Intervalles de confiance pour l'estimation de superficies à partir d'images satellitaires

Benjamin LAMBERT[1,2] Florence FORBES[3] Senan DOYLE[2] Alan TUCHOLKA[2] Michel DOJAT[2]

[1]Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000, FR

[2]Pixyl, Laboratoire de Recherche et Développement, 38000 Grenoble, FR

[3]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, FR

Résumé – La segmentation des images satellites permet d'extraire des métriques de haut-niveau telle que la superficie d'une région d'intérêt. Cependant, ces estimations sont rarement équipées d'intervalles de confiance, ce qui met en cause la fiabilité des algorithmes. Dans cette étude, nous proposons un nouveau modèle de segmentation, TriadNet, qui permet d'obtenir la segmentation de l'image associée à un intervalle de confiance concernant les superficies, en moins de $0.05s$ par image. L'intérêt de notre approche est démontré sur deux tâches distinctes : la segmentation de bâtiments et de routes, à partir d'images satellite.

Abstract – Segmentation of satellite images allows the obtention of high-level metrics such as the area of a region of interest. However, these estimations are rarely equiped with confidence intervals, which undermine the reliability of the algorithms. In this study, we propose a new segmentation model, TriadNet, which allows to obtain the segmentation of the image associated with a confidence interval on estimated areas, in less than 0.05s per image. The interest of our approach is demonstrated in two distinct tasks: the segmentatcion of buildings and roads, from satellite images.

1 Introduction

L'imagerie satellitaire permet de couvrir de larges portions de la surface terrestre et offre une aide précieuse pour la surveillance de la déforestation [10], des feux de forêt [10] ou encore des catastrophes naturelles [3]. Ces données de grande dimensionnalité peuvent être analysées avec des outils Deep Learning (DL), permettant de segmenter automatiquement les images en différentes régions d'intérêt (ROI) [13]. À partir de la segmentation, une mesure de la superficie de la ROI peut facilement être obtenue en multipliant le nombre de pixels segmentés par la résolution du pixel. Cependant, cette approche ne permet pas d'obtenir directement un intervalle de confiance (IC) associé à cette mesure de superficie, ce qui peut mettre en cause la fiabilité de la prédiction.

Deux démarches méthodologiques ont été proposées afin d'obtenir des intervalles de confiance à partir d'un modèle DL : l'approche échantillonnage [6, 7], et l'approche régression [14, 16]. Dans l'approche échantillonnage, divers masques de segmentation plausibles sont générés pour la même image d'entrée, à partir desquels une distribution sur la quantité d'intérêt est calculée (e.g superficie ou volume), dont la moyenne et l'écart-type peuvent être extraits pour définir un IC. Pour générer ces masques variés, des méthodes telles que le Monte Carlo Dropout [9] ou encore l'Augmentation de Données [17] peuvent être employées. L'inconvénient est que le temps d'inférence est sensiblement augmenté, car plusieurs dizaines de prédictions pour la même image doivent être générées. Dans l'approche régression, un réseau de neurones auxiliaire est entraîné à prédire les 3 éléments de l'IC (médiane, borne inférieure et borne supérieure) à partir de l'image et/ou de sa segmentation. Cependant, cela nécessite d'entraîner un modèle auxiliaire de régression en plus du modèle de segmentation.

Pour résoudre les limitations respectives de ces deux approches, nous proposons un nouveau modèle de segmentation baptisé TriadNet. TriadNet est un modèle de segmentation des images satellite, qui fournit également des intervalles de confiance associés à l'estimation de superficie, et cela sans besoin d'échantillonnage. Nous validons notre méthode sur deux tâches : estimation de la surface de routes, et de bâtiments, à partir d'images satellites.

2 Définition mathématique

Nous considérons un problème de segmentation binaire d'images 2D. L'objectif est d'estimer la superficie Y de la région d'intérêt, à partir de la segmentation de l'image. Pour une estimation donnée de la superficie X, considérée comme une variable aléatoire, nous définissons un IC $\Gamma_\alpha(X)$ comme un intervalle de valeurs conditionné pour contenir Y, la vraie superficie, avec un certain degré de confiance $1-\alpha$ (e.g. 90%, 95%). C'est-à-dire, compte tenu d'une série d'estimations de superficie $X_1 \dots X_n$, et des vraies valeurs $Y_1 \dots Y_n$ qui leur sont associées, $\Gamma_\alpha(\cdot)$ doit être construit de manière à satisfaire :

$$P(Y_{\text{test}} \in \Gamma_\alpha(X_{\text{test}})) \geq 1-\alpha \tag{1}$$

pour toute paire $(Y_{\text{test}}, X_{\text{test}})$ suivant la même distribution que les (Y_i, X_i). Cette propriété est appelée la *couverture marginale*, comme la probabilité est marginale sur l'entièreté de la base de test [1].

