

Apprentissage par transfert entre images aériennes actuelles et historiques par un modèle génératif adapté à la segmentation sémantique du couvert arboré

Thierry GAUBERT^{1,2} Valentine BELLET¹ Yousra HAMROUNI¹ Thierry FEUILLET² Déborah BIRRE³ David SHEEREN¹

¹Université de Toulouse, INRAE, UMR DYNAFOR, 31320 Castanet-Tolosan, France

²Université de Caen Normandie, CNRS – UMR 6266 IDEES, Caen, France

³FRB-CESAB, 34000 Montpellier, France

Résumé – Les problèmes d’apprentissage supervisé nécessitent des données annotées en quantité importante. L’adaptation de domaine permet de transformer les données d’entraînement d’un problème analogue en des données utilisables pour un problème ciblé et pallier le manque éventuel de données annotées. Cette étude propose de réaliser l’adaptation de domaine entre images aériennes actuelles acquises dans les années 2010-2020 et des images aériennes historiques produites autour des années 1950, afin d’entraîner un algorithme de segmentation sémantique pour cartographier le couvert arboré passé en France métropolitaine. L’objectif est de réaliser l’adaptation de domaine avec un modèle génératif de type CycleGAN adapté à la tâche de segmentation sémantique, et d’évaluer son apport sur les performances de segmentation des images historiques par rapport à d’autres méthodes d’adaptation de domaine. Les deux tâches d’adaptation et de segmentation sémantique sont couplées afin d’obtenir des données appropriées au problème ciblé de la détection du couvert arboré. Les résultats obtenus mettent en évidence deux apports du couplage des tâches. Premièrement, le modèle génératif converge plus rapidement vers la production d’images vraisemblables. Deuxièmement, les résultats de segmentation sont plus pertinents lorsque l’apprentissage des deux tâches sont couplées en comparaison avec l’apprentissage successif. Toutefois, les performances de segmentation actuelles sont bien en deçà des performances attendues en utilisant directement des images du domaine cible pour l’apprentissage de la tâche de segmentation.

Abstract – Supervised learning problems require large amounts of annotated data. Through domain adaptation, training data from an analogous problem can be transformed into data usable for a targeted problem, and to compensate for any lack of annotated data. This study proposes to perform domain adaptation between recent aerial images acquired in the years 2010-2020 and historical aerial images produced around the 1950s, in order to train a semantic segmentation algorithm to map past tree cover in mainland France. The aim is to perform domain adaptation using a CycleGAN generative model adapted to the semantic segmentation task, and to evaluate its contribution to historical image segmentation performance compared with other domain adaptation methods. The two tasks of adaptation and semantic segmentation are coupled in order to obtain data appropriate to the targeted problem of tree cover detection. The results highlight two benefits of task coupling. Firstly, the generative model converges more quickly towards the production of plausible images. Secondly, the segmentation results are more relevant when the two tasks are coupled, compared with successive learning. However, current segmentation performance is well below that expected when using images from the target domain directly to train the segmentation task.

1 Introduction

Les performances atteignables par les algorithmes de segmentation sémantique dépendent principalement de la disponibilité et de la qualité de larges bases de données, comprenant des images annotées. L’adaptation de domaine vise à contourner le manque d’annotations des images répondant à un problème ciblé, en utilisant des images annotées provenant d’une autre source et avec des propriétés analogues. L’enjeu est ensuite d’effectuer une transformation des images annotées pour qu’elles suivent au mieux la distribution des images non-annotées et réaliser l’apprentissage du problème de segmentation ciblé. Des transformations simples d’augmentation de données peuvent potentiellement pallier le manque d’images annotées mais ne garantissent pas de reproduire les propriétés radiométriques, texturales et contextuelles adéquates. Des méthodes de conversion plus avancées basées sur des techniques d’adaptation de domaine telles que le transport optimal ou les modèles généra-

tifs profonds peuvent être utilisés pour assurer un transfert de style cohérent [8]. En particulier, les modèles génératifs basés sur le principe des réseaux génératifs adverses (GAN) tels que Pix2Pix [7] ou l’architecture CycleGAN [12] ont permis de réaliser le transfert de styles sur des images de nombreux domaines. Plus spécifiquement, dans le domaine de la télédétection, l’architecture CycleGAN a été employée avec succès dans divers contextes, incluant la transformation d’images dans le domaine visible à des images dans le proche infrarouge [2] et la transformation d’images aériennes récentes en images historiques [9].

