ONERA
THE FRENCH AEROSPACE LAB

www.onera.fr

PROPOSITION DE SUJET DE THÈSE

Intitulé : Neural networks for multimodal (aerial / streetview / text) geospatial analysis

Référence : TIS-**DTIS-2019-xxx**
(à rappeler dans toute correspondance)

Laboratoire d'accueil à l'ONERA :

Domaine : Traitement de l'Information et Systèmes — Lieu (centre ONERA) : Palaiseau

Département : Traitement de l'Information et Systèmes

Unité : Image Vision Apprentissage — Tél. +33 1 80 38 65 73

Responsable ONERA : Bertrand Le Saux — Email :bertrand.le_saux@onera.fr

Directeur de thèse envisagé :

Nom : Bertrand Le Saux

Adresse : ONERA, Université Paris Saclay

Tél. : — Email : bertrand.le_saux@onera.fr

Sujet :

Objective

More and more data are now geo-localized, and this opens a whole new research area at the intersection of remote sensing (aerial or satellite images), computer vision (standard images shot from the ground) and machine learning (text and structured information). Hence, this relationship between heterogeneous data leads to ask questions with many practical applications. For example:

Where was a given streetview image taken ? **Vo et al. 2017]** It will be useful for self localization in autonomous driving, but also to check and disambiguate fake news if the photo is incoherent with its caption…

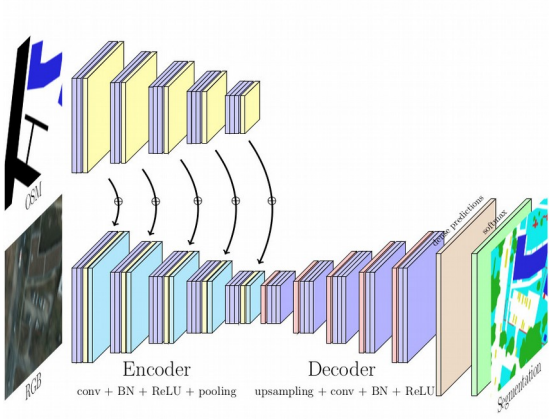
What is visible from the air, given aerial imagery and geo-localized text describing this location (for example from Wikipedia) ? **[Sheehan et al., 2018]** This allows precise land use and landcover classification, much better than standard Earth-observation (EO).

How to describe a place seen from above, or with a few snapshots ? Similar to image or video captioning, "place captioning" is a way to make land classification understandable for humans.

The objective of this research project is to design and develop methods for classification of geolocalized, multimodal aerial / streetview / text data. It rises several scientific problems related to hot topics of deep learning and Earth-observation.

Problem #1: How to build convolutional network models for highly heterogeneous data?

It will build on multimodal convolutional neural networks (CNNs) for semantic segmentation of EO data , developed at ONERA/DTIS **[Audebert et al., 2017][Audebert et al., 2018]**. These networks were designed for various aerial data fusion and aerial imagery and map fusion. They will be expanded to integrate even sparse geolocalized data. Indeed, the adaptability of current neural networks make them particularly relevant for combining multimodal sources simply by designing new net architectures.



Multimodal fusion network for cartography and optical imagery [Audebert et al., 2017]

Problem #2: How to build a representation space suitable for various tasks?

Projecting multiple related data in a single, common, high-dimensional representation space through neural nets creates a rich encoding. However, it has to be carefullly designed to ensure it is optimal and generic. In particular, we will investigate multi-task learning which offers promising solutions to make the models more statistically robust **[Zamir et al., 2018]**.Indeed, optimizing various tasks simultaneously acts as regularization by preventing the model from becoming over-specialized.

Problem #3 : How to solve geospatial problems which benefit to everyone?

Using the previously developed models, a particular care will be given to imagine and design tools able to solve practical applications in geospatial analysis. Some are already well defined, such as land use classification, and we will investigate how more precise solutions can be proposed on specific tasks. Others still have to be conceived and clearly specified, as place captioning, visual geo-localization with a few shots **[Workman et al., 2015][Brahmbhatt et al., 2018]**, ground lay-out prediction and obstacle prediction from geo-localization and aerial imagery **[Zhai et al., 2017],** photo geolocation retrieval **[Wewand et al. 2016]**, fact checking and fake news disambiguation.

