

HMSN208 - Rapport de stage

Guilhem BARTHES

20 juillet 2017

Stage effectué au sein de l'unité de recherche HortSys (département PerSyst) du Cirad

Encadrants

Philippe BORIANNE

Frédéric BORNE

Émile FAYE

Supervision pédagogique

Sylvain DAUDÉ



UNIVERSITÉ
DE MONTPELLIER

Table des matières

1	Introduction	3
2	Adaptation du réseau	4
2.1	Projets connexes	4
2.2	L'environnement	5
2.3	Spécialisation du réseau	5
3	Évaluation du réseau	6
3.1	Sélection de l'annotateur et choix de l'indicateur	6
3.1.1	Choix de l'indicateur de performance : la F-mesure	6
3.1.2	Reproductibilité	7
3.1.3	Variabilité inter-individuelle	8
3.2	Influence des conditions d'acquisitions sur les résultats	9
3.2.1	La luminosité	9
3.2.2	La saturation	10
3.3	Impact du paramétrage du réseau sur les résultats	11
3.3.1	nombre d'itérations	11
3.3.2	Seuil de sélection	12
3.3.3	Non-Maximum Suppression	12
4	Conclusion	13
	Annexes	I
A	Exemple de resultat de segmentation d'image	I
B	Exemple de fichier d'annotations utilisé par notre réseau de neurones	I
C	Exemple des modifications des conditions d'acquisition	II
	Références	III
	Glossaire	IV
	Acronymes	IV

1 Introduction

L'analyse de la productivité des cultures représente un enjeu stratégique pour les pays en voie de développement, autant au niveau économique (amélioration de la production)[1] qu'au niveau de la sécurité alimentaire (autonomie).

L'estimation du rendement est une information nécessaire et stratégique pour les acteurs publics et privés. Cependant, les outils d'estimation de récolte restent inexistantes ou précaires pour la plupart des cultures dans les pays du Sud. Le projet PixYeld au sein de l'unité de recherche HortSys¹ vise donc à combler ce manque par une analyse d'image multi-échelle, fournissant des estimations spatialisées de rendement des manguiers sénégalais, utiles autant pour l'export mais aussi pour le marché local.[2]

L'analyse d'image est maintenant une voie majeure de recherche pour l'estimation du rendement d'arbres fruitiers. Certains travaux mettent en avant des résultats prometteurs : la segmentation d'images nocturnes sous lumière artificielle a permis de détecter jusqu'à 78% des mangues comptées[3]. Certaines autres études montrent des résultats très probants, avec 90% de pommes détectées sur leurs arbres[4].

Les méthodes habituelles de classification présentent des limites que ce projet vise à dépasser. La première est la gestion des objets aux limites des classes. Par exemple, dans le cadre de l'utilisation d'un classificateur de type k -NN, celui-ci aura plus de difficultés à classer les mangues ne correspondant pas aux critères définis par l'utilisateur. De plus, ce type de technique nécessite de définir les caractéristiques prises en compte pour l'identification.

Les réseaux de neurones répondent à cette problématique d'estimation de rendement. L'objectif initial de ce stage était de déployer un réseau permettant d'effectuer de la segmentation[5], c'est-à-dire prédire pour chaque pixel la classe associée (exemple disponible en annexe A). On pourrait donc différencier la partie de l'image correspondant à des mangues du reste.

La limite majeure de la segmentation est l'absence de la gestion des occlusions. En effet, les prédictions se font sur chaque pixel et il n'y a pas de comptage d'objet possible (les prédictions se faisant par pixel). Or ce type d'information est essentiel afin de pouvoir estimer la production de mangues.

Des études montrent le potentiel d'extrapolation de ce type d'approche. Certains chercheurs ont réussi à re-spécialiser un réseau de neurones pour la détection de poivrons alors qu'il avait été entraîné pour reconnaître des melons, [6].

Malgré le nombre important d'articles évoquant l'utilisation d'un réseau de neurone pour le comptage de fruits, notamment sur des arbres fruitiers, aucun ne s'intéresse à l'adaptation d'un réseau de neurone à un cas spécialisé, ni à l'influence de divers paramètres sur les résultats fournis. Ce sont donc les deux grands points que nous allons traiter.

