

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DOTA-2025-19**
(à rappeler dans toute correspondance)

Lieu : Palaiseau

Département/Dir./Serv. : DOTA/MPSO

Tél. : 01 80 38 63 76

Responsable(s) du stage :

Email. : Selim.ollivier@onera.fr /
sidonie.lefebvre@onera.fr

S. Ollivier ; S. Lefebvre

DESCRIPTION DU STAGE

Thématique(s) : Deep learning, diffusion models, ViT, Synthèse de fonds texturés

Type de stage : Fin d'études bac+5 Master 2 Bac+2 à bac+4 Autres

Intitulé : Synthèse de fonds texturés par modèles de diffusion

Sujet :

Contexte :

L'évaluation des performances des capteurs optiques et SAR nécessite de disposer de scènes de référence et de prendre en compte la variabilité des objets d'intérêt et du fond, donc de pouvoir générer rapidement des images reproduisant les niveaux radiométriques et la texture des fonds naturels réels, que ce soit pour compléter les bases de données expérimentales ou pour remplacer des codes de simulation coûteux pour des évaluations de grandeurs statistiques.

Nos travaux portent plus particulièrement sur les fonds naturels, comme les nuages, pour différentes applications :

- prédiction de couverture nuageuse afin de décider de déclencher ou non une prise de vue, et donc optimiser la planification des constellations de satellites agiles pour l'observation de la terre, mais aussi pour prédire la disponibilité des liens de communications optiques entre les stations sol et les satellites ;
- détection d'anomalies : la synthèse de fond naturels, notamment pour plusieurs longueurs d'onde ou bandes de fréquence, permettra de générer à volonté des images pour entraîner des algorithmes de détection.

Objectifs :

Notre objectif est d'être en mesure de générer une base de données d'images de fonds naturels à partir d'un petit nombre de données mesurées ou issues de codes coûteux, en conservant les mêmes niveaux radiométriques et le même taux de couverture nuageuse pour les nuages, mais avec une répartition spatiale différente. Outre des temps de calculs faibles, le principal point dur est de garantir à la fois la diversité des images générées et leur réalisme physique, notamment au niveau du caractère multi-échelle des textures.

Depuis les travaux de Gatys et al. [Gatys2015], les réseaux de neurones convolutionnels se sont imposés comme l'état de l'art en synthèse de texture par l'exemple. L'idée principale de ces travaux repose sur l'utilisation de réseaux convolutionnels déjà entraînés pour une tâche de reconnaissance d'objets pour extraire des caractéristiques d'une image et générer de nouvelles textures à partir d'un bruit blanc, de façon à minimiser l'écart entre les matrices de Gram dans l'espace des caractéristiques de l'image d'origine et de l'image synthétique.

Dans le cadre de de la thèse de Pierrick Chatillon (2020-2023), nous avons proposé une implémentation d'un ensemble de contraintes statistiques associées à des a priori physiques sur les textures, notamment sur les distributions des histogrammes et sur les pentes spectrales qui, combinées à des réseaux de type Gatys ou StyleGAN [Karras2019], conduisent à des résultats de synthèse très satisfaisants sur des images RGB de nuages [Chatillon2023]. Une extension à des données hyperspectrales et à des cubes 3D est en cours dans le cadre de la thèse de Sélim Ollivier (2023-2026).

Les progrès récents des méthodes de synthèse d'images par modèles de diffusion avec une architecture ViT [Hatamizadeh2024][Peebles2023] font de ces architectures des candidates très prometteuses pour la synthèse de fonds texturés naturels. Les modèles de diffusion apprennent à produire des données réalistes en inversant un processus de diffusion pas à pas. Ce processus commence par un bruit gaussien et, au fil d'une séquence d'étapes, se transforme progressivement pour représenter la distribution de données souhaitée. Ces modèles conduisent à des résultats remarquables dans la synthèse d'images, mais les versions traditionnelles, basées sur des réseaux de type U-net, peuvent conduire à une complexité importante sur des données 3D. Les

architectures Transformers de type ViT, qui permettent une modélisation des dépendances à longue distance au sein des données, améliorent la qualité des images générées par les modèles de diffusion mais aussi les coûts de calculs associés.

Le but du stage est d'adapter les architectures de type diffusion ViT à nos données de fonds naturels texturés. Les principaux points durs sont :

- La prise en compte d'images multispectrales, dans une dizaine de longueur d'onde, ou d'images SAR dans plusieurs bandes de fréquences et de cubes d'images 3D de contenu en eau ou en glace, au lieu des images RGB traditionnellement synthétisées ;
- Le respect de contraintes physiques, comme la propriété de décroissance en puissance du spectre de Fourier ou la couverture fractionnaire pour les nuages ;
- La gestion de la diversité des images générées.

Afin d'évaluer les apports des architectures de type Diffusion ViT mises en œuvre dans le cadre du stage par rapport aux approches de type Gatys et GAN que nous avons développées, nous nous appuyerons sur des bases de données d'images de nuages (issues de simulations ou du satellite Sentinel2) et de textures (bases de données publiques HyTeXiLa ou MacroTextures).

Collaboration : Luc Vignaud - ONERA DEMR

Références bibliographiques :

[Gatys2015] L. Gatys et al, Texture synthesis using convolutional neural networks. In Advances in neural information processing systems, pages 262–270, 2015.

[Karras2019] T. Karras, et al, Analyzing and Improving the Image Quality of StyleGAN, arXiv: 1912.04958, 2019.

[Chatillon2023] P. Chatillon, Y. Gousseau, S. Lefebvre, A geometrically aware auto-encoder for multi-texture synthesis, SSVM 2023, ArXiv: 2302.01616.

[Vignaud2021] IGAN: Inferent and Generative Adversarial Networks, NATO Specialist Meeting SET-273 RSM on Multidimensional Radar Imaging for ATR, arXiv 2109.13360.

[Peebles2023] W Peebles et S. Xie, Scalable Diffusion Models with Transformers, ICCV 2023.

[Hatamizadeh2024] A. Hatamizadeh et al, DiffiT : Diffusion Vision Transformers for Image Generation, ECCV 2024.

Est-il possible d'envisager un travail en binôme ? Non

Méthodes à mettre en oeuvre :

- | | |
|---|--|
| <input checked="" type="checkbox"/> Recherche théorique | <input type="checkbox"/> Travail de synthèse |
| <input checked="" type="checkbox"/> Recherche appliquée | <input checked="" type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale | <input type="checkbox"/> Participation à une réalisation |

Possibilité de prolongation en thèse : Oui (sur sujet proche, mais pas dans la continuité directe)

Durée du stage : Minimum : 4 mois Maximum : 5 mois

Période souhaitée : début courant premier semestre 2025

PROFIL DU STAGIAIRE

Connaissances et niveau requis :
Deep Learning, modèles génératifs, Pytorch

Ecoles ou établissements souhaités :
Master 2 ou école d'ingénieur