthogonalité :
 - Poids proches de l'orthogonalité : facilite la convergence et améliore la généralisation (Pennington et al. 2017)
 - Préserve les normes des gradients → limite leurs explosions & disparitions (Rodriguez et al. 2017, Vorontsov et al. 2017, Massart et al. 2021)
 - Alternative à la batch normalization (Trockman et al. 2021)
 - Peut aider à se rappeler les dépendances au long terme dans les RNN
 - Augmente la robustesse aux exemples adverses (faibles constantes de Lipschitz)(Tsuzuku et al. 2018)

Motivation

- Pour les ConvNets :
- Possibles intérêts de **contraindre des filtres** à l'orthogonalité :
 - Facilite la convergence et améliore la généralisation ([Bansal et al. 2018](#))
 - Assure la diversité des filtres → moins redondant donc meilleure utilisation de la capacité du réseau ([Wang et al. 2020](#))

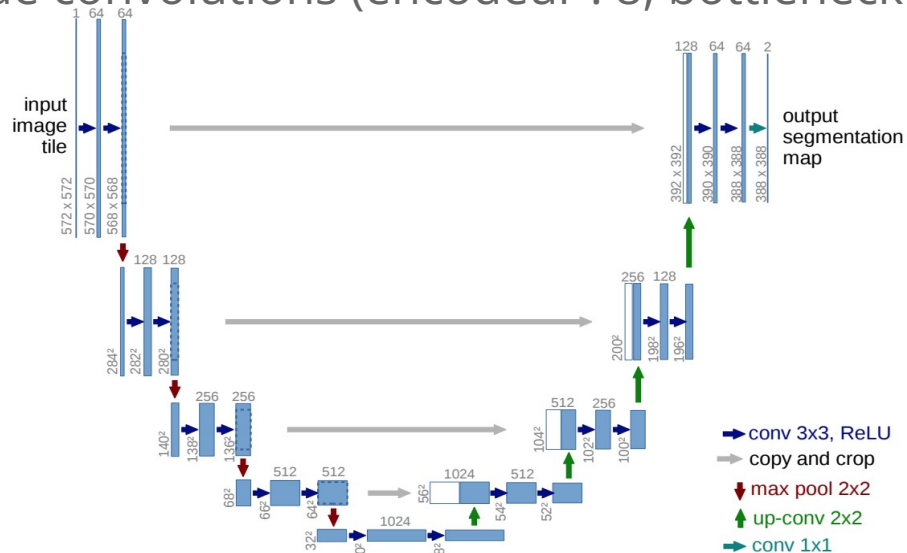
Motivation

- Pour les ConvNets :
- Possibles intérêts de **contraindre des convolutions** à l'orthogonalité :
 - Meilleurs scores (Trockman et al. 2021, Wang et al. 2020)
 - Augmente la robustesse aux exemples adverses (faibles constantes de Lipschitz)(Trockman et al. 2021)
- L-Lipschitzité du réseau f :

$$\|f(x_1) - f(x_2)\| \leq L \|x_1 - x_2\|$$

Réseau de segmentation

- Que donne cette régularisation sur des problématique de **segmentation** ?
- Tache de segmentation médicale → utilisation d'un **UNet** basique (Ronneberger et al. 2015)
 - 18 couches de convolutions (encodeur : 8, bottleneck : 2, décodeur : 8)

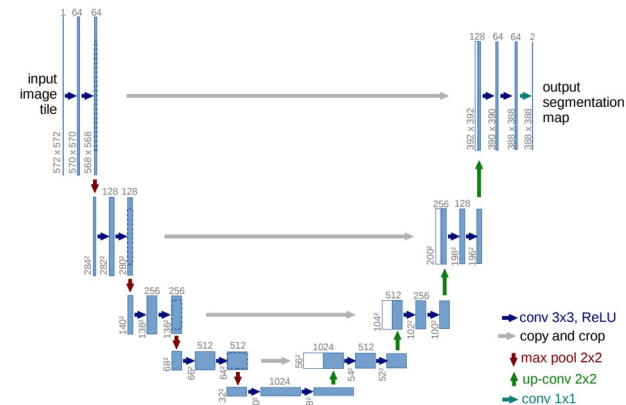
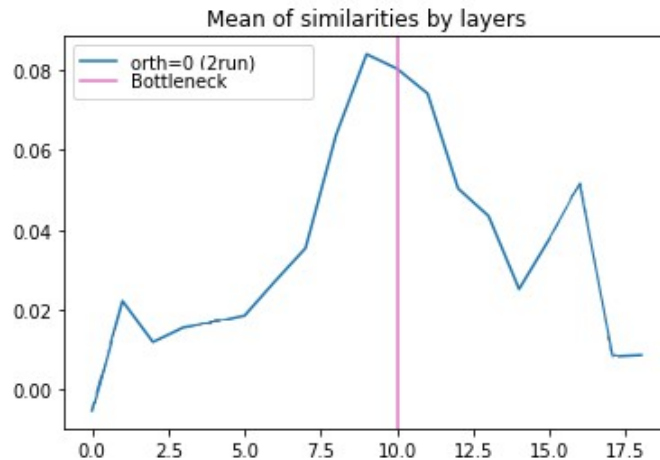