Workplan and practical information

The work program will comprise: study of neural networks for multimodal classification and semantic segmentation; coding (python) and experiments with CNNs using open libraries (Pytorch) on large-scale geospatial datasets built at ONERA. Practical sense and a creative mind will be required to imagine new potential applications. This thesis It will take place at the ONERA centre in Palaiseau (near Paris).

References:

[Audebert et al., 2018] Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks *Nicolas Audebert, Bertrand Le Saux, Sébastien Lefèvre*, ISPRS Journal of Photogrammetry and Remote Sensing, 2018

[Audebert et al., 2017] Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps *Nicolas Audebert, Bertrand Le Saux, Sébastien Lefèvre*, CVPR/Earth Vision workshop, Hawaï, USA, July 2017

[Brahmbhatt et al., 2018] MapNet: Geometry-Aware Learning of Maps for Camera Localization,

Samarth Brahmbhatt, Jinwei Gu, Kihwan Kim, James Hays, and Jan Kautz, CVPR, Salt Lake City, June 2018.

[Sheehan et al., 2018] Learning to interpret satellite images using wikipedia, Sheehan, Evan; Uzkent, Burak; Meng, Chenlin; Tang, Zhongyi; Burke, Marshall; Lobell, David; Ermon, Stefano, eprint arXiv:1809.10236.

[Vo et al. 2017] Revisiting IM2GPS in the Deep Learning Era. Nam Vo, Nathan Jacobs, and James Hays.ICCV, Venice, Oct. 2017.

[Wewand et al. 2016] PlaNet - Photo Geolocation with Convolutional Neural Networks, Tobias Weyand, Ilya Kostrikov, James Philbin, ECCV, Amsterdam, Netherlands, 2016

[Workman et al., 2015] Wide-Area Image Geolocalization with Aerial Reference Imagery. Workman S., Souvenir R., Jacobs N., ICCV 2015.

[Zamir et al., 2018] Taskonomy: Disentangling Task Transfer Learning. Zamir, Sax, Shen, Guibas, Malik, Savarese, CVPR, Salt Lake City, USA, June 2018

[Zhai et al., 2017] Predicting Ground-Level Scene Layout from Aerial Imagery. Zhai M., Bessinger Z., Workman S., Jacobs N., CVPR, Hawaï, USA, July 2017.

Collaborations extérieures :

PROFIL DU CANDIDAT

Formation : Ms. Eng. (CS, EE, …), M.Sc. with outstanding results.
French Grandes Écoles, Master 2 recherche learning / computer vision

Spécificités souhaitées : Machine Learning, Deep Learning, Image Processing.
Programming experience (python, etc.)

PARTIE DESTINÉE EXCLUSIVEMENT À LA DSG

ANNEXE À LA PROPOSITION DE SUJET DE THÈSE

Les rubriques suivantes doivent être dûment renseignées :

1. **Titre de la thèse**

   Neural networks for multimodal (aerial / streetview / text) geospatial analysis

2. **Domaine et thématique scientifique principale**

   Perception et Traitement de l'Information / Intelligence Artificielle et Décision

3. **Contexte de l'étude (en 1 à 2 pages)**

   a. à l'ONERA (préciser notamment les personnes participant à l'encadrement en plus du responsable ONERA)

      B. Le Saux

   b. à l'extérieur

   c. bibliographie succincte

      cf. sujet.

4. **Description des travaux (en 1 à 2 pages)**

   a. plan de thèse prévisionnel

      La thèse commencera le mardi 1er octobre 2019. Le sujet de thèse ne pouvant pas être défini de manière exhaustive et définitive (sinon ce ne serait plus un projet de recherche) le octorant approfondira pendant 1 mois sa réflexion sur les motivations et intérêts, le contexte et les domaines d'application, ainsi que les perspectives. Cette réflexion se poursuivra de manière plus espacée jusqu'à T0+4mois.

