

Utilisation du Deep Learning pour la Morphométrie et la Systématique

Nicolas Parisey

UMR IGEPP - Équipes Démécologie

En collaboration avec : équipes IGEPP DEBI & PMB & MIR, UMRs IRHS - LaBRI - IRISA - EcoBio - ESE - LISAH, Univ Picardie, ANSES Unité Nématologie

Webinaire CATI IMOTEP

Un des thèmes de l'équipe depuis 2017

Aider à répondre aux questions épidémiologiques en santé des plantes en augmentant le volume et/ou la précision des données.

Un des thèmes de l'équipe depuis 2017

Aider à répondre aux questions épidémiologiques en santé des plantes en augmentant le volume et/ou la précision des données.

En pratique

- conception et mise en place de protocoles optimisés pour les acquisitions d'images uni ou multimodales

Un des thèmes de l'équipe depuis 2017

Aider à répondre aux questions épidémiologiques en santé des plantes en augmentant le volume et/ou la précision des données.

En pratique

- conception et mise en place de protocoles optimisés pour les acquisitions d'images uni ou multimodales
- adaptation et développement d'algorithmes dédiés à l'extraction de caractéristiques pertinentes pour la résolution des problèmes posés (e.g. classification) et de chaînes de pré et post-traitements (e.g. recallage)

Un des thèmes de l'équipe depuis 2017

Aider à répondre aux questions épidémiologiques en santé des plantes en augmentant le volume et/ou la précision des données.

En pratique

- conception et mise en place de protocoles optimisés pour les acquisitions d'images uni ou multimodales
- adaptation et développement d'algorithmes dédiés à l'extraction de caractéristiques pertinentes pour la résolution des problèmes posés (e.g. classification) et de chaînes de pré et post-traitements (e.g. recallage)
- assimilation des données d'imagerie dans des modèles mécanistes répondant formellement aux questions posées (e.g. spatio-temporelles)

Un des thèmes de l'équipe depuis 2017

Aider à répondre aux questions épidémiologiques en santé des plantes en augmentant le volume et/ou la précision des données.

En pratique

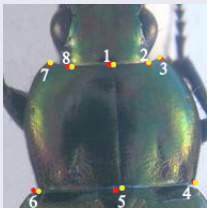
- conception et mise en place de protocoles optimisés pour les acquisitions d'images uni ou multimodales
- **adaptation et développement d'algorithmes dédiés à l'extraction de caractéristiques pertinentes pour la résolution des problèmes posés** (e.g. classification) et de chaînes de pré et post-traitements (e.g. recallage)
- assimilation des données d'imagerie dans des modèles mécanistes répondant formellement aux questions posées (e.g. spatio-temporelles)

Morphométrie et Compagnie

- Morphométrie géométrique e.g. SPE DevMAP
- Taxonomie intégrative e.g. ANSES-INRAE NemAlliance
- Morphogénèse des plantes e.g. FEDER GWASSICA
- Symptomatologie de maladies fongiques (detect, id, quanti) e.g. CASDAR ATIPICAL
- Systématique des plantes e.g. Labex Agro Agrichannels + Bourse M2 PAYOTE

But : Comprendre les relations "Morphotypes X Environnements" chez des auxiliaires et ravageurs des cultures (2D CNN regressions)

Automated landmarking for insects morphometric analysis



(a) Morphométrie géométrique ...

	From scratch		Fine-tune	
	cor	p-value	cor.	p-v.
Pronotum	0.4118	0.001	0.7424	0.001
Head	0.5145	0.001	0.7337	0.001
Elytra	0.2513	0.407	0.3025	0.069

(b) ... et régressions sur variables prédites

VL Le, M Beurton-Aimar, A Zemhari, N Parisey, MAPR (IEEE), 2018

VL Le, M Beurton-Aimar, A Zemhari, A Marie, N Parisey, Ecol Inf, 60, 2020

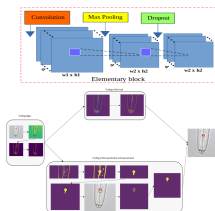
Collab : UMR 5800 LaBRI

Vers une classification et identification automatique des espèces de nématodes de quarantaine *Globodera* basée sur l'analyse d'images

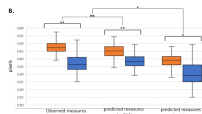
Caractériser et extraire automatiquement de nouvelles métriques



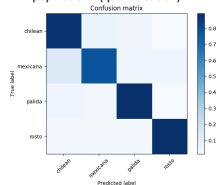
Comparer des approches deep learning et vision numérique