L’adaptation de domaine à l’aide d’un CycleGAN est une approche inspirante mais l’adaptation n’est pas dans ce cas optimisée pour la tâche ciblée de segmentation, les tâches y sont découplées. Néanmoins, des travaux antérieurs ont montré que le couplage des tâches de génération et de classification donne des résultats plus précis que les tâches découplées [11]. Dans le domaine de la télédétection, le couplage de l’adaptation

via un CycleGAN avec une tâche de classification a déjà été proposé et semble montrer une amélioration des résultats sur le problème ciblé [10]. Cependant, le couplage avec une tâche de segmentation sémantique n'a pas encore été évalué à notre connaissance.

Dans les dernières années, l'accès aux images aériennes de grandes campagnes historiques a été rendu possible par la numérisation des images argentiques. En France, les images acquises autour des années 1950 ont été traitées pour fournir le produit BD ORTHO® Historique, librement accessible, couvrant l'ensemble de la France métropolitaines et regroupant plus de trois millions de clichés [1]. Les images aériennes historiques sont une source d'information cruciale pour remonter à l'occupation du sol dans le passé. En particulier, elles peuvent permettre la cartographie du couvert arboré passé, essentielle pour analyser la dynamique de la végétation et évaluer l'état de conservation des habitats.

Les images aériennes historiques, présentent une grande variabilité liée aux conditions d'acquisition, au capteur utilisé, à l'état de conservation et au paramètre de numérisation. Cette variabilité est un défi pour effectuer leur analyse automatique. Cependant leur résolution spatiale assez fine, en dessous du mètre, permettant d'extraire des caractéristiques pertinentes pour effectuer leur segmentation sémantique. En effet, les méthodes automatiques d'apprentissage profond développées pour la segmentation sémantique s'appuyant sur des architectures de réseaux convolutifs (Fully Convolutional Networks, FCN) avec une structure de type encodeur-décodeur telle que l'architecture U-Net, ont été utilisées avec succès sur des images aériennes historiques [4].

Si les annotations d'images aériennes pour la couverture du sol sont assez répandues pour des images récentes, elles sont rares pour les images aériennes historiques, en particulier avec des annotations spécifiques à la détection du couvert arboré. Le manque d'images historiques annotées peut être contourné en s'appuyant sur des images aériennes récentes et annotées et en réalisant une adaptation de domaine pour transformer ces dernières en de fausses images historiques.

Par ailleurs, les changements du paysage et de la couverture du sol entre les deux périodes rendent impossible la constitution de paires d'images pour apprendre le transfert de style entre images actuelles et historiques. L'architecture CycleGAN qui réalise la transformation d'image à image non appariée (unpaired image-to-image translation) [12] est donc pertinente pour cette adaptation de domaine. Un CycleGAN a ainsi permis à Van den Broeck *et al.* [9] de transformer des images aériennes récentes en fausses images historiques afin de pouvoir entraîner un algorithme de segmentation sémantique en utilisant des annotations des images récentes.

Ainsi, cette étude propose d'évaluer l'apport d'une contrainte liée à la segmentation sémantique dans l'architecture de type CycleGAN [12] pour réaliser la transformation des images actuelles et entraîner un algorithme de segmentation sémantique. En particulier, la tâche de segmentation sémantique sera appliquée à la détection du couvert arboré dans les images aériennes historiques. L'objectif, en utilisant cette architecture, est d'obtenir des images transformées en fausses images historiques qui non seulement sont visuellement cohérentes, mais également interprétables et adaptées à un modèle de segmentation sémantique pour répondre à l'objectif final.