Les méthodes d'estimation d'IC basées sur l'échantillonnage se basent sur l'hypothèse que X suit une loi normale. Sous cette hypothèse, la valeur médiane μ_X ainsi que la déviation standard σ_X de la distribution sont obtenus en échantillonnant différentes prédictions pour la même image d'entrée. Puis, l'IC est obtenue comme étant $\Gamma_\alpha(X) = [\mu_X - z\sigma_X, \mu_X + z\sigma_X]$, où z correspondant au nombre de déviation standard,

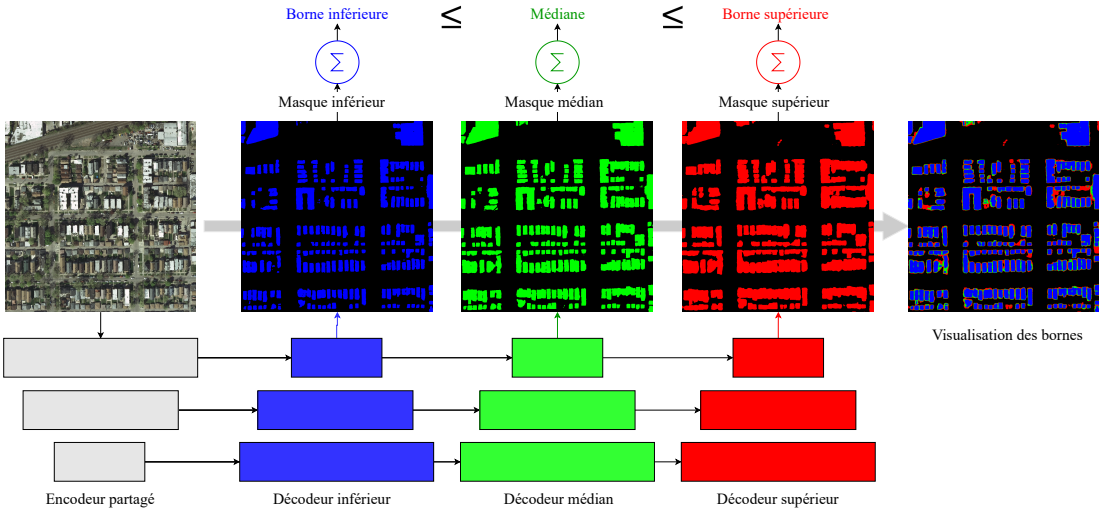

FIGURE 1 : Illustration du TriadNet sur un exemple de segmentation de bâtiment. Chacun des trois décodeurs fournit un masque respectif, permettant d'obtenir respectivement la borne inférieure, la médiane, et la borne supérieur de l'IC.

stipulant le degré de confiance souhaité de l'intervalle. Par exemple, pour un IC à 90% de confiance, z correspond à 1.65. En opposition, les estimateurs directs d'IC (comme notre solution proposée, TriadNet), prédisent directement la valeur médiane μ, ainsi que les bornes inférieures l_b et supérieures u_b ($l_b \leq \mu \leq u_b$), sans échantillonnage.

3 Solution proposée

TriadNet est basé sur une architecture de segmentation du type U-Net [8], composée initialement d'un encodeur suivi d'un décodeur. Nous modifions cette architecture en démultipliant les décodeurs, de manière à en obtenir 3, chacun relié au même encodeur (voir Figure 1). Pour une image donnée, TriadNet fournit trois masques de segmentation différents, utilisés pour extraire un IC : le masque inférieur, le masque médian, et le masque supérieur, construits de manière à ce que la superficie S de chaque masque soit croissante : $S_{\text{inférieur}} \leq S_{\text{médiane}} \leq S_{\text{supérieur}}$. Pour obtenir ce résultat, le modèle est entraîné avec une nouvelle fonction de coût, que nous nommons *fonction de coût triadique* (FCT).