1. HortSys est une équipe de recherche spécialisée dans l'horticulture. Cette discipline est désormais considérée comme une composante essentielle de la sécurité et de l'équilibre alimentaire mondiaux. Les systèmes horticoles sont également une source importante de revenus et d'emplois pour les populations les plus pauvres en zone tropicale. Mais sensibilité des cultures aux bio-agresseurs et les méthodes de lutte contre ceux-ci font le plus souvent appel à des pesticides, entraînant des risques sur la santé humaine et l'environnement. Le défi global consiste donc à concilier à la fois une production horticole de qualité en quantité suffisante pour satisfaire une demande mondiale en croissance et permettre le développement économique et social des paysans des pays du Sud, et la limitation des risques pour la santé humaine et les écosystèmes. Cette unité de recherche fait parti du département PerSyst (Performances des systèmes de production et de transformation tropicaux) du CIRAD, étudiant les productions tropicales.

2 Adaptation du réseau

Nous commencerons par évoquer les différents projets existants dans la littérature. Nous décrirons ensuite l’environnement utilisé pour mener les expériences, avant d’évoquer le principe de spécialisation d’un réseau de neurone.

2.1 Projets connexes

La plupart des projets similaires utilisent non pas de la segmentation, mais des réseaux de détection et localisation. Cette approche permet de résoudre la problématique de la gestion des occlusions. le réseau est capable dans la plupart des cas de distinguer les différents fruits se chevauchant, fournissant une estimation précise du nombre de fruits présents dans l’arbre.

Un autre avantage est la rapidité d’annotations. L’annotation consiste à déterminer une boîte englobante pour chaque fruit détecté, qu’il soit partiellement ou totalement visible. Cette annotation ne nécessitant pas de distinguer les pixels de l’image, le traitement est de ce fait très rapide : on peut approximativement annoter 848 mangues sur 200 images (de 500 par 500 pixels) en 1 minute.

Les différents projets de détection de mangues sur des manguiers mettent en évidence des caractères communs nuisant à la qualité des résultats : les fruits sont variable au niveau de la couleur, mais aussi de la forme, de la taille, de la texture et sont souvent partiellement visibles[6, 7]. Cela rend plus difficile à appréhender les caractéristiques à conserver dans les méthodes de *machine learning*. Ce problème est simplifié avec les réseaux de neurones, ceux-ci sélectionnant les éléments clés à retenir (on parle de *feature engineering*).

Deux projets nous ont particulièrement intéressés, tout deux reposant sur le principe de détection des fruits plutôt que la segmentation.

Le premier a avoir attiré notre attention est celui disponible dans l’article [7]. Celui-ci propose un objectif similaire, avec prise de vue automatisée et analyse effectuée par un réseau de neurones, mais l’article met surtout en avant la qualité des résultats obtenus, ne s’interrogeant pas sur la validation de ceux-ci.

Un autre projet a plus particulièrement mis en évidence l’adaptabilité des réseaux de neurones[6]. En effet, la réutilisation de l’entraînement effectué sur des poivrons, en appliquant la même méthode à d’autres fruits, montre des résultats très concluants.

Ces deux projets partagent le même modèle de réseau de neurones, montrant la pertinence des modèles de détection et localisation, et plus particulièrement de leur modèle commun, *py-faster-RCNN*, dans la détection de fruits. Au vu des résultats obtenus dans ces deux publications, nous avons décidé de l’utiliser pour la suite du projet.

Les données d’entraînement du projet australien[7] étant mises à disposition de la communauté scientifique², il nous a été possible d’aborder le lien entre « jeu d’apprentissage » et pertinence des classifications. Ici, le jeu de données fourni est constitué de mangues de variétés différentes de celles concernées par le projet PixYield. De plus, les conditions d’acquisition sont totalement différentes, avec des prises de vue nocturne en Australie (visible en Figure 1a) alors que toutes les prises de vue au Sénégal se font de jour (comme montré en Figure 1b). Les résultats sont disponibles en figure 1. Déjà satisfaisant, il nous posent la question de l’influence des modifications des jeux d’apprentissage et sur les stratégies d’enrichissement de ces jeux de données.

2. Données disponibles à l’adresse suivante : <http://data.acfr.usyd.edu.au/ag/treecrops/2016-multifruit/>.



(a) Test sur les données australiennes

(b) Test sur nos données

FIGURE 1 – Exemple de résultats obtenus en entraînant notre réseau sur les données australiennes.