2 Méthodologie

Dans cette étude, nous proposons une version modifiée de l'architecture CycleGAN pour réaliser le transfert de style des images aériennes actuelles vers des images historiques. L'architecture CycleGAN, initialement présentée par Zhu *et al.* [12], permet de réaliser une transformation bidirectionnelle entre deux domaines distincts : un domaine source noté X et un domaine cible noté Y . L'architecture repose sur une paire de GAN. Un premier GAN effectue la transformation, notée G , du domaine source X vers le domaine cible Y , le second GAN effectue la transformation, notée F , du domaine Y vers le domaine X . Lors de l'entraînement, la fonction de coût est calculée d'une part à travers la capacité des deux réseaux discriminateurs, et d'autre part en évaluant la reconstruction d'une image lors d'un cycle complet. Les cycles de reconstruction permettent de garantir la préservation des caractéristiques essentielles des données et d'obtenir une correspondance cohérente entre les deux domaines.

Nous proposons une implémentation d'une architecture CycleGAN adaptée à la segmentation sémantique, que nous nommons SegCycleGAN. L'architecture se base sur un CycleGAN comme décrit précédemment, auquel sont ajoutés deux modules de segmentation sémantique, S_X et S_Y , pour contraindre le modèle et tel qu'illustré en Figure 1. Les deux modules apprennent à partir des annotations du domaine source. Le module du domaine cible S_Y est entraîné à partir des fausses images générées par G . Le module du domaine source S_X est entraîné à partir des images reconstruites, générées par F . La fonction coût issue des modules de segmentations, calculée par une entropie croisée binaire, est ajoutée aux autres fonctions coût.

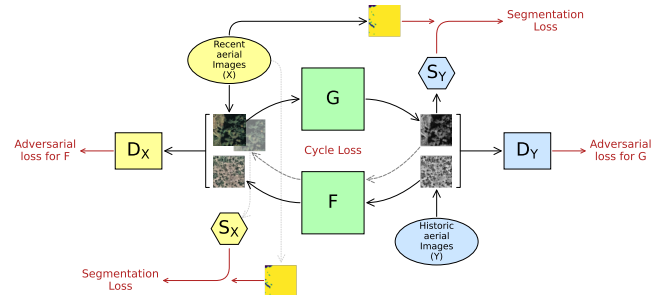


FIGURE 1 : Architecture CycleGAN adaptée à la segmentation sémantique : X désigne le domaine des images récentes et Y celui des images historiques. $G : X \rightarrow Y$ et $F : Y \rightarrow X$ sont les transformations entre ces domaines. D_X et D_Y sont les discriminateurs des domaines X et Y , tandis que S_X et S_Y traduisent la segmentation sémantique dans X et Y .

A partir du produit BD ORTHO® Historique [1], qui fournit des images à une résolution de 0.5 m, 5000 imageries carrées de 512 par 512 pixels ont été extraites pour permettre l'entraînement du SegCycleGAN. Les positions des 5000 imageries ont été sélectionnées aléatoirement pour couvrir tout le territoire français et tenir compte de la diversité des paysages, des capteurs et des conditions d'acquisition. De la même manière, 240 imageries ont été extraites et annotées manuellement sur l'ensemble du territoire pour valider les performances de segmentation sémantique. Nous nommons par la suite ce jeu de

données HistoTrees. De plus, deux autres jeux de données comprenant des images historiques annotées serviront pour l'évaluation de la segmentation sémantique. Le premier jeu de données, nommé SpatialTreeP, comprend des images acquises dans la moitié est des Pyrénées et représente une surface totale de 19 km² [3]. Le deuxième, nommé Pygar, issu de clichés aériens de 1942, couvre une surface de près de 100 km² dans le sud-ouest de la région toulousaine [4].