3.1 Fonction de coût triadique

La fonction d'entraînement de TriadNet part du constat que le masque inférieur doit être plus restrictif (*i.e.* plus grande précision et plus faible rappel) que le masque médian. De manière similaire, le masque supérieur doit être plus permissif (*i.e.* plus grand rappel et plus faible précision). Pour garantir cela, nous proposons de nous baser sur la fonction de coût Tversky [15], qui permet un contrôle direct sur le compromis entre précision et rappel. La fonction de coût Tversky $T_{\alpha,\beta}$ est une extension de la populaire fonction Dice [12], très couramment utilisée en segmentation, avec 2 hyper-paramètres additionnels α et β qui contrôlent respectivement le poids des faux positifs (FP) et faux négatifs (FN) dans la fonction de coût. À noter qu'avec $\alpha = \beta = 0.5$, la fonction de coût est strictement équivalente à la fonction Dice.

En écrivant p_{infr}, p_{med} et p_{sup} les sorties de chaque décodeur, et y le masque de segmentation de la vérité-terrain, la FCT est définie comme :

$$\text{FCT} = T_{1-\gamma,\gamma}(p_{inf,y}) + T_{0.5,0.5}(p_{med,y}) + T_{\gamma,1-\gamma}(p_{sup,y}) \quad (2)$$

avec γ un hyper-paramètre fixé à 0.3 dans nos fexpériences. En d'autres termes, le décodeur médian est entraîné avec la fonction Dice standard. Pour obtenir un masque plus restrictif (et donc une superficie plus faible), le décodeur inférieur est entraîné à minimiser les FP au détriment d'un plus grand nombre de FN. De manière similaire, pour obtenir des masques plus permissifs (et des superficies plus importantes), le décodeur supérieur est entraîné à minimiser les FN au prix d'un plus grand nombre de FP.

4 Matériel et Méthodes

4.1 Bases de données

Nous démontrons l'intérêt de notre solution pour deux problèmes de segmentation binaire d'images satellites. Tout d'abord, nous étudions la segmentation de bâtiments à l'aide de la base Inria Aerial Image Labeling [11] [1]. La partie en libre accès de cette base comporte 180 images satellites faisant 5000×5000 pixels, acquis avec une résolution de $0.3m$, que nous avons sous-divisées en patchs de 1024×1024 pixels. Pour augmenter le nombre de patchs en vue de l'apprentissage, nous avons utilisé un recouvrement de 256 pixels entre les patchs. Pour les patchs de test, le recouvrement est de 0. Ce processus permet d'aboutir à 4176 patchs pour l'apprentissage, 504 pour la validation, et 800 pour le test. La seconde base de données, Deep Globe Road Extraction [5] [2], se concentre sur la segmentation de routes. Les données en libre accès représentent 6226 images satellites de taille 1024×1024. Nous les avons réparties en 4657 images pour l'apprentissage, 569 pour la validation, et 1000 pour le test.

[1] https ://project.inria.fr/aerialimagelabeling/

[2] https ://www.kaggle.com/datasets/balraj98/deepglobe-road-extraction-dataset

4.2 Méthodes comparées

Nous avons comparé TriadNet avec deux autres méthodes basées sur l'échantillonnage : le MC Dropout et l'Augmentation de données.

Monte Carlo Dropout est une technique populaire de quantification de l'incertitude d'un modèle DL [9]. Elle est basée sur l'utilisation du Dropout, qui consiste à éteindre une partie des paramètres du modèle pour éviter le sur-apprentissage. Pour implémenter cette technique, le Dropout est gardé actif au moment de l'inférence. En conséquence, plusieurs passages de la même image dans le réseau mènent à des prédictions différentes (et donc des estimations de superficies différentes), car les paramètres éteints changent de manière aléatoire à chaque itération.

Augmentation de données (AD) consiste à créer des versions alternatives d'une image en appliquant des transformations telles que les rotations ou encore les modifications de contraste. Chaque instance est ensuite analysée par le modèle de segmentation, ce qui mène à une estimation de la superficie. En répétant le processus, une distribution sur la superficie est obtenue, à partir de laquelle l'IC peut être dérivé.