2.2 L'environnement

Le cadre applicatif retenu pour nos expérimentations est *Caffe*[8]. Il est conçu comme une plateforme pour différents modèles de réseau de neurone.

Il utilise massivement la parallélisation permise par les Graphics Processing Unit (GPU). C'est lui qui se charge de définir les communications avec le matériel et met à disposition des outils afin de rendre la création (et l'utilisation) de réseaux de neurones plus facile.

Comme évoqué précédemment, le modèle finalement retenu est *py-Faster-RCNN*. Il permet d'obtenir des prédictions en 300 millisecondes pour une image de 500 pixels par 500 pixels.

Les annotations qu'il nous donne en résultat (ou qu'on lui fournit) sont des coordonnées formant des boîtes. Un exemple est disponible en annexe B

py-faster-RCNN propose 300 objets par image. Ceci peut rapidement devenir une limite pour des images de grande taille avec de nombreux objets.

2.3 Spécialisation du réseau

La méthode utilisée pour l'entraînement de notre réseau se nomme spécialisation (ou « *fine-tuning* »). Elle consiste à prendre le résultat d'un apprentissage sur une base d'image généraliste (ImageNet, COCO ...) et modifier les dernières couches de neurones afin d'améliorer la connaissance de l'objet cible.

Cette méthodologie permet de ré-utiliser une partie de l'entraînement d'un réseau de neurone, engendrant une diminution du temps d'apprentissage et une amélioration des résultats.

3 Évaluation du réseau

Les facteurs influençant les résultats et la significativité des scores fournis par les réseaux de neurones sont des problématiques peu étudiées dans la littérature. Nous avons donc décidé d'appréhender ceux-ci de manière empirique : nous avons fait varier des conditions d'acquisitions et le paramétrage du réseau de neurones pour voir leurs influences sur les résultats.

N'ayant pas de connaissances sur les indicateurs fournis par le réseau de neurone, la première étape est de déterminer la vérité terrain et de mettre en place un indicateur connu afin d'évaluer les variations de performance du réseau de neurone.

3.1 Sélection de l'annotateur et choix de l'indicateur

Nos expérimentations consistent à comparer des annotations d'images. Un des fichiers d'annotations est considéré comme la référence, tandis que le deuxième est l'élément à comparer. Ces fichiers sont constitués des coordonnées des rectangles entourant nos objets d'intérêt, appelées boîtes englobantes (un exemple est disponible en sous-section B). On compare alors les différentes boîtes englobantes entre elles. La mise en place d'un indicateur externe nécessite l'utilisation de diverses valeurs statistiques. Pour nos expérimentations, nous aurons besoin de trois valeurs de la matrice de confusion : les vrai positif (VP), les faux positif (FP) et les faux négatif (FN). Ceux-ci se définissent de la manière suivante :

VP Le barycentre de la boîte comparée est dans une boîte de la référence.

FP Le barycentre de la boîte comparée n'est dans aucune boîte de référence.

FN Aucun barycentre de l'image comparée tombe dans une boîte de la référence.

3 humains ont annotés les images de manguiers sur pied : Philippe Borianne, Julien Sarron et Guilhem Barthes. Deux d'entre eux, Philippe Borianne et Guilhem Barthes, n'ont que peu d'expérience dans l'analyse d'image de mangues, tandis que le troisième travaille sur l'identification sur le même type d'image grâce à une autre méthode de *machine learning*.

Nous avons aussi intégré les résultats du réseau de neurones dans les analyses suivantes, afin de pouvoir comparer la variabilité de celui-ci par rapport aux annotateurs experts. Deux versions ont été intégrées : la première, notée réseau 80img, a été entraînée sur l'ensemble des données annotées. Le deuxième, réseau 39img, a été uniquement entraîné sur les clichés où les experts étaient unanimes.

Dans un premier temps, nous avons sélectionné un indicateur de performance pour pouvoir évaluer nos différents experts annotateurs.

3.1.1 Choix de l'indicateur de performance : la F-mesure

Nous avons vu que nous avons accès uniquement à 3 des 4 variables de la matrice de confusion. Cette limitation nous empêche d'utiliser certains indicateurs connus, tel que la courbe ROC[9].

