



Simulations de données afin d'améliorer la segmentation automatique de la tavelure du pommier

Clément Douarre¹²³, David Rousseau², Laure Tougne¹, Anthony Gelibert³, Gérald Germain³

1. LIRIS (Université Lumière Lyon 2)
2. LARIS (Université d'Angers)
3. Carbon Bee Valence

Orateur : Clément DOUARRE

La tavelure du pommier est une maladie fongique affectant les arbres du genre *Malus* et se caractérisant par des lésions brunâtres sur les feuilles et les fruits. Au vu de la perte potentielle de production, la détection de cette maladie le plus tôt possible et de façon la plus automatisée possible est de grande importance pour les agriculteurs. C'est dans ce but que nous travaillons, avec l'entreprise Carbon Bee, à la détection automatique de tavelure sur des images de pommiers, par apprentissage automatique supervisé avec des réseaux de neurones profonds. Pour que l'apprentissage fonctionne, il est cependant nécessaire d'annoter une grande quantité de données, une tâche largement reconnue comme chronophage, fastidieuse et coûteuse. En conséquence, les bases annotées comprennent souvent peu de données.

Nous avons travaillé sur la problématique de l'augmentation du jeu de données à partir du jeu de données initial. Il s'agit de générer des images annotées nouvelles, en se basant sur des informations et métriques tirées du jeu de données initial. Le but est de réduire la charge d'annotation de vraies données, en faisant apprendre les réseaux (en partie) sur les données synthétiques.

Nous avons développé à ces fins deux simulateurs de données. Un simulateur « modèle » est basé sur une génération de scène de champs à partir de photos de feuilles de pommier isolées. Le deuxième est basé sur les Generative Adversarial Networks, des réseaux capables de générer des images synthétiques ressemblant, du point de vue d'un réseau de neurones, à des images réelles également fournies en entrée.

Dans les deux cas de simulateur, les données sont générées avec leur vérité terrain, ce qui permet de les utiliser pour l'entraînement sans avoir à les ré-annoter. Nous montrons que ces données générées peuvent aider à l'apprentissage dans le cas de bases d'apprentissage de petite taille.