

Le deep learning appliqué au phénotypage des cultures

B. de Solan

ARVALIS, LPA CAPTE

Avignon



Le phénotypage haut-débit et la phénomique

Depuis 15 ans environ, le phénotypage haut-débit a connu une attention particulière avec la perspective de mieux comprendre la fonction des gènes dont le séquençage est devenu une activité de routine : le phénotypage est « le goulot d'étranglement de la sélection »

- 1^{ère} conférence sur la « Phenomics » : 2009, CSIRO, Canberra
- Objectif de la phénomique : établir le lien entre l'information génétique, le fonctionnement de la plante et les traits variétaux en contexte agricole (Furbank & Tester, 2011).
- En pratique, développement d'outils et de méthodes pour :
 - Acquérir les données au champ
 - Traiter et organiser ces données
 - Intégrer ces données dans des modèles pour identifier les caractères fonctionnels des variétés.
- Une communauté s'est structurée au travers de projets et réseaux nationaux (PHENOME en France), européens (EPPN et Emphasis) et internationaux (IPPN).
- Un intérêt croissant des sélectionneurs privés
- De nouvelles applications apparaissent, notamment en lien avec l'agroécologie



Pourquoi et comment Arvalis adopte ces outils dans ses métiers

- Objectif : numériser l'acquisition de références au champ

- Aide à la sélection / évaluation variétale
- Aide à la conduite des cultures
- Aide à l'évaluation des nouveaux systèmes de cultures

- Avantages attendus importants :

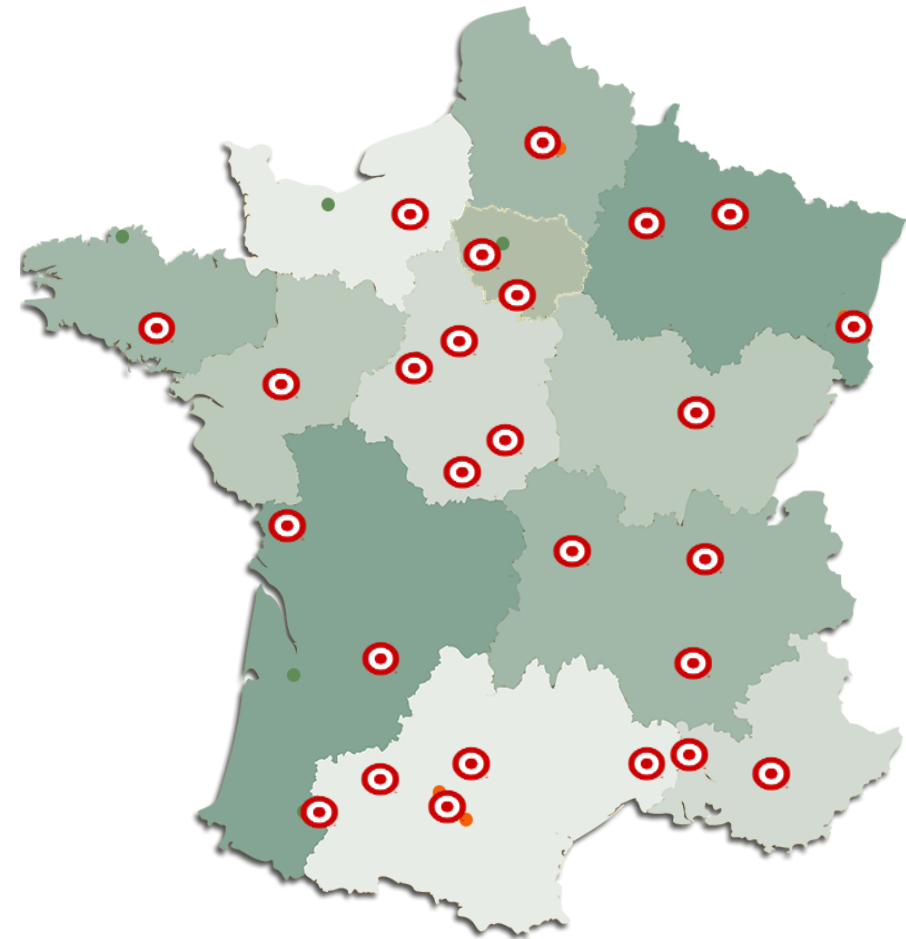
- Gain en temps de travail sur les activités de phénotypage traditionnelles (évaluation variétale)
- Objectiver les mesures par des techniques plus répétables
- Répondre aux nouveaux contextes d'acquisition de références, notamment agroécologiques, mettant en œuvre des systèmes cultivés plus complexes (e.g. cultures associées)
- Rendre le travail de technicien plus attractif et limiter la pénibilité

- Mais avec des freins à l'adoption :

- Des mesures différentes de celles dites de référence
- Une validation pas toujours aisée
- Une modification de l'organisation des équipes

Outils de phénotypage

Une stratégie de déploiement national d'outils de phénotypage, adaptés aux problématiques locales