Cependant, certains indicateurs composites nous sont accessibles, tel que la précision, le rappel et la F-mesure[9]. Le premier est aussi appelé valeur prédictive positive (VPP). Il se définit comme $\frac{VP}{VP+FP}$ et représente la proportion de bons résultats parmi les résultats fournis. Le deuxième est aussi connu sous le nom de sensibilité (*recall* en anglais) : $\frac{VP}{VP+FN}$. C'est la proportion de résultats trouvés dans l'ensemble des résultats à trouver.

Le troisième indicateur d'intérêt est la F-mesure (aussi retrouvé dans la littérature sous le nom de F₁-score. Il est souvent utilisé dans l'étude de classification, que se soit

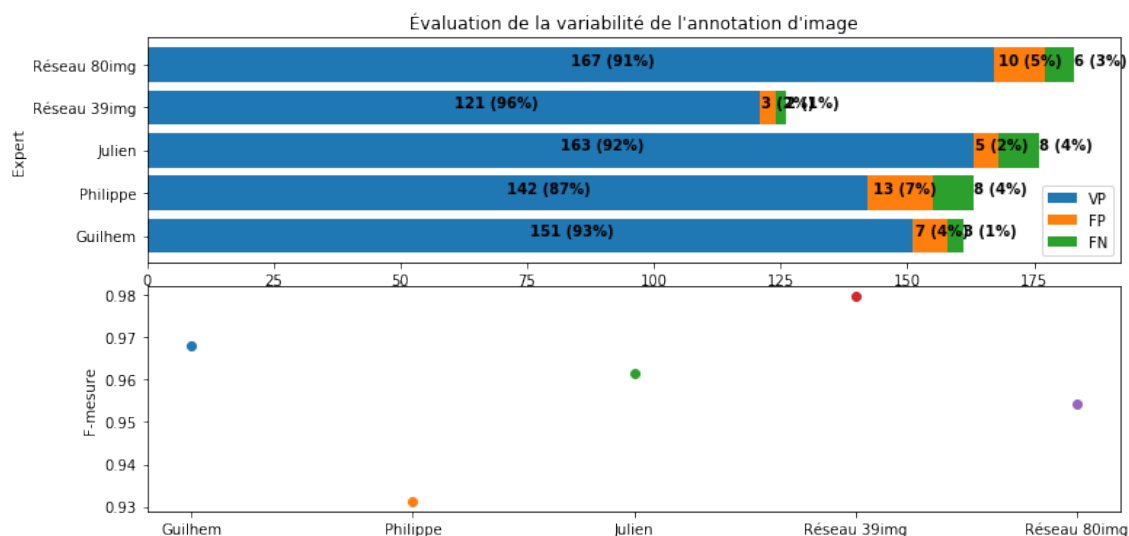


FIGURE 2 – Évaluation de la reproductibilité des annotateurs. Le réseau 39img a été entraîné sur les images où les annotateurs humains étaient d'accord. Le réseau 80img a été entraîné avec toutes les images annotées.

pour des méthodes de *machine learning* ou de *deep learning*. C'est la moyenne harmonique de la précision et du rappel, et s'exprime de la manière suivante : $2 \frac{P \cdot R}{P + R}$. C'est une mesure de convergence qui permet d'avoir une vue synthétique sur la précision et le rappel.

Une fois l'indicateur de performance choisi, on va s'intéresser à la « vérité terrain » des images de manguiers sur pied. Nous avons étudié en premier lieu la reproductibilité des différents experts.

3.1.2 Reproductibilité

Nous avons comparé les résultats des annotateurs humains mais aussi ceux obtenus par les réseaux de neurones (l'ensemble des deux catégories étant nommés « les annotateurs »). Ils ont eu 40 images et leur miroir à annoter. La Figure 2 représente les résultats obtenus et se divise en deux sous-parties. On retrouve, pour chaque annotateur, le nombre de VP, FP et FN en haut et la F-mesure obtenue en dessous.

Au niveau des annotateurs humains, on peut voir une assez bonne reproductibilité. Ceux avec les meilleures F-mesures sont Guilhem ($F_1 = 0,9679$) et Julien ($F_1 = 0,9614$). La variation entre le meilleur annotateur humain et le moins bon met en évidence une variation de 5%. On peut voir apparaître une part de subjectivité de la vérité terrain car les différents experts ont présenté des différences de jugement. Les mangues ayant créé le plus de variation sont celles situées dans des zones d'ombre (peu de luminosité) et celles constituant des grappes, dont la distinction entre les différents fruits s'avère souvent délicate.