      En parallèle, dès T0, le doctorant s'attèlera à l"étude de la bibliographie du domaine. Dans un premier temps, elle consistera en une étude générale des papiers du domaine, afin de bien fixer les idées. Elle sera dévelopée avec les outils standards (biblatex ou zotero) et aura vocation à être enrichie et approfondie tout au long de la thèse. Le premier focus concernera les réseaux multimodaux, raster/raster, et les réseaux ayant des sorties géolocalisées, tels IM2GPS.

      Afin de nourrir la réflexion par des expérimentations pratiques, le doctorant codera en parallèle un réseau multimodal image aérienne (BDORTHO) / images streetview (LUCAS) pour la classification sémantique de géolocalisations. Ce travail sera documenté (approches, expériences et analyse des résultats).

      Le premier jalon de la thèse sera à T0+3mois (fin décembre 2019) avec d'une part la rédaction d'un résumé pour la JDD ONERA (qui présentera le sujet, le contexte et les premières expériences) et d'autre part la rédaction d'un article 4 page (le document sus-mentionné mis en forme) à soumettre à la conférence IGARSS 2020 au tout début janvier. À T0+4 mois, le doctorant préparera un poster pour la JDD qui aura lieu début février 2020. Ainsi s'achèvera le premier sprint de la thèse.

      À partir de janvier 2020 (T0+3 mois) et jusqu'à fin mars (T0+6mois), le doctorant cherchera à adjoindre une branche dédiée au texte à son réseau multimodal. Pour ce faire, il constituera une base de textes géolocalisés (issus de wikipedia) correspondant aux zones précédemment identifiées pour enrichir son jeu de données. Ensuite, il codera un réseau pour la classification de texte (type recurrent network) ayant pour but la prédiction de géo-localisation. Ce travail s'appuiera sur l'enrichissement de la bibliographie, dont le deuxième focus sera la classification et le résumé (sumarization) de texte. Enfin il combinera ce réseau textuel simple (noté txt2gps dans la suite) avec le réseau multimodal.

Ce travail sera documenté afin de soumettre un article de 12 pages à la conférence ECCV 2020, dont l'échéance est le 1er avril 2020, et qui constitue le 2e jalon principal de la thèse, qui finit le 2e sprint. (T0+6mois)

La fin de la première année de thèse consistera en l'approfondissement de ce travail pour aboutir à la soumission d'un article revue en télédétection à T0+1an (septembre 2020, 3e jalon principal). Les expériences comprendront des études d'ablation, des focus sur les problèmes rencontrés et des comparatifs des solutions de l'état de l'art ou imaginées à propos. Cette partie plus libre permettra de doctorant de faire preuve de créativité et d'imagination, en développant son autonomie. À T0+1 an, une première boîte à outils logicielle pour la classification géospatiale multimodale sera livrée, par exemple sous forme de dépôt github.

En parallèle, plusieurs jalons secondaires seront à respecter: rédaction d'un article 6 pages reprenant les travaux IGARSS/ECCV en français pour le colloque RFIAP2020 (échéance avril 2020), la présentation de cette communication en juin 2020 (T0+9 mois), et la présentation de la communication IGARSS en juillet 2020 (T0+10 mois).

La deuxième année sera dédiée aux travaux sur l'apprentissage multi-tâche et visera à évaluer les apports de l'apprentissage conjoint. La bibliographie sera enrichie sur ce point (3e focus). Le code sera amélioré avec de nouvelles applications (classification de texte, géolocalisation d'image streetview, etc.). Une première échéance sera T0+15 mois (décembre 2020) avec la rédaction d'un résumé pour la JDD 2021, et la rédaction d'un article 4 page pour la conférence IGARSS 2021. Le matériel de cette communication IGARSS sera constitué des travaux supplémentaires pour l'article revue ou de l'application jugée innovante parmi les premiers travaux de l'année 2. Le doctorant présentera ensuite ses travaux à T0+16 mois lors de la JDD 2021 sous la forme d'un exposé court.