PhenoField



PHENOMOBILE



ALPHI



PHENOMOBILE 2



Drone RVB



Drone multispectral



Minirhizotron


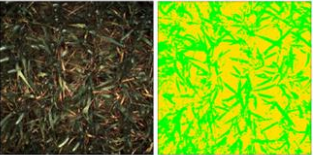

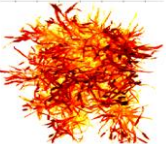

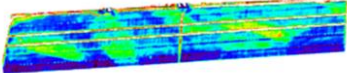

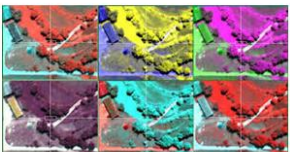

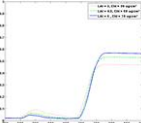


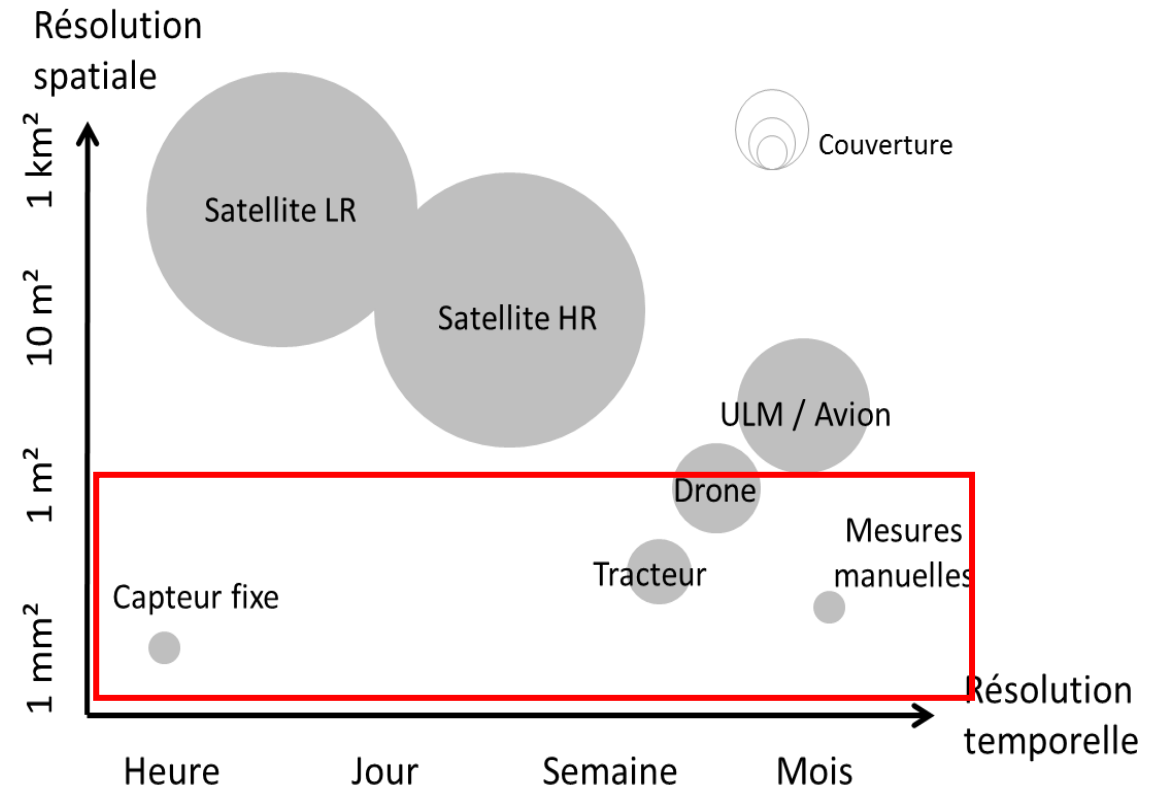
Literal



Les principaux capteurs optiques

Une mesure de l'interaction lumière / végétation

		Utilisations	
	Camera haute résolution/RVB		<ul style="list-style-type: none"> • Identification adventices • Enherbement • Taux de couverture verte • Développement foliaire • Comptage de plante • Dégâts de nuisibles • Biomasse, risque de verse
	LIDAR Mesure de distance et profondeur		<ul style="list-style-type: none"> • Structure • Port du couvert • Interception lumineuse • Expérimentation
	Caméra IR thermique		<ul style="list-style-type: none"> • Température de surface • Stress hydrique • Pilotage de l'irrigation
	Camera multispectrale		<ul style="list-style-type: none"> • Développement • Statut azoté/chlorophylle • Stress • Pilotage de l'azote
	Spectromètre		<ul style="list-style-type: none"> • Chlorophylle/statut azoté • Stress (hydrique, ...) • Développement • Composés de la plante



Un traitement mobilisant massivement le deep learning

- Donne accès à des variables très proches des variables mesurées par les équipes techniques
- Permet une standardisation des méthodes de traitement
- /!\ Besoin en données d'entraînement
 - Diversité
 - Qualité
- /!\ Robustesse des modèles
 - Adéquation entre données d'entraînement et domaine d'application
- /!\ Volume des données à gérer et puissance de calcul

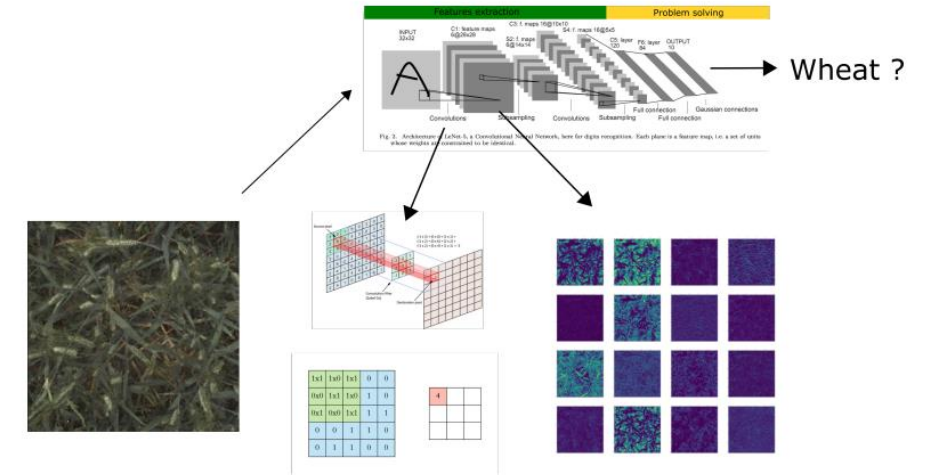
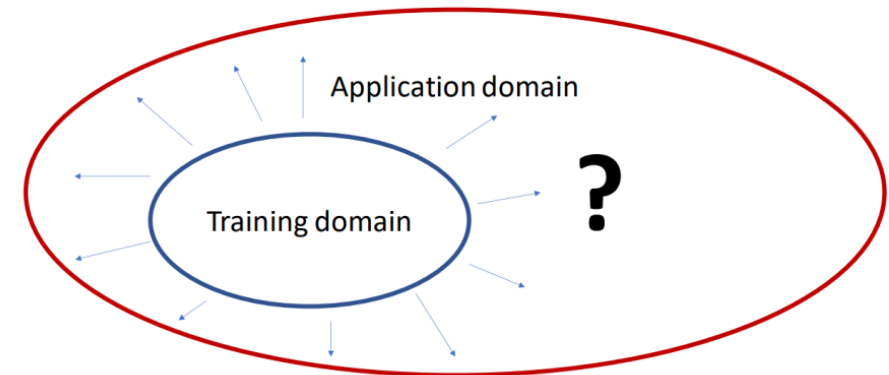


Figure 11. Illustration of the processing of a RGB image by a CNN



Classification, détection, localisation, segmentation : que veut-on voir ?