Nous allons maintenant regarder plus en détail la reproductibilité des deux réseaux (résultats toujours disponibles en Figure 2). On peut constater que le réseau 39img se montre légèrement plus robuste que le meilleur des annotateurs humains, mais avec un nombre de mangues trouvées bien inférieur à tout les autres annotateurs. Le réseau 80img se comporte comme un humain, c'est-à-dire qu'il n'est pas meilleur qu'eux, mais ne présente pas des résultats inférieurs.

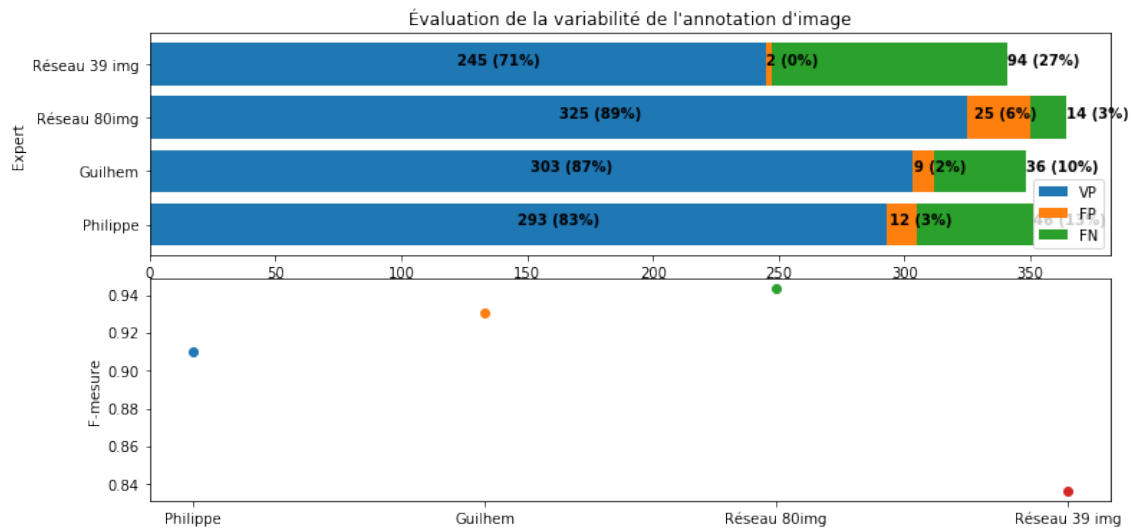


FIGURE 3 – Évaluation de la variabilité entre les différents annotateurs. La référence est l'annotateur « Julien »

3.1.3 Variabilité inter-individuelle

Pour évaluer la variabilité entre les différents annotateurs à l'aide de la F-mesure, il a fallu sélectionner ce qui représente notre vérité terrain. Nous avons vu dans la partie précédente que les annotateurs humains n'étaient pas forcément unanimes sur la vérité terrain. Les annotations retenues comme vérité terrain sont celles de Julien car il a montré une bonne reproductibilité et possède la meilleure connaissance de la mangue et du manguier parmi les experts. Les résultats de la comparaison à Julien sont disponibles en figure 3. De la même manière que dans la partie précédente, on y retrouve deux graphiques. Le premier représente le nombre de VP, de FP et de FN pour chaque annotateur. Le deuxième représente la F-mesure par expert. L'indicateur utilisé étant la F-mesure, les résultats obtenus montrent la divergence des annotations avec celles de Julien : un score bas indique une grande divergence entre les annotations comparées et les annotations de Julien.

Au niveau du nombre de VP, les extrêmes sont représentés par les réseaux de neurones : le réseau 39img possède le minimum tandis que le réseau 80img est le maximum de VP. En regardant plus précisément les résultats obtenus par le réseau 39img, on se rend compte que malgré le peu de VP détecté, le nombre de FP est le plus faible des annotateurs. Cela signifie que celui-ci a oublié beaucoup d'annotations comparé à Julien, mais qu'il est tout de même resté très proche du jeu de référence. D'un autre côté, le réseau 80img possède le plus grand nombre de FP. Cela signifie qu'il a eu tendance à trouver des mangues là où la vérité terrain n'en trouvait pas.