Mesure de diversité

- Qu'en est il de la **diversité des filtres** dans un **UNet entraîné** → choisir une **mesure de similarité**
 - Utilisation du cosinus entre les deux vecteurs :

$$\frac{u \cdot v}{\|u\| \|v\|}$$

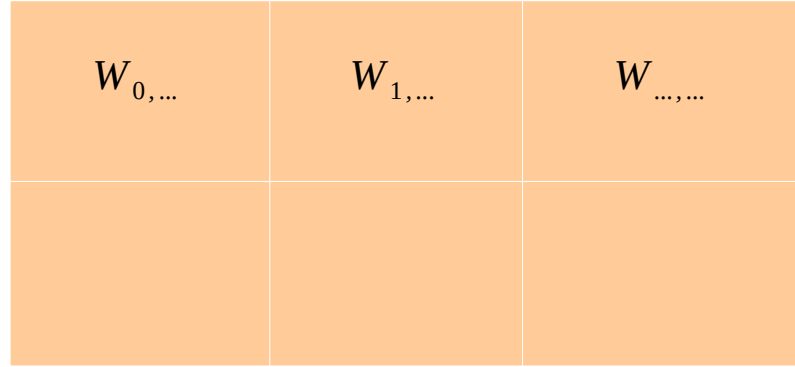
- Sur UNet standard : filtres plus profonds très corrélés donc redondant ([Wang et al. 2020](#))



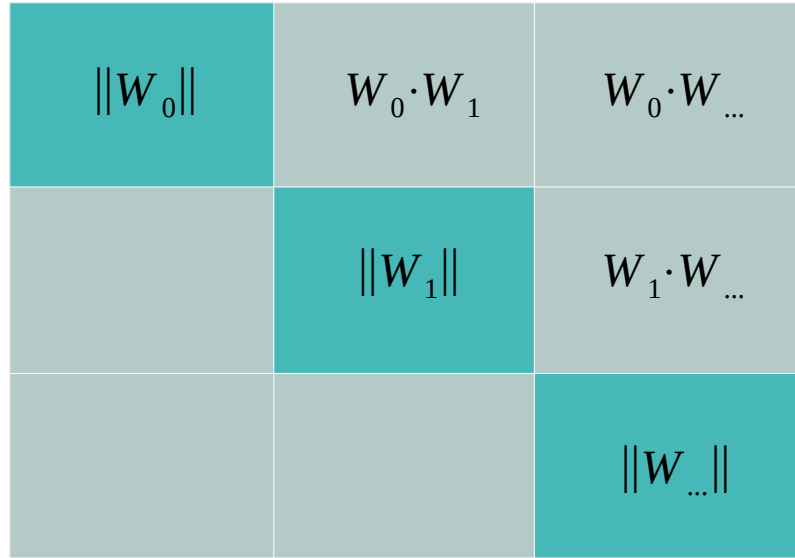
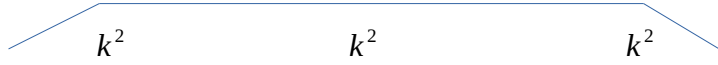
Impact de contraintes sur la diversité

$$W W^T = I_{\# \text{filtres}}$$

canaux $\times k^2$



canaux $\times k^2$



Impact de contraintes sur la diversité

- Cosinus = produit scalaire normalisé, si = 0 → vecteurs seront dissimilaires.
- Forcer l'orthogonalité revient donc à diminuer la redondance entre filtres.

$\ W_0\ $	$W_0 \cdot W_1$	$W_0 \cdot W_{\dots}$
	$\ W_1\ $	$W_1 \cdot W_{\dots}$
		\dots

Méthode utilisée

Comment appliquer cette orthogonalité de filtres dans un ConvNet ?