Quantifier les dissemblances



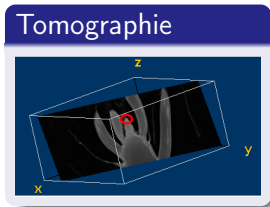
Maximiser l'efficacité d'assignation des populations (quantification)



Thèse de R Thevenoux 2017-2020
Thevenoux et al, *Computer and Electronics in Agriculture*, 2021

- **Systématique** : Une assignation sans erreur à *G. pallida* et *G. rostochiensis* à l'échelle des populations
- **Taxonomie intégrative** : Des résultats morphométriques concordants avec l'identification moléculaire d'une nouvelle espèce au sein de ce genre
- Collaboration **RA** **Demecol** / ANSES

But : Détecter, à des stades précoces et de façon non-invasive, des défauts liés à l'impact du changement climatique sur certaines variétés cultivés.



- Machine learning (ML) avec caractéristiques dédiées ,
- 2D CNN (slices) avec transfer learning (TL) ou fine-tuning ,
- 3D CNN 'from scratch'.

Best = ML ; **Second Best** = 3D CNN

Zhou et al, 2021, in revision for Computers and Electronics in Agriculture

Pour aller plus loin : TL sur 3D ; Deep Multimodale ; Deep "Spectres Dédiés" (mais pré-étude hyperspectrale peu concluante)

Collab : UMR OPAALE & IGEPP-DEBI

But : Détecter, identifier et quantifier les maladies fongiques de plantes cultivées

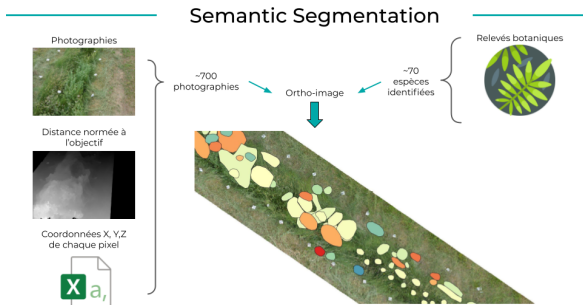
Historiquement plutôt traiter dans l'équipe par Machine learning uni- ou multi-modale (Bousset et al, Plant Pathology, 2019 ; Bousset et al, 2021, in revision for Fungal Ecology) mais passage au Deep.

Exemple : Mimétique et ambiguïté des classifications expertes (en cours)

- Utilisation de CNN et ML pour l'identification de symptômes sur un dataset mono-expert mono-annotation mais de grande taille ;
- Prédiction des symptômes sur un autre dataset multi-experts multi-annotations mais de petite taille ;
- Evaluation des discrédances entre prédiction morphologique 'pure' (e.g. CNN) et classification expertes (images + contextes + autres)

Collab : UMR BioGER

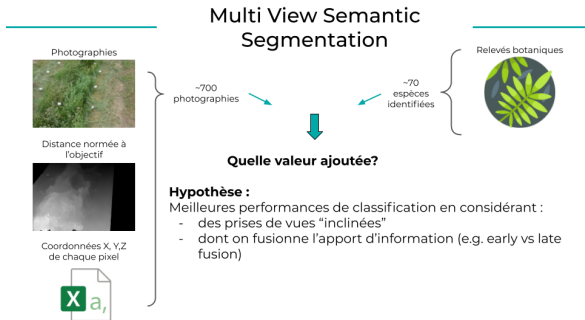
But : Identifier automatiquement des espèces végétales d'intérêt écohydrologique *in situ*



Travail préliminaire sur jeux de données réduit :
CNN \geq ML

Collab : UMR LISAH

But : Identifier automatiquement des espèces végétales d'intérêt écohydrologique *in situ*



Travail préliminaire sur jeux de données réduit :
Multi-View CNN \geq (?) CNN \geq ML

Collab : UMR LISAH

Take home message ?

Le Deep,

- ça peut servir à (presque) tout (CNN, RNN, R-CNN, ...)
- ça peut (aussi) se faire sur des jeux de données de taille modeste.

Le Deep,

- ça peut servir à (presque) tout (CNN, RNN, R-CNN, ...)
- ça peut (aussi) se faire sur des jeux de données de taille modeste.
- c'est gourmand.
- ça n'est pas une panacée : pas de gestion 'auto-magic' des jeux de données non balancés, trop bruitées, pas *systématiquement* les algos les plus performants ...