Au niveau de la F-mesure, on retrouve la même tendance que dans la précédente partie, les deux versions du réseau de neurones formant les observations extrêmes. Le réseau 39img possède la plus faible F-mesure, tandis que le réseau 80img possède la meilleure F-mesure. Ce dernier a donc des annotations plus proches de Julien que le reste des annotateurs.

Nous venons de voir que le choix de la vérité terrain était une décision arbitraire car il existait une variabilité entre les différents annotateurs humains. De plus, nous avons aussi

montré que le réseau était au moins aussi reproductible qu'un être humain. Il pourrait donc remplacer l'annotateur humain de part ses caractéristiques similaires.

3.2 Influence des conditions d'acquisitions sur les résultats

Les conditions d'acquisitions semblent avoir un impact sur les résultats obtenus par les réseaux de neurones. Cependant, les auteurs restent peu précis à propos des raisons ayant guidés leur choix de prise de vue. Dans le projet de *Bargoni et Underwood*[7], on apprend que les clichés sont fait « en journée ». Une autre étude nous affirme que la prise de vue de nuit (sous éclairage contrôlé) permet d'améliorer leurs résultats[10]. Ils n'apportent cependant pas d'évaluation de l'influence de ces conditions dans les résultats qu'ils obtiennent.

Ces informations sont utiles afin de fournir des recommandations pour la prise de cliché (par exemple, déterminer l'heure idéale pour la prise de vue).

Deux facteurs ont été évalués : la luminosité et la saturation des images. Le premier est surtout lié à l'ensoleillement du manguier et de la mangue. Le second permet de connaître l'importance des informations de couleur pour le réseau de neurone.

En absence de méthode d'évaluation des valeurs réelles de luminosité et de saturation de nos clichés, il a été décidé de faire varier artificiellement les valeurs de -100% à +100%. Un exemple de ces transformations est visible en Annexe C.

3.2.1 La luminosité

Nous avons alors observé l'évolution de la F-mesure en fonction de cette variation. Les résultats de l'expérimentation sont disponibles en Figure 4 et les 21 états testés sont visibles en Figure 10a (Annexe C).

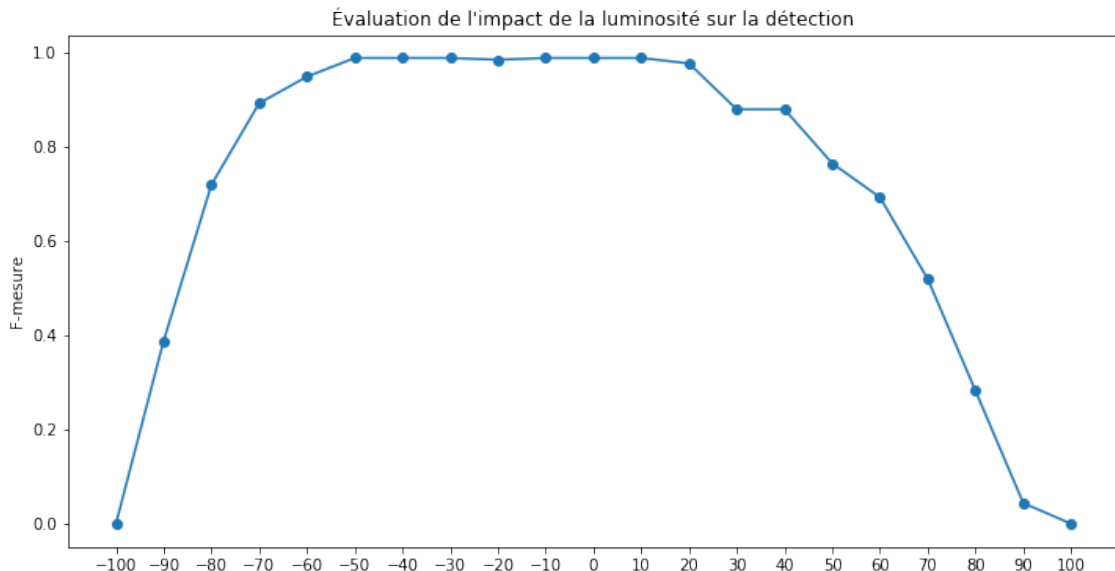


FIGURE 4 – Évaluation de l'influence de la luminosité sur la F-mesure

On peut voir l'existence d'un plateau entre -50 et +20. Dans cette plage de luminosité, la F-mesure de notre réseau change très peu. Ce résultat est intéressant car il montre que

notre réseau est peu sensible à un certain écart de luminosité. On peut aussi voir que la décroissance est plus rapide dans l'intervalle $[-100, -50[$ que dans l'intervalle $] +20, +100]$. Cela signifie que celui-ci sera plus efficace sur un objet trop peu illuminé que sur un objet trop illuminé.