- Pénalité utilisée dans la loss → Ex. «**soft orthogonality**» (Bansal et al. 2018)

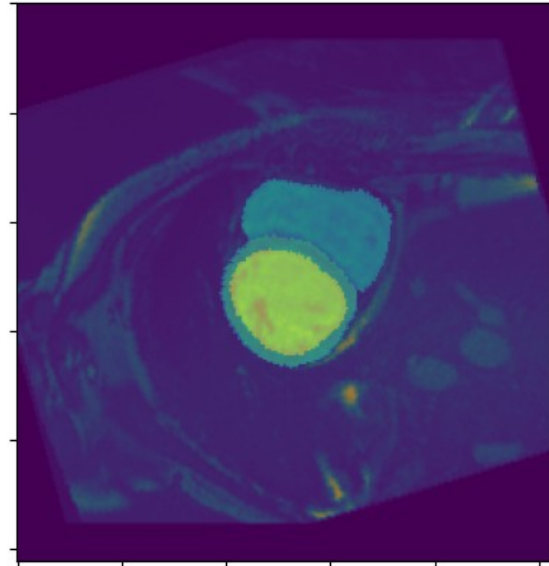
$$Orth(L) = \frac{1}{|L|} \sum_{l \in L} \|W_l \cdot W_l^T - I\|_F$$

$$Loss(x, y, L) = 0,8 \times CE(x, y) + 0,2 \times DSC(x, y) + \lambda Orth(L)$$

- Utilisation de la transformée de Cayley (Li et al. 2020, Trockman et al. 2021)
- Optimisation dans la variété de Stiefel (Vorontsov et al. 2017)

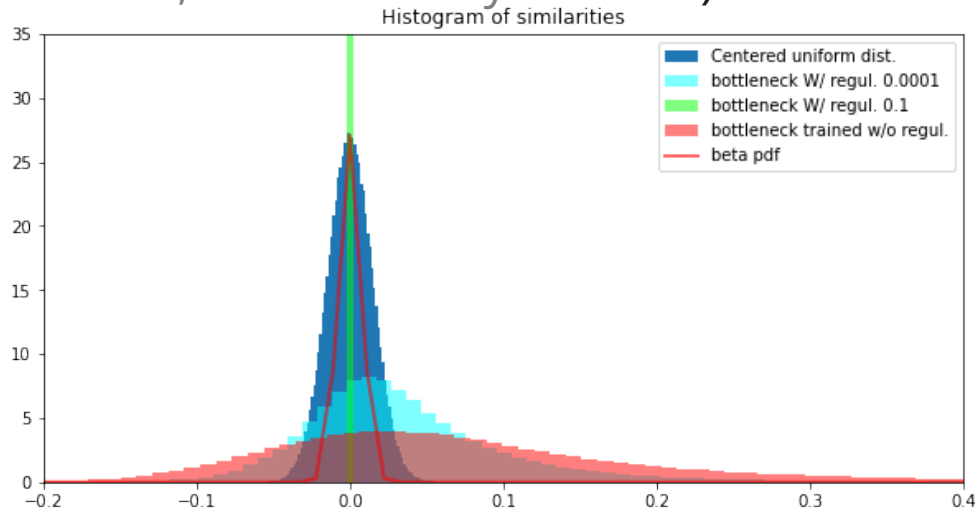
Jeu de données

- On utilise le jeu de données **ACDC** (Automated Cardiac Diagnosis Challenge)
 - 100 sujets en entraînement (951 images)
 - Split en 776 train/ 175 val
 - 4 classes :
 - fond,
 - cavité du ventricule droit,
 - cavité du ventricule gauche,
 - myocarde gauche