Les images actuelles utilisées dans cette étude proviennent du jeu de données FLAIR [5]. FLAIR comprend des images aériennes manuellement annotées, couvrant environ 810 km² réparties en 950 zones sur la France métropolitaine. Les images FLAIR sont initialement à une résolution spatiale de 0.2 m. Afin d'aligner les données FLAIR sur les caractéristiques de la BD ORTHO® Historique, un rééchantillonnage bilinéaire a été appliqué pour obtenir des images à une résolution de 0.5m, de même un rééchantillonnage au plus proche voisin est appliqué aux annotations sous la forme de masques. Les images ont ensuite été divisées pour obtenir près de 13000 images carrées de 512 pixels par 512 pixels.

Concernant la segmentation sémantique, l'utilisation de l'architecture SegCycleGAN est évaluée en deux parties. D'une part, nous évaluons les performances de notre modèle en le comparant avec les performances obtenues avec un découplage des tâches, c'est-à-dire un apprentissage d'une architecture CycleGAN suivi d'un apprentissage d'un module de segmentation. D'autre part, nous définissons des scores de références : (1) une référence basse en entraînant le modèle sur le jeu de données FLAIR converti en niveaux de gris, (2) une référence dite d'adaptation par augmentation de données (AAD) s'inspirant du travail de Gui *et al.* [6], consistant à entraîner le modèle sur le jeu de données FLAIR converti en niveaux de gris et en appliquant des transformations simples pour imiter les images historiques, (3) une référence haute en entraînant le modèle de segmentation directement avec des images historiques annotées (jeux de données HistoTrees, SpatialTreeP et Pygar). Pour la référence haute, les scores sont évalués séparément sur chaque jeu de données, à l'aide d'une stratégie de validation croisée Monte-Carlo avec 10 répétitions. Les scores de segmentation sémantique sont tous calculés à l'aide de l'indice de Jaccard (IoU).

L'implémentation des modèles est réalisée avec pytorch. Les générateurs sont des U-Net avec 8 blocs de sous-échantillonnage, 7 blocs de sur-échantillonnage et une couche finale de convolution avec activation tanh. Les réseaux discriminateurs sont issus de l'architecture PatchGAN [7] avec 4 blocs de sous-échantillonnage. Tous les modules de segmentation utilisés sont des réseaux U-Net avec un encodeur ResNet-34 pré-entraîné sur la base de données Imagenet. Les différentes fonctions de coût sont pondérées par les poids suivants : 10, 5 et 1 respectivement pour les fonctions de coût adversaires, de cycle et de segmentation. L'entraînement du SegCycleGAN est effectué sur 100 epochs, avec une taille de batch de 8 et un learning rate de 10^{-6} . Les mêmes paramètres sont utilisés pour l'entraînement de CycleGAN. Pour calculer les scores de références et évaluer l'adaptation de domaine du CycleGAN, les modules de segmentation sont entraînés sur 10 epochs avec une taille de batch de 16 et un learning rate de 5.10^{-4} .

	HistoTrees	SpatialTreeP	Pygar
Référence basse	0.23	0.23	0.49
Référence AAD	0.40	0.27	0.48
Référence haute	0.73	0.69	0.80
CycleGAN	0.17	0.06	0.01
SegCycleGAN Cible (S_Y)	0.33	0.35	0.45
SegCycleGAN Source (S_X)	<u>0.48</u>	<u>0.40</u>	<u>0.54</u>

TABLE 1 : Indices de Jaccard calculés pour chaque jeu de données d'images historiques annotées et pour les différentes expériences considérées

3 Résultats

Dans un premier temps, la vraisemblance des images produites par les modèles génératifs est évaluée de manière qualitative. Les modèles CycleGAN et SegCycleGAN fournissent des images cohérentes et vraisemblables comme l'illustrent les exemples d'images transformées en Figure 2. Comparées aux images générées par le CycleGAN, les images issues du SegCycleGAN paraissent plus réalistes et comportent moins d'artefacts pour le même nombre d'itérations à l'apprentissage.