Au niveau de sa variation globale, on peut voir que ce facteur a un impact important. En effet, la F-mesure va de 0 à un maximum proche de 1 lorsqu'on fait varier ce paramètre.

3.2.2 La saturation

Les résultats de l'expérimentation sont disponibles en Figure 5 et les états testés sont aussi disponibles en Figure 10b (Annexe C).

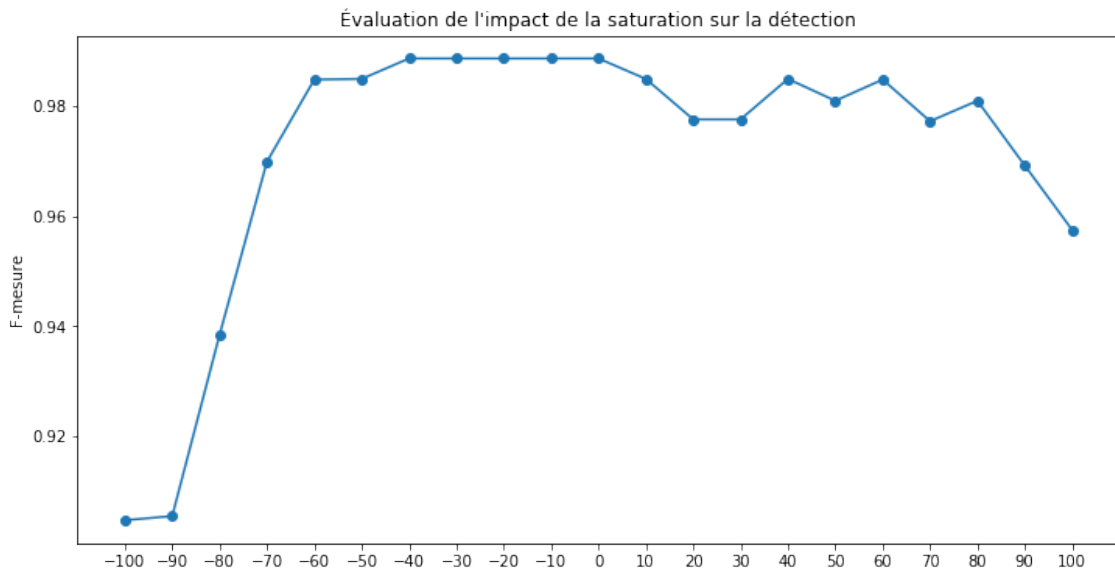


FIGURE 5 – Évaluation de l'influence de la saturation sur la F-mesure

On peut observer un plateau sur l'intervalle $[-50, 0]$, similaire à la luminosité. Cependant le reste de la courbe ne possède pas le même type de tendance régulière. De plus, la variation maximale est de 0.07 point, ce qui nous met dans une configuration différente de celle de la luminosité. En effet, une variation aussi faible pourrait être liée à un cas erratique et ne pas refléter une tendance générale. Cela montre que le réseau est peu sensible au variation de saturation, même si la différence est plus sensible si l'on va vers -100 que vers +100.

Cette expérience permet aussi de voir comment le réseau peut traiter des photos sans information de couleur alors que son jeu d'entraînement était coloré. En effet, lorsque la la photo est totalement désaturée (stade -100), on obtient une photo en noir et blanc (*cf sous-section C*). On peut constater que la F-mesure est certes plus basse que dans le cas où la saturation est réglée au niveau standard, mais celle-ci s'élève tout de même à 0.915. On en déduit que dans notre cas, la couleur ne semble pas être un facteur important pour la détection de mangue sur les photos.

