

U-Net: Convolutional Networks for Biomedical Image Segmentation.

Amar BACHA, Razik DJAFRI

Décembre 2020

Résumé :

Ce document est un résumé de l'article intitulé « U-Net : Convolutional Networks for Biomedical Image Segmentation » publié le 8 Mai 2015 par Olaf Ranneberger, Philipp Fischer et Thomas Brox.

Dans cet article les auteurs nous parlent des approches qui existait à l'époque pour la segmentation des images dans le domaine biomédical. Ils nous présentent leurs propres approches, il s'agit des réseaux de convolution U-Net. L'article se termine par une partie expérimentale où les auteurs abordent les très bons résultats obtenus lors de leurs participations à 2 challenges de segmentation médicales.

1. Introduction

Le deep-learning nécessite de grandes quantités de données étiquetées pour entraîner des modèles, il n'y a malheureusement pas assez de données dans le secteur biomédical. C'est pourquoi ils ont eu l'excellente idée du réseau U-Net, ils se sont concentrés spécifiquement sur l'exigence d'une petite quantité de données pour l'entraînement et il se sont basés sur l'utilisation intensive de l'augmentation des données afin de créer un modèle plus général.

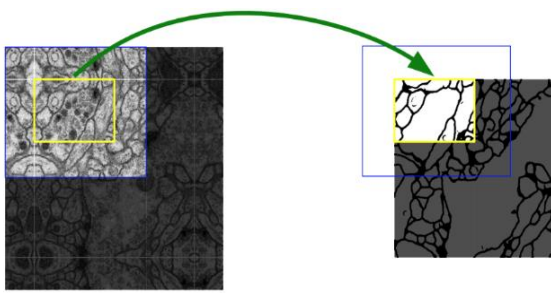
Les auteurs ont abordé les approches précédentes existantes, telle que le réseau convolutif de fenêtres coulissantes qui était considéré à l'époque comme solution de l'état de l'art pour la segmentation sémantique. Et la façon dont ils l'ont utilisé est assez intéressante car ils ont essentiellement préparé des correctifs comme des régions locales autour du pixel d'entrée à prédire (classer) et utiliser la technique des fenêtres coulissantes pour classer chaque pixel.

2. Le réseau U-Net

2.1 La théorie derrière l'architecture de U-net :

- Proposition de réseau entièrement convolutif.
- L'idée de base est d'utiliser un chemin de sous-échantillonnage pour l'extraction d'entités, suivi d'un suréchantillonnage pour une localisation précise de ces entités dans des couches à plus haute résolution.
- En concaténant les cartes d'entités de sous-échantillonnage avec les couches de suréchantillonnage correspondantes, nous aidons les couches successives des couches convolutives à assembler une sortie plus précise.

- **Overleap tile strategy** : La prédiction de la segmentation dans la zone jaune nécessite des données d'image dans la zone bleue comme entrée.



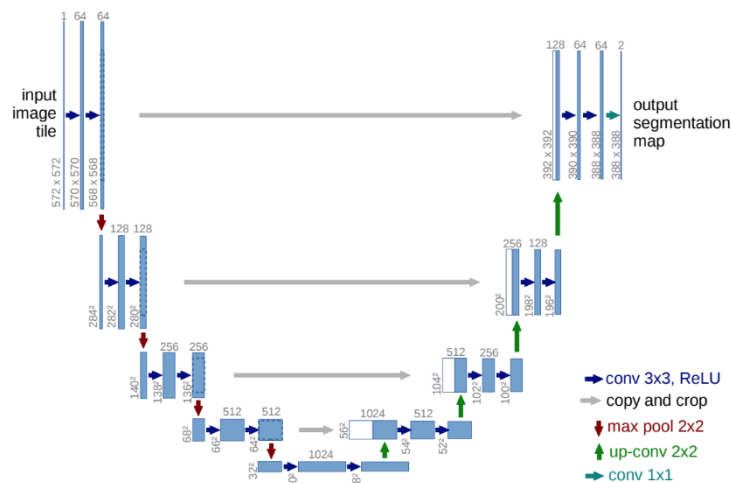
Input > output : car pas de padding dans chaque couche de *conv*.