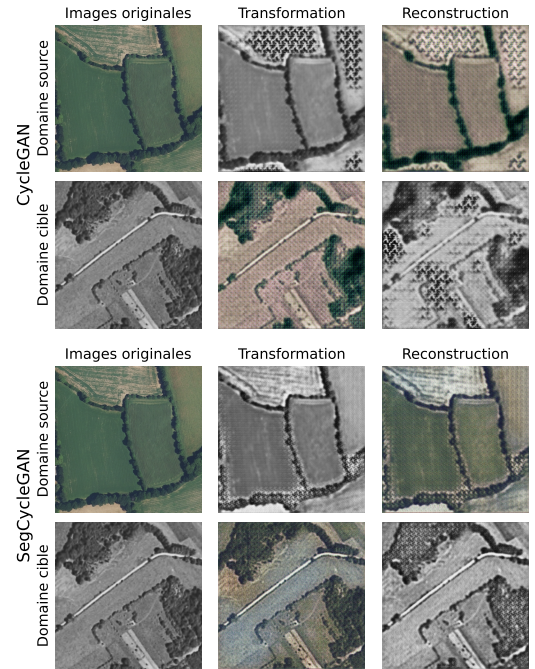


FIGURE 2 : Exemples de transformations issues du modèle CycleGAN et SegCycleGAN pour l'adaptation entre images aériennes actuelles et historiques. Première et troisième lignes, de gauche à droite : une image actuelle, sa transformation en fausse image historique et sa reconstruction. Seconde et quatrième lignes, de gauche à droite : une image historique, sa transformation en fausse image actuelle, et sa reconstruction

Les indices de Jaccard calculés selon les différentes expériences considérées sont regroupés dans la Table 1.

Les segmentations apprises à partir des images générées par le CycleGAN obtiennent des scores faibles. Ces scores montrent que les fausses images historiques, bien que visuellement réalistes, ne correspondent pas suffisamment à la distribution des vraies images pour effectuer l'apprentissage de la segmentation. Les scores de segmentation obtenus avec le

SegCycleGAN semblent indiquer que l'ajout de la contrainte de segmentation permet de générer des images qui sont plus facilement interprétables pour le module de segmentation et permettent une meilleure généralisation. Cependant les scores obtenus avec le SegCycleGAN restent comparables aux scores de références AAD et bien en deçà des scores de référence haute. On peut également noter avec le SegCycleGAN une différence de performances entre les prédictions effectuées par le module domaine cible, et celui coté domaine source. Comme l'illustrent les exemples de prédiction en Figure 3, la différence de performance s'explique par une part importante de non détection dans le domaine cible.

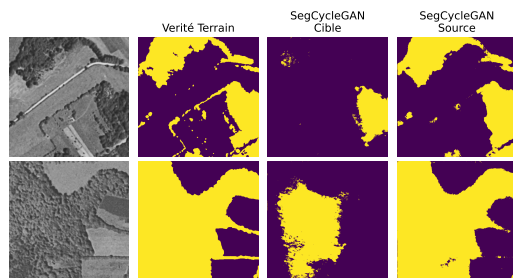


FIGURE 3 : Exemples de prédictions de segmentation pour le SegCycleGAN

4 Conclusion

Le couplages des tâches d'adaptation de domaine et de segmentation par l'ajout de contraintes de segmentation sémantique au CycleGAN (SegCycleGAN) permet de : (i) produire des fausses images historiques plus réalistes (ii) réduire la quantité d'artefacts produits par le générateur et leur prééminence pour le même nombre d'itérations, (iii) améliorer la précision des prédictions de segmentation, et en particulier par rapport à un entraînement avec de fausses images historiques obtenues avec une simple transformation (score de référence avec l'augmentation de données). Cependant, à ce stade, la précision des prédictions de segmentation reste inférieure aux résultats escomptés avec des images historiques annotées (score de référence haut). Les prédictions de segmentation sémantique avec SegCycleGAN sont plus précises dans le domaine source que dans le domaine cible. Nous émettons l'hypothèse que les distributions des images récentes reconstruites et des images historiques colorisées sont similaires et peuvent expliquer cette différence.