- **Classification : identifier un objet dans une image**

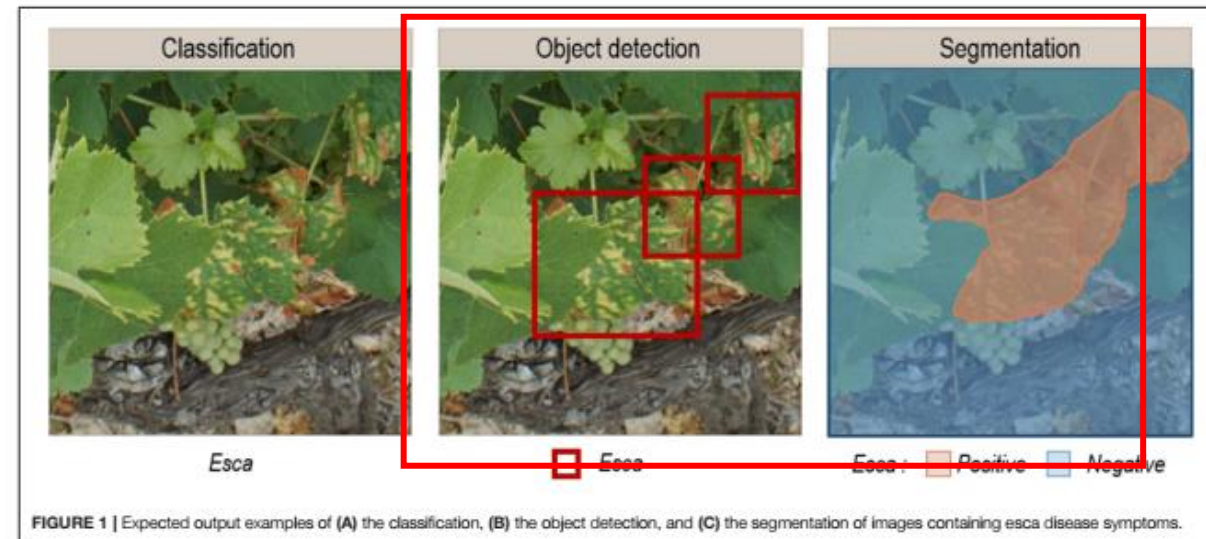
- Y-a-t-il de la maladie cette image ?
- O / N

- **Localisation / détection d'objets**

- Où y-a-t-il de la maladie dans cette image ?
- Coordonnées d'un rectangle sur les zone malades

- **Segmentation sémantique**

- Quels sont les pixels de l'image qui appartiennent à la classe « maladie » et ceux appartenant à la classe « sain » ?
- Nombre / coordonnées de pixels

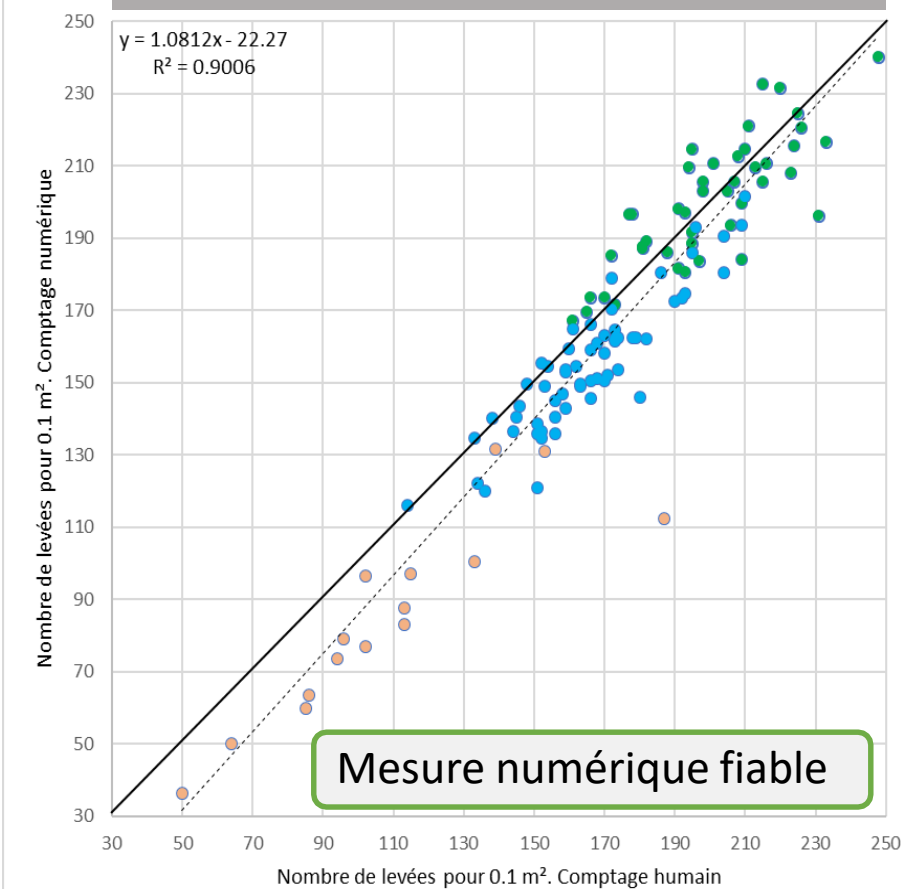


Remplacer des mesures existantes

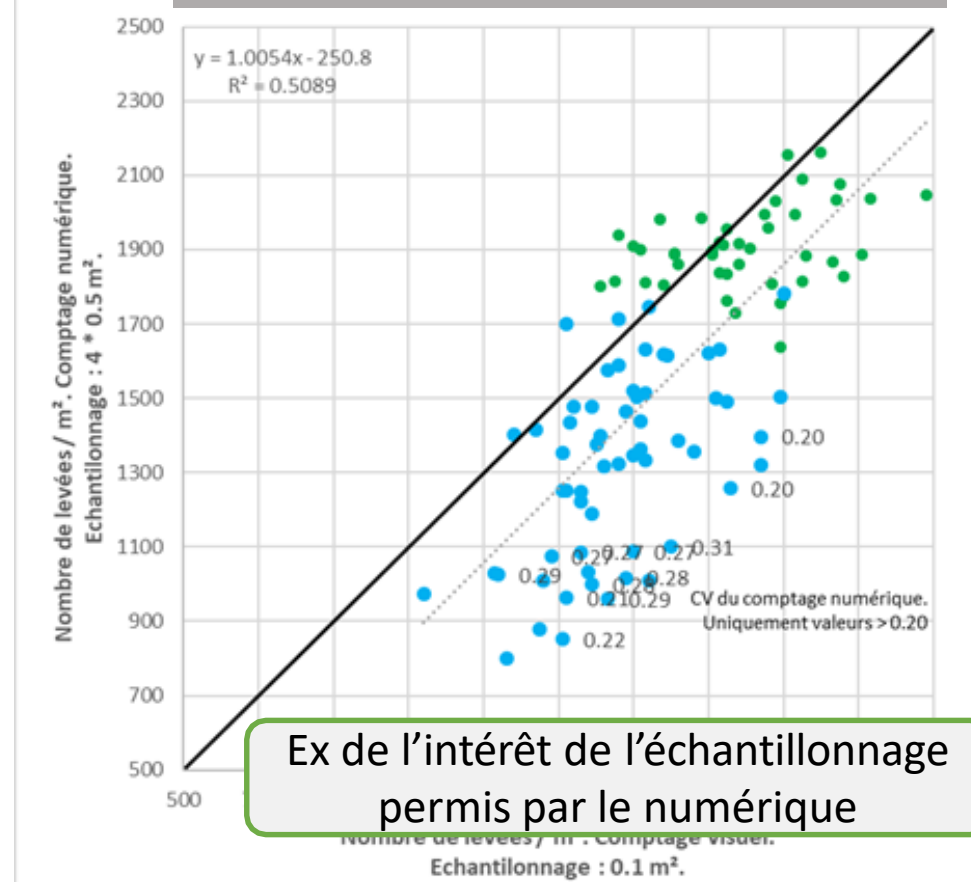
Expl : Comptage de plantes de lin



Comparaison comptage numérique et humain sur les zones de référence de 0.1 m².