La carte de segmentation en sortie contient uniquement les pixels pour lesquels le contexte complet est disponible dans l'image d'entrée.

Le contexte manquant sur les bords de l'image d'entrée est extrapolé par mise en miroir

2.2 Architecture du réseau :

Le U-net comprend deux parties : un chemin de codeur/contractant (côté gauche) pour capturer le contexte et un chemin décodeur/expansif (côté droit), ce qui lui confère une architecture en forme de « U ».



Chemin de contraction :

- 4 blocs de *conv* (2 couches de *conv* chacun) suivis de couches de *maxpooling 2x2* avec stride de 2 pour le sous-échantillonnage.
- 5ème bloc de *conv* sans *maxpooling* (connexion avec le chemin d'expansion).
- Premier bloc *conv* avec 64 filtres sur chaque couche *conv*.
- Le nombre de filtres est doublé à chaque bloc de *conv* consécutif.
- Réduire la résolution, augmenter la profondeur (le nombre de couches).
- Pas de padding.
- Filtres 3X3 avec activation relu.

Chemin d'expansion :

- Relativement symétrique au chemin de sous-échantillonnage

- 4 blocs de *conv* (2 couches de *conv* chacun) suivis de couche convolutive transposée (up \ *conv*)
- Le nombre de filtres de chaque bloc *conv* consécutif est égale à la moitié des filtres du bloc *conv* précédent.
- Réduire la résolution, augmenter la profondeur (nombre de couches).
- Pas de padding.
- Filtres 3X3 avec activation relu.
- Concaténation de la carte d'entités à partir des couches de sous-échantillonnage correspondantes pour une localisation plus précise.
- La couche finale est une *conv* 1x1 utilisée pour mapper chaque vecteur d'entités de 64 composants au nombre souhaité de classes (dans ce cas, cellule et arrière-plan).

3. Entraînement

Les images fournies en entrée et leurs caractéristiques correspondantes sont utilisées pour entraîner le réseau avec le gradient stochastique.

Pour la fonction de perte, ils l'ont basée sur le soft max par pixel sur la carte de caractéristiques finale combinée avec la fonction de perte d'entropie croisée, ce qu'ils ont fait ici, c'est introduire la carte de poids pré-entraînée qui a été utilisée pendant l'entraînement,

Cette carte de poids précalculée est obtenue pour chaque vérité terrain de segmentation afin de compenser la fréquence des pixels de certaines classes, parce que l'apparence des pixels de différentes classes était déséquilibrée car il y avait plus de pixels blancs représentant les cellules réelles et les pixels d'arrière-plan ou de bordure et ils ont donc en quelque sorte calculé cette carte de poids précalculée pour chaque segmentation réelle et l'ont utilisée dans la fonction de perte finale pour non seulement forcer le réseau à s'améliorer mais détecter à la fois les classes de bordure et les cellules mais aussi pour améliorer la reconnaissance de la bordure les frontières entre les cellules, ils se sont vraiment attachés à forcer le réseau à détecter ces frontières le plus précisément possible. Ils ont donc introduit ce type d'approche où ils ont pénalisé le réseau pour ne pas détecter ou pour fusionner des cellules et en quelque sorte ignorant la frontière et donc la frontière recevait généralement des poids plus élevés pour la rendre plus importante.

4. Data augmentation

Concernant l'augmentation des données, ils ont mentionné que le plus grand succès qu'ils ont eu était d'utiliser la génération de déformations en utilisant des facteurs de déplacement aléatoires et en appliquant une fenêtre de taille 3*3 et en déplaçant les pixels en les décalant et ainsi de suite et c'était le plus réussi et le plus influent pour la l'entraînement.

5. Conclusion

L'architecture U-Net réalise de très bonnes performances sur de divers problèmes de segmentation biomédicale. Le réseau ne nécessite que peu de d'images d'entraînement et ce grâce à l'augmentation des données.

- Lien vers l'article : <https://arxiv.org/pdf/1505.04597v1.pdf>