4.3 Calibration des intervalles de confiance

En pratique, les intervalles de confiance sont souvent calibrés à l'aide d'une base de calibration [1] (dans notre cas, les images de validation), de manière à s'assurer que la couverture marginale désirée est respectée. Cette étape a pour but de trouver une valeur corrective q, telle que les IC calibrés respectent la couverture désirée de $(1-\alpha)\%$ sur la base de calibration. Dans le cas des IC estimés par échantillonnage, q prend la forme d'un facteur multiplicatif appliqué sur la déviation standard (Équation 3). Dans le cas des IC estimés directement (comme pour TriadNet), q correspond à un facteur additif appliqué sur les bornes inférieures et supérieures (Équation 4) :

$$\Gamma_{\alpha,\text{cal}}(X) = [\mu_X - q\sigma_X, \mu_X + q\sigma_X] \quad (3)$$

$$\Gamma_{\alpha,\text{cal}}(X) = [l_b - q, u_b + q] \quad (4)$$

4.4 Évaluation

Nous menons nos expériences avec $\alpha = 0.1$, signifiant que nous nous concentrons sur des IC à 90%. Pour évaluer la performance de segmentation, nous calculons le score Intersection sur Union (IoU) entre le masque prédit et la vérité terrain (pour TriadNet, le score IoU est calculé à partir du masque *médian*). Nous reportons également l'erreur absolue moyenne (AEM) entre la superficie estimée, et la vraie superficie. Des IC optimaux doivent respecter deux conditions. Tout d'abord, ils doivent atteindre la *couverture marginale* désirée de 90%. Secondement, les intervalles doivent être les plus étroits possible, de manière à être informatifs. Pour vérifier cela, nous calculons deux métriques sur les IC : la couverture empirique (f) ainsi que la taille moyenne des intervalles (W) [7]. f correspond à la proportion empirique de superficies vérité-terrain qui sont contenues dans les ICs prédits. W correspond à la distance moyenne entre les bornes supérieures et inférieures. Pour terminer, nous reportons également le temps moyen de prédiction (T), correspondant au temps nécessaire pour générer la segmentation d'une image et pour dériver les intervalles prédictifs.

4.5 Détails d'implémentation

Toutes nos expériences sont implémentées avec PyTorch, en utilisant un GPU RTX A5000 de Nvidia. Trois types de modèles de segmentation sont implémentés pour chacune des deux bases de données. Tout d'abord, le modèle *Base* correspond à un U-Net standard, implémenté avec la librairie MONAI [4], qui permet ensuite d'obtenir les IC avec la technique AD. Ensuite, pour implémenter la méthode MC, un second U-Net *Dropout* est entrainé avec un taux de Dropout de 10% dans chacune des couches de l'encodeur et du décodeur. Le dernier type de modèle correspond à notre TriadNet. Tous les modèles sont entrainés avec l'algorithme ADAM, avec un taux d'apprentissage de $1e-4$, utilisant la fonction de coût Dice pour les modèles *Base* et *Dropout*, et la FCT pour TriadNet. Pendant l'apprentissage, les modèles analysent les images 4 par 4. Nous utilisons de l'augmentation de données modérée pendant l'apprentissage, consistant en des rotations et des altérations de contraste, implémenté avec la librairie Albumentation [2].

Pour la méthode AD, 20 versions altérées de l'image d'entrée sont obtenues en application des rotations et des modifications de contraste, et le modèle *Base* est utilisé pour obtenir les prédictions. Pour la méthode MC Dropout, 20 inférences sont effectuées pour chaque image d'entrée avec le modèle *Dropout*.

TABLE 1 : Tableau récapitulatif de performances

	IoU	AEM (m^2)	f ($\%$)	W (m^2)	T (s)
Bâtiment					
MC	.67	1065	87.4	4290	0.51
AD	.66	**1008**	89.6	4904	1.63
Triad	**.70**	1100	**89.8**	**4083**	**0.05**
Route					
MC	.62	2002	88.5	8583	0.52
AD	.62	2044	**90.5**	10924	1.56
Triad	.62	**1670**	88.9	**6649**	**0.05**

5 Résultats et conclusion

Le tableau 1 présente les performances de segmentation (IoU), d'estimation de superficie (AEM) ainsi que la qualité des intervalles de confiance (f et W), pour les deux bases de données testées. La Figure 2 propose une représentation visuelle des IC calculés sur la tâche de segmentation de routes. Il apparaît que notre approche, TriadNet, produit des IC qui sont les plus étroits, tout en approximant correctement la *couverture marginale* désirée de 90% (f atteint 89.8% sur la segmentation des bâtiments et 88.9% pour les routes). Par ailleurs, notre méthode se démarque nettement par sa rapidité, étant 10 fois plus rapide que MC Dropout, et environ 30 fois plus rapide que AD. Par ailleurs, les modifications d'architecture et de fonction de coût nécessaires pour entrainer TriadNet n'ont pas de répercussion négative sur la performance de segmentation, TriadNet surpassant même les autres modèles sur le dataset