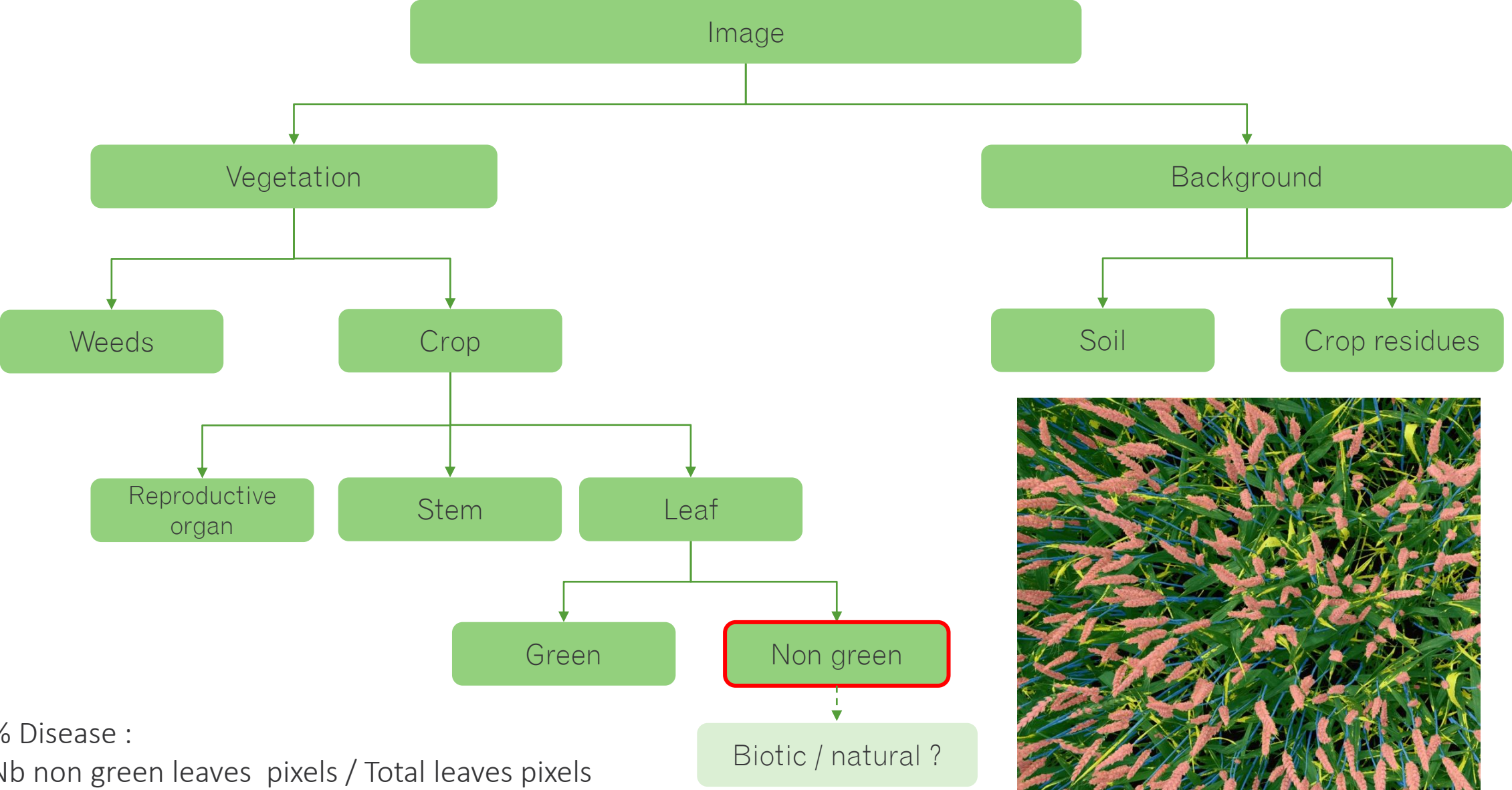


Comparaison comptage numérique et humain par microparcelle



Objectiver des estimations visuelles

Expl : estimation de la sénescence des cultures

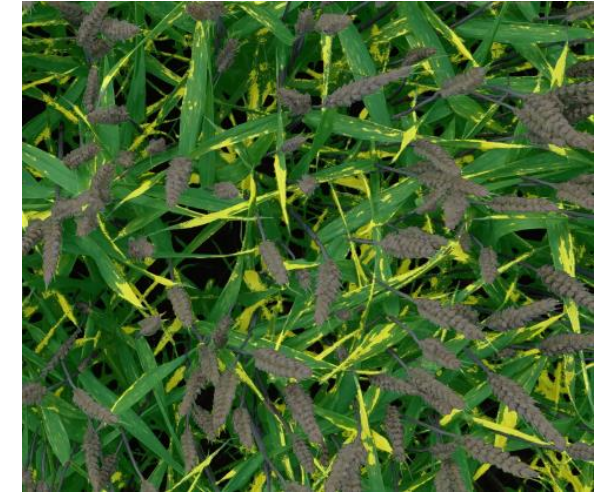
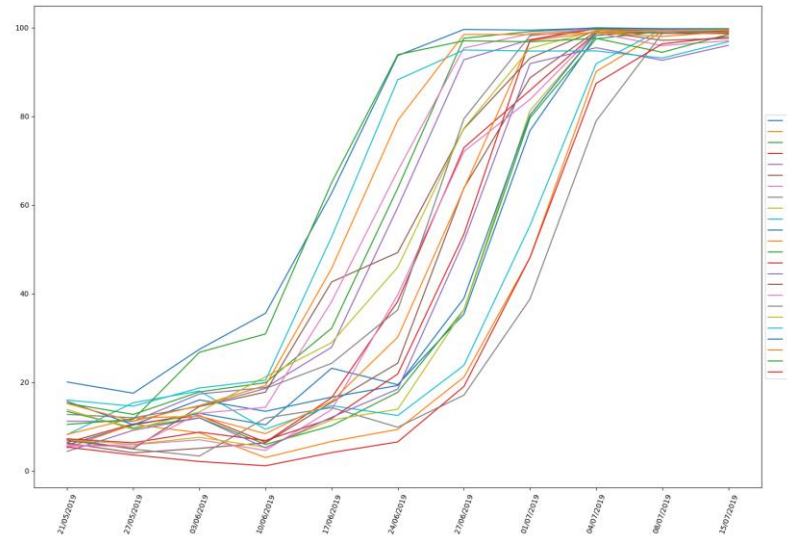