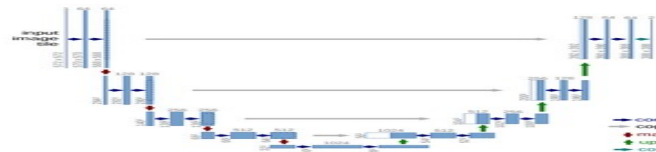
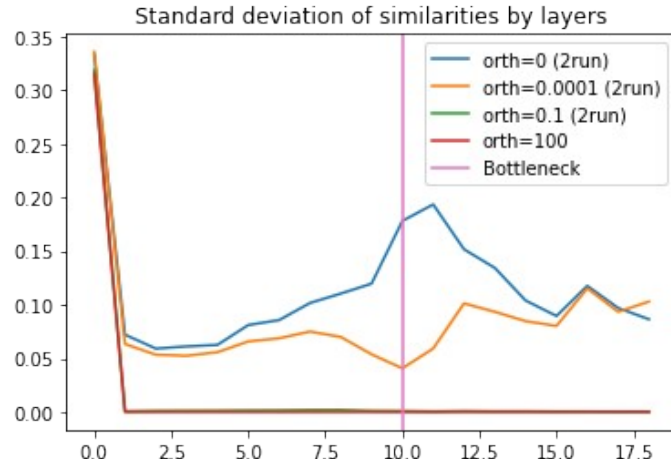
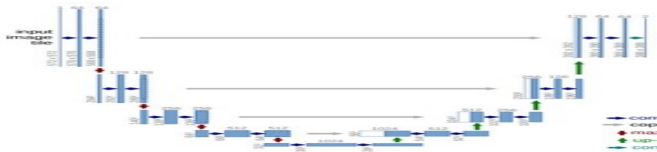
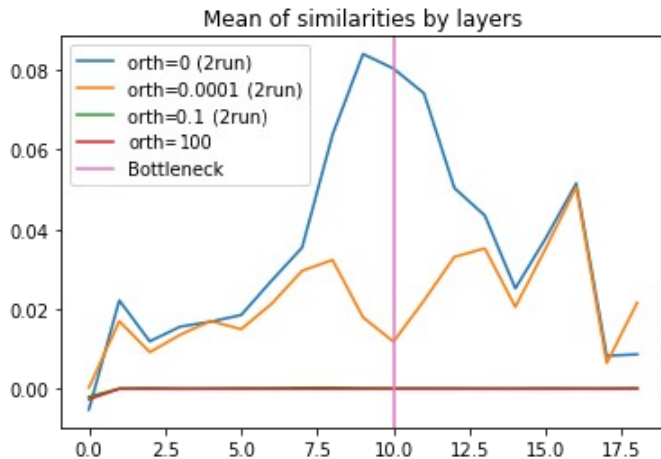
Résultats

- L'entraînement tend à rendre plus de filtres similaires (plus grand écart type de la distribution de similarité)
- Il semble également augmenter la moyenne des similarités, à étudier.
- Augmenter la régularisation enlève bien la redondance (réduit l'écart type de la distribution de similarité, et met la moyenne à 0)



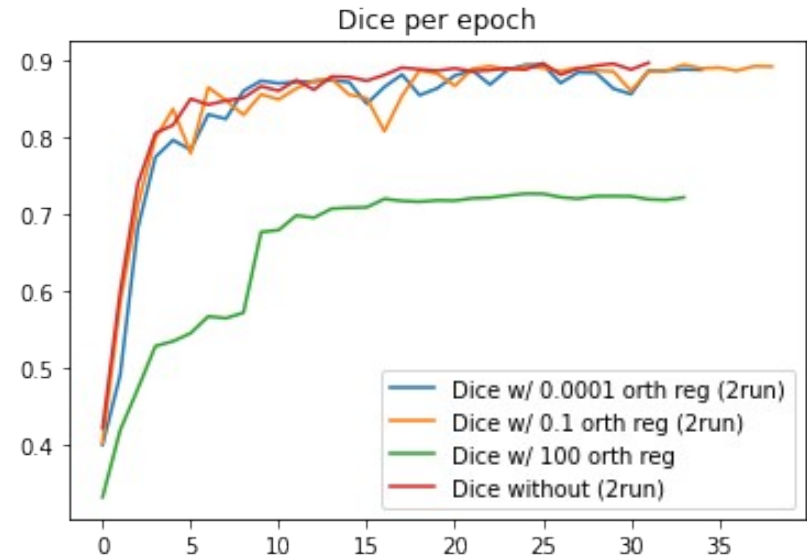
Résultats sur la diversité

- Plus de similarité au niveau du bottleneck, réduite par la régularisation
- Augmenter λ entraîne la moyenne et l'écart type des distributions vers 0
 - ↳ Toute redondance entre les filtres est supprimée !



Résultats sur le score

- Pas de différences notables en score ou convergence observés.
- Problème trop facile ? → diversité inutile ?
 - Essais en cours sur CityScape



- Reste à étudier la résistance aux problèmes adverses

Merci !