Le quatrième jalon principal (et le 1er de l'année 2) sera à T0+ 17 mois (février 2021) avec la soumission d'un article 12 pages à la conférence ICCV 2021, ou à défaut au workshop Eath Vision associé au congrès CVPR 2012 (échéance mars 2021). Ce travail documentera les travaux pour des tâches multiples, et notamment les expériences avec des réseaux multi-entrée / multi-sorties qui permettent donc l'apprentissage multi-tâche pour la télédétection. La livraison du jeu de données à la communauté sera un des atouts de l'article. Le travail collaboratif avec d'autres doctorants de DTIS/IVA (par exemple Javiera Castillo-Navarro – thèse sur l'apprentissage large-échelle pour l'observation de la Terre) sera également à envisager pour parvenir au niveau d'exigence de ces conférences.

Plusieurs jalons secondaires se succèderont jusqu'à la fin de l'année universitaire; conseil de suivi de thèse de l'université Paris Saclay (à mi parcours, soit T0+18 mois, en mars/avril 2021); Soumission d'un article 6 pages en français à la conférence GRETSI 2021, reprenant les travaux de 2e année, fin mai 2021 (T0+20 mois), qui sera présenté en septembre 2021.

Cependant, le doctorant aura soin de réserver la majeure partie de son temps à l'approfondissement de l'étude des réseaux multi-tâches, qui fera l'objet de la rédaction d'un article revue, dont la soumission devra survenir avant fin 2021.

En parallèle, le doctorant réfléchira à une application attractive (résolvant un problème émergent) qui pourra être entraînée en multi-tâche mais qui sera également intéressante par elle-même. Les expériences afférentes et la programmation du démonstrateur seront documentées en vue d'une soumission à CVPR 2022 (novembre 2021, T0+25 mois, 5e jalon majeur). Cette réflexion se sera nourrie de la poursuite de l'étude biblographique, notamment en ce qui concerne les applications géospatiales.

Le 6e jalon majeur sera à T0+27 mois, fin décembre 2021. Les livrables correspondants seront la soumission du 2e article sus-mentionné, si possible dans une revue de type CVIU ou Pattern Recognition, la soumission d'un article sur une application géospatiale à IGARSS 2022, et la rédaction du résumé pour la JDD 2022, qui aura lieu un mois plus tard, à T0+28 mois, où le doctorant fera une présentation longue de ces travaux de thèse.

La période de janvier à mai 2022 sera consacrée à la rédaction du manuscrit de thèse, et du traitement des affaires courantes (y compris la soumission d'une dernière

communication à la conférence RFIAP 2022, échéance en mars 2022, T0 + 30 mois, et la livraison d'une boîte à outils logcielle pour l'apprentissage multi-tâche, sous forme de dépôt github). Le manuscrit sera envoyé fin mai 2022, soit à T0+32mois.

La période de juin 2022 à septembre 2022 sera consacrée à la préparation de la soutenance et la valorisation des travaux : communications à CVPR et RFIAP (juin 2022), mais également dans diverses réunions informelles (GDR, séminaires, etc.)

La thèse sera soutenue le 30 septembre 2022, soit T0+36 mois, ce qui constitue le dernier jalon majeur.

Au fil de l'eau, diverses actions auront été entreprises : formation (notamment en suivant les cours de Masters voisins, tels le Master MVA, ou bien école d'été – Gretsi, etc. - ou bien formations ONERA).

b. techniques à mettre en œuvre

Apprentissage automatique dont deep learning

Vision par ordinateur, Traitement d'image avancé

Télédétection

+ techniques citées dans le plan de travail ci-dessus.

c. résultats attendus

Publications dans des journaux et des conférences de références en vision par ordinateur, optique et robotique (2 articles revue + 6 articles conférence)

Boîte à outils logiciels (2 minimum)

5. Financement envisagé

Cocher dans la colonne de droite

Type de bourse	
ONERA	☒
DGA	☒
Région	☐
Contrat doctoral	☐
CIFRE	☐
CNES	☐
Autre	☐

Commentaire éventuel :

6. Avis de l'Adjoint Scientifique du Département