% Disease :
$$\frac{\text{Nb non green leaves pixels}}{\text{Total leaves pixels}}$$

Suivi de la sénescence

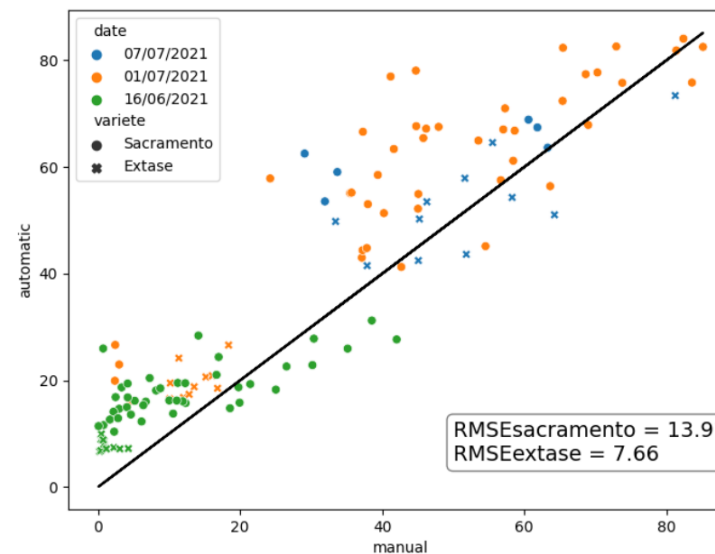
- Suivis de dynamiques de sénescence naturelle cohérents

Graphe : évolution de la fraction de sénescence foliaire, PHENOFIELD 2019



- Bonne cohérence avec notations visuelles de maladie

Graphe : comparaison notation visuelle de sévérité de septoriose vs notation automatique, Villers St Christophe 2021



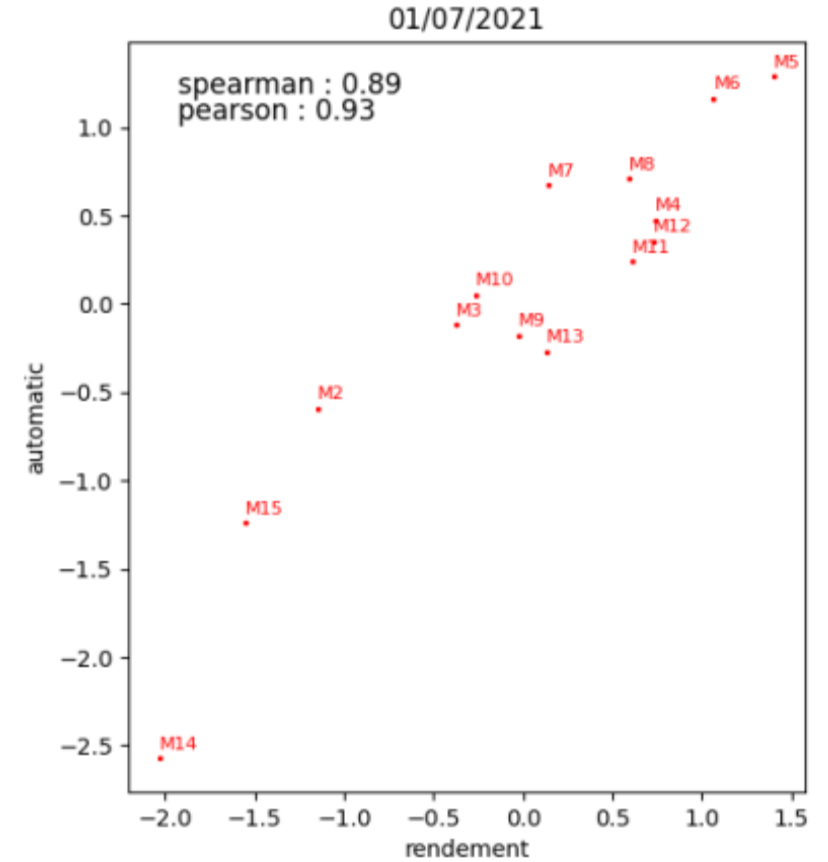
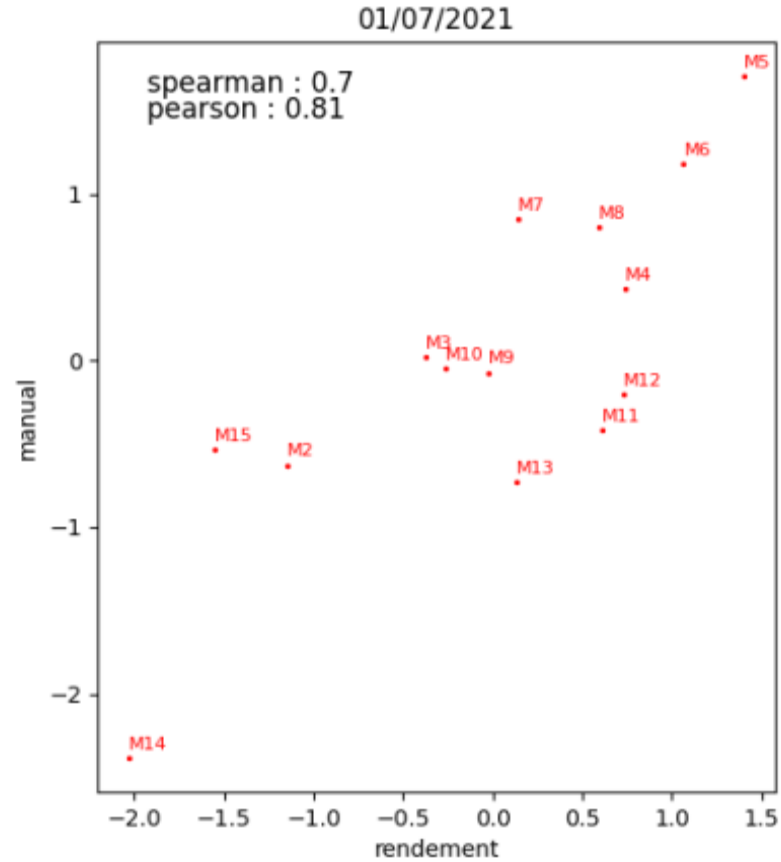
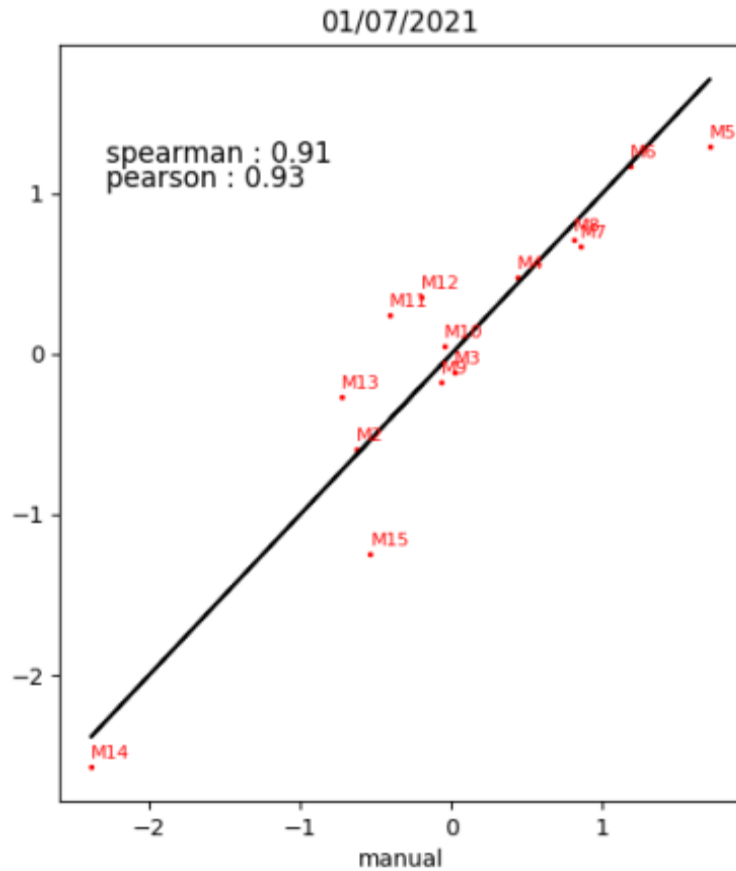
→ 2 varieties:
- 1 with beard (Sacramento)
- 1 without (Extase)