5 Remerciements

Ce travail est financé par le projet ANR SpatialTreeP (ANR-21-CE03-0002). Nous tenons à remercier les contributeurs de ce projet ainsi que les équipes des UMR DYNAFOR et IDEES qui ont fourni les données essentielles pour l'entraînement et la validation des modèles. Ces travaux ont bénéficié d'un accès aux ressources de calcul en IA et de stockage au TGCC au travers de l'allocation de ressources 2024-AD011015621 attribuée par GENCI sur la partition V100 du calculateur Joliot Curie.

Références

- [1] BD ORTHO® Historique | Géoservices.
- [2] Bilel BENJDIRA, Yakoub BAZI, Anis KOUBA et Kais OUNI : Unsupervised Domain Adaptation Using Generative Adversarial Networks for Semantic Segmentation of Aerial Images. *Remote Sensing*, 11(11):1369, janvier 2019. Number : 11 Publisher : Multidisciplinary Digital Publishing Institute.
- [3] Déborah BIRRE, Thierry FEUILLET, Roman LAGALIS, Johan MILIAN, Frédéric ALEXANDRE, David SHEEREN, Roberto SERRANO-NOTIVOLI, Matthieu VIGNAL et Maaïke Y. BADER : A new method for quantifying treeline-ecotone change based on multiple spatial pattern dimensions. *Landscape Ecology*, 38(3):779–796, mars 2023.
- [4] Vitória Barbosa FERREIRA, David SHEEREN, Sébastien LEFÈVRE et Stefan LANG : More labels or better labels ? a semantic segmentation study case using historical aerial images for tree delineation. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 9604–9608, 2024.
- [5] Anatol GARIOUD, Nicolas GONTHIER, Loïc LANDRIEU, Apolline De WIT, Marion VALETTE, Marc POUPÉE, Sébastien GIORDANO et Boris WATTRELOS : FLAIR : a Country-Scale Land Cover Semantic Segmentation Dataset From Multi-Source Optical Imagery. In *Advances in Neural Information Processing Systems (NeurIPS) 2023*, 2023.
- [6] Yuanyuan GUI, Wei LI, Xiang-Gen XIA, Birgit EBEN, Christian GINZLER et Zuyuan WANG : Mapping countrywide historical tree cover using semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–1, 2025. Conference Name : IEEE Transactions on Geoscience and Remote Sensing.
- [7] Phillip ISOLA, Jun-Yan ZHU, Tinghui ZHOU et Alexei A. EFROS : Image-To-Image Translation With Conditional Adversarial Networks. pages 1125–1134, 2017.
- [8] Devis TUIA, Claudio PERSELLO et Lorenzo BRUZZONE : Domain Adaptation for the Classification of Remote Sensing Data : An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, juin 2016. Conference Name : IEEE Geoscience and Remote Sensing Magazine.
- [9] Wouter A. J. Van den BROECK, Toon GOEDEME et Maarten LOOPMANS : Multiclass Land Cover Mapping from Historical Orthophotos Using Domain Adaptation and Spatio-Temporal Transfer Learning. *Remote Sensing*, 14(23):5911, janvier 2022. Number : 23 Publisher : Multidisciplinary Digital Publishing Institute.
- [10] Claire VOREITER, Jean-Christophe BURNEL, Pierre LASALLE, Marc SPIGAI, Romain HUGUES et Nicolas COURT : A Cycle Gan Approach for Heterogeneous Domain Adaptation in Land Use Classification. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 1961–1964, septembre 2020. ISSN : 2153-7003.
- [11] Xiaofeng XU, Xianglin BAO, Xingyu LU, Ruiheng ZHANG, Xinquan CHEN et Guifu LU : An end-to-end deep generative approach with meta-learning optimization for zero-shot object classification. *Information Processing & Management*, 60(2): 103233, mars 2023.
- [12] Jun-Yan ZHU, Taesung PARK, Phillip ISOLA et Alexei A. EFROS : Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. pages 2223–2232, 2017.