→ Good correlation for both case but much better for variety without beard.

L. Bernigaud, 2022

Suivi de la sénescence

Septoriose sur feuilles (Villers St Christophe, 2021)



Automatic & manual : Sensibility relative to mean of notations.

Rendement: Loses of yield relative to mean yield.

Ranking with automatic notations is similar to ranking with manual notations and it's better correlated to yield than manual ranking.



Enjeu 3 : Estimation de nouveaux traits - Expl des couverts complexes



Plot406_Camera2_1



Plot106_Camera2_3



Plot304_Camera2_2



Plot103_Camera2_1



Plot110_Camera2_2



Plot108_Camera1_2



Plot305_Camera2_4



Plot215_Camera1_2



Plot213_Camera1_4



Plot408_Camera1_4



Segmentation sémantique plurispécifique

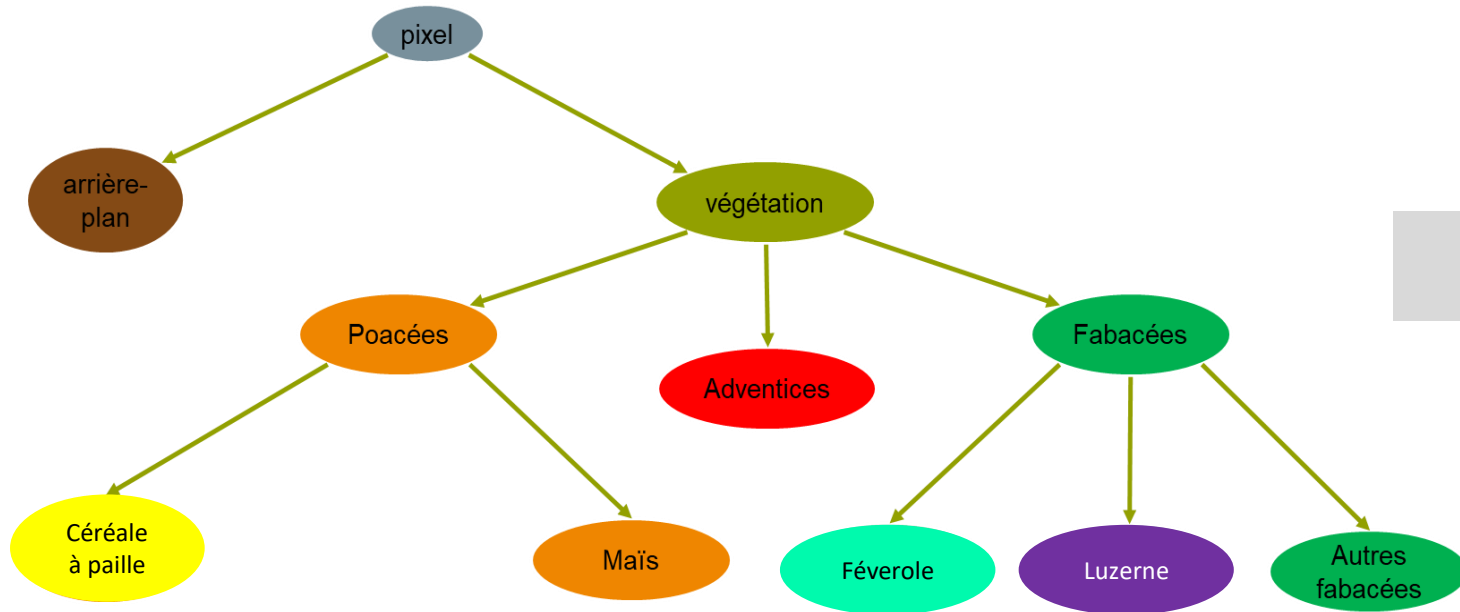


Image RVB



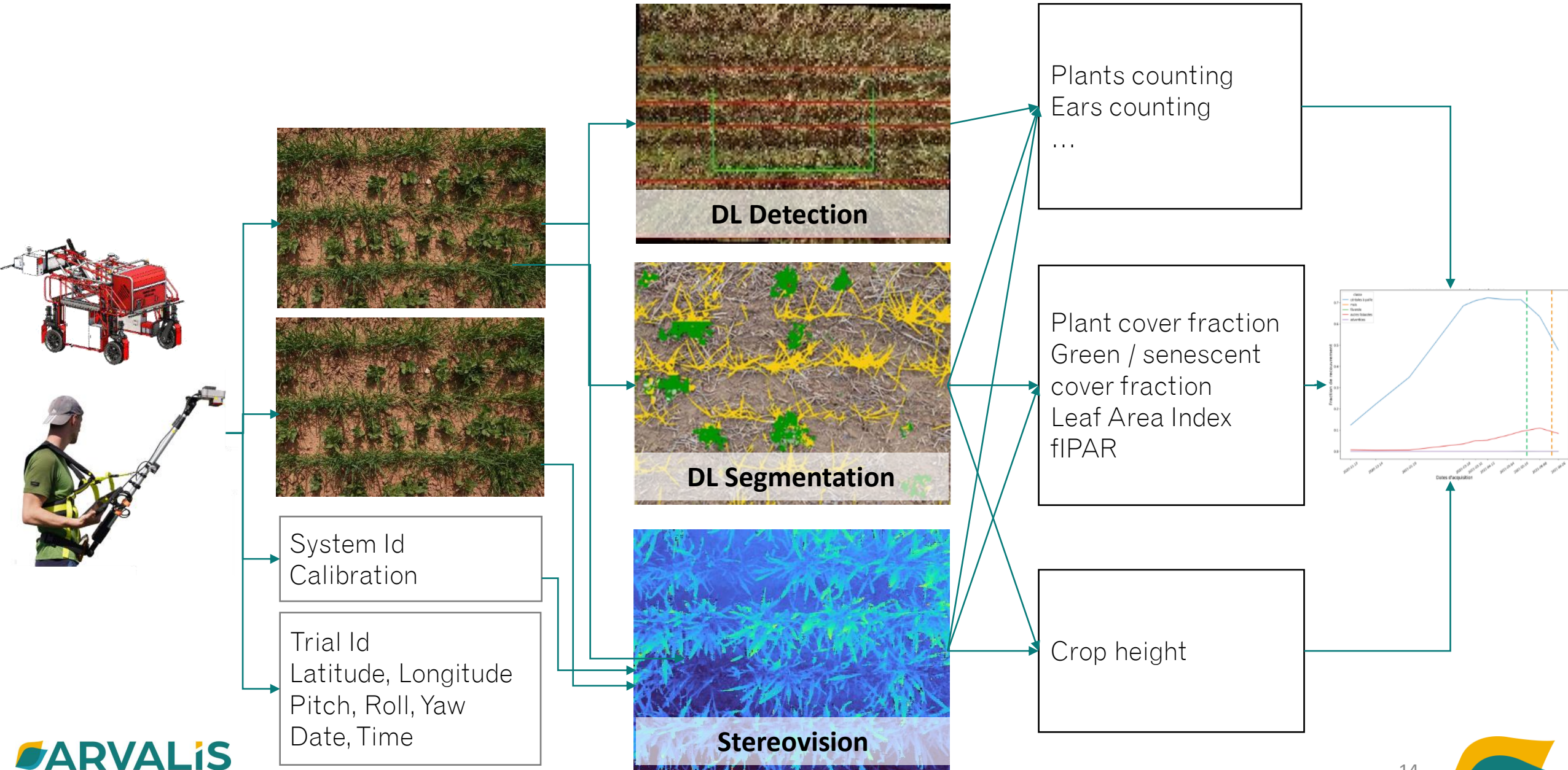
Segmentation sol / végétation



Segmentation sémantique multi-espèces



Estimation de traits complexes



Enjeu 1

Collecter des données diverses et de bonne qualité

- Initiatives coordonnées :
 - Data challenges (Kaggle, AICrowd)
 - Sites de partage de jeux de données (Zenodo, OpenML)



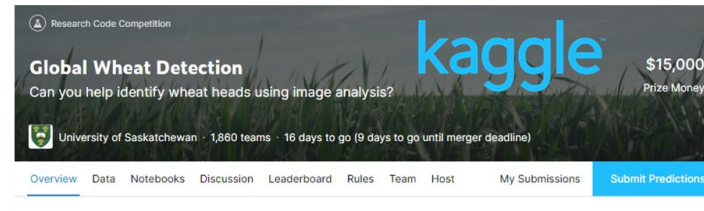
7 countries

4700 images

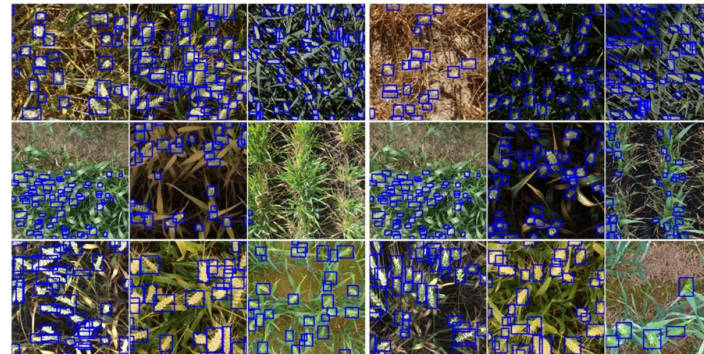
185.000 labeled wheat head



E. David, 2020



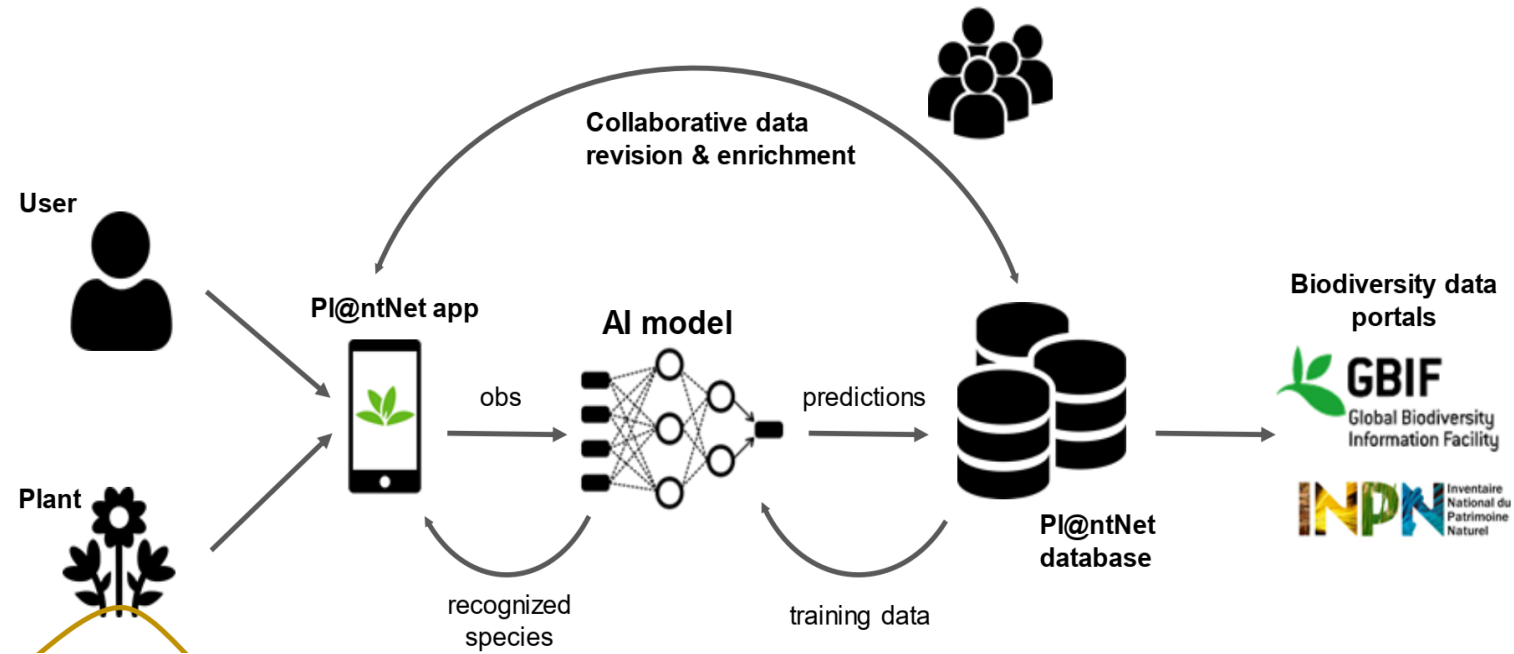
- Mai – Août 2020
- 2235 équipes participantes
- Les 3 meilleurs modèles mis à disposition



Enjeu 1

Collecter des données diverses et de bonne qualité

- Apprentissage coopératif (PI@ntnet)



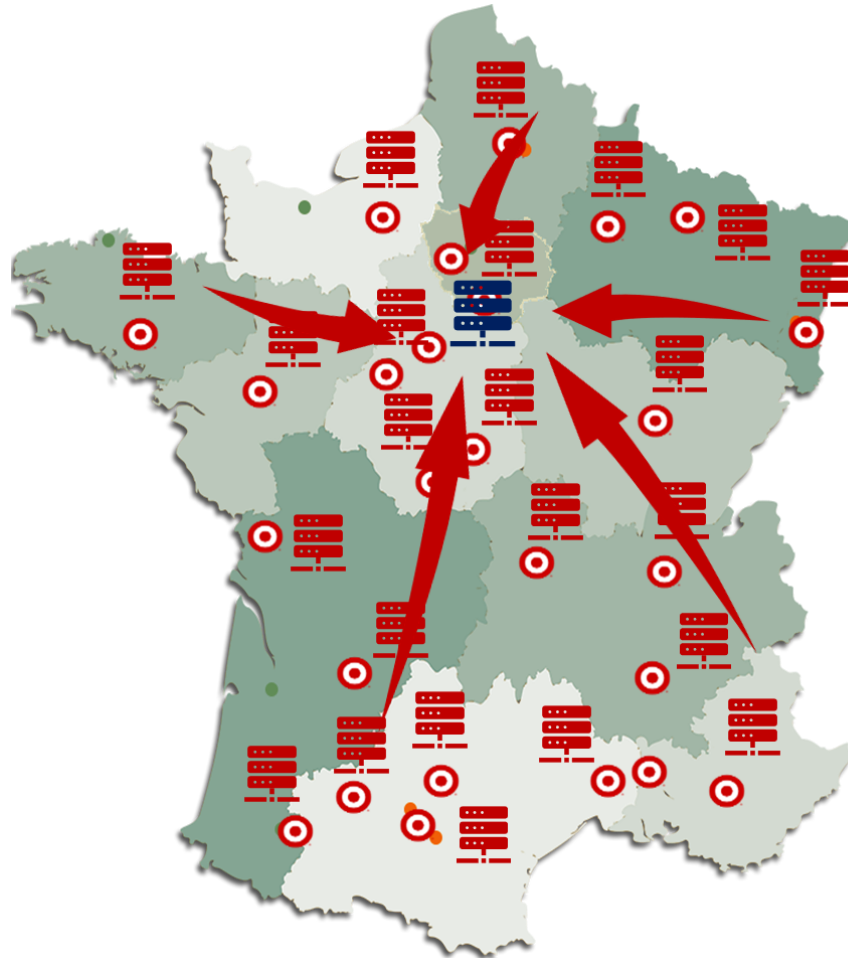
- Intérêts :


- Collecte et révision des jeux de données d'apprentissage
- Collecte également des observations de manière distribuée
- Source complémentaire aux réseaux d'observation du territoire (BSV, Vigicultures®)




Enjeu 2









Gérer les volumes de données générés



 Copie locale des données brutes acquises sur le terrain dans un répertoire dédié sur le NAS de la station expérimentale (~ 20 To / an)

Transfert par mise en place de synchronisations automatiques entre NAS

 Centralisation à Boigneville sur le NAS dédié aux données capteur (60 To)

- ▼  RAWDATA
- ▼  SITE
- ▼  ANNEE
- ▼  SYSTEME D'ACQUISITION
- ▼  ESPECE
- ▼  ESSAI
 -  Session de mesure 1
 -  Session de mesure 2

Duplication automatique de l'ensemble des données vers un site SharePoint par campagne pour archivage

 SharePoint Archivage (2 Po)

- A terme : traitement embarqué pour simplifier la gestion des données



Enjeu 3

Analyser les larges de bases de données produites

- Données générées par les traitements
 - Pour ARVALIS : 1 million d'observations unitaires / an en 2023 -> forte augmentation
 - Nécessite des procédures de validation automatique poussées
- Les données générées permettent de nouveaux types d'analyses
 - Analyses spatiales et temporelles
 - Fusion de données multi-échelles et multi-sources

On en est au commencement !



Conclusion

- Le deep learning a révolutionné le phénotypage des cultures, il s'impose comme la méthode de référence pour l'analyse de données issues de capteurs
- Il constitue une étape initiale d'analyse de ces données, complétée par d'autres traitements plus classiques
- Il permet d'automatiser des tâches répétitives mais donne aussi accès à de nouveaux traits, notamment pour l'analyse de scènes complexes
- Il comporte des contraintes opérationnelles fortes (volumes de données, puissance de calcul, nécessité de large jeux de données d'entraînement)
- Les acteurs du domaine apportent cependant constamment de nouvelles solutions techniques. C'est donc un domaine en constante et très rapide évolution
- L'étape à venir consiste à valoriser pleinement les larges volumes d'observations générées... via des méthodes de machine learning !

Merci pour votre attention



Contact: b.desolan@arvalis.fr

Membre de



Partenaire technique ACTIA

Avec la contribution financière du compte d'affectation spéciale développement agricole et rural CASDAR



MINISTÈRE DE L'AGRICULTURE ET DE LA SOUVERAINETÉ ALIMENTAIRE

Liberté
Égalité
Fraternité

PHENOME
EMPHASIS | FRANCE

anr[©]
agence nationale de la recherche

En partenariat avec les filières

(Intercéréales, SEMAE, FNPSMS, CNIPT, GIPT, CIPALIN, FNAMS, FNPT)

#DigitAc