Nous venons de voir l'impact qu'ont deux paramètres des conditions d'acquisition, et qu'un seul des deux avait un impact significatif : la luminosité. Nous avons aussi vu que celui-ci était peu variable sur une plage assez étendue, signe de la robustesse du réseau de neurones et donnant des pistes pour des recommandations pour des prises de vue.

L'évaluation de l'impact de la saturation nous a permis de voir sa faible importance et la réponse du réseau de neurone sur des images sans information de couleur.

3.3 Impact du paramétrage du réseau sur les résultats

Maintenant nous allons nous attarder sur le rôle du paramétrage. En effet le réseau permet de personnaliser de nombreuses variables afin d'améliorer les résultats obtenus. Nous avons décidé d'évaluer le rôle de 3 paramètres qui nous semblaient majeurs : le nombre d'itérations, le seuil de sélection et le seuil de Non-Maximum Suppression (NMS).

3.3.1 nombre d'itérations

Lors de l'apprentissage du réseau de neurone on peut décider du nombre d'itération d'apprentissage. Le nombre d'itération correspond au nombre de « tours » effectué par le réseau afin de minimiser sa fonction de perte. Nous avons observé les variation de la F-mesure selon le nombre d'itérations, et les résultats sont disponibles en Figure 6.

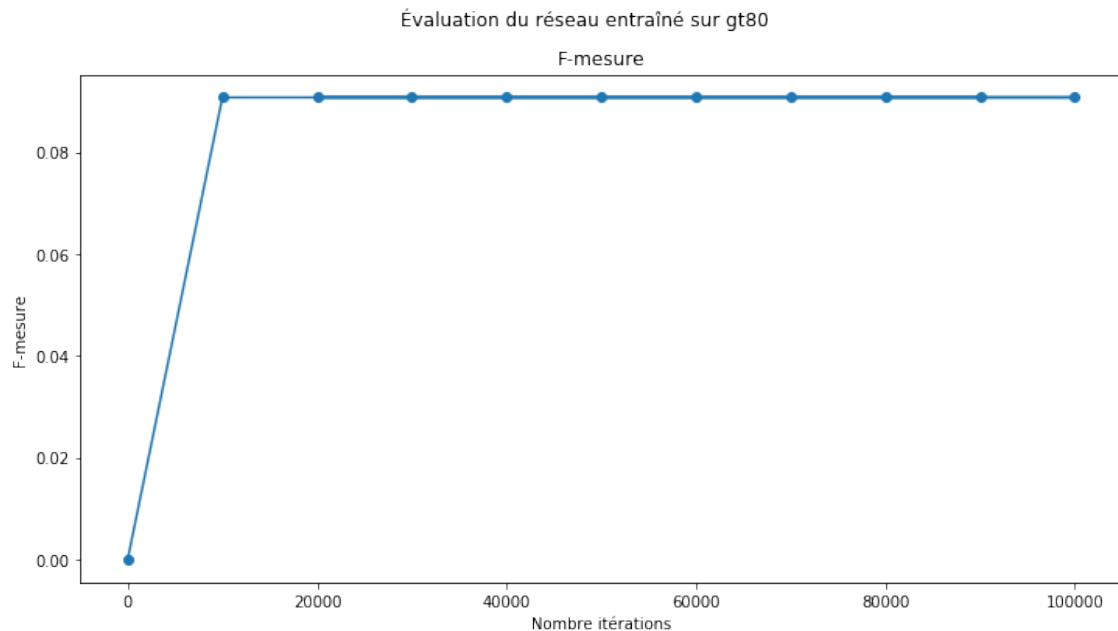


FIGURE 6 – Modification de la F-mesure en fonction du nombre d'itérations effectuées

On peut constater que la F-mesure reste stable dès que le nombre d'itération est supérieur à 0. Ceci nous apporte une information importante pour les futurs entraînements. L'entraînement étant la phase la plus coûteuse en temps, il est donc intéressant de pouvoir estimer à quel moment on peut stopper l'apprentissage.

Dans notre cas, il est intéressant de voir que l'on peut s'arrêter à partir de 10000 itérations. Cela montre que nos entraînements peuvent maintenant s'arrêter à partir de ce seuil là, il ne présente pas d'intérêt de continuer entraînement. Dans notre cas, passer de 100000 à 10000 itérations permet de gagner environ 18 heures (19 heures 40 minutes contre 2 heures).

3.3.2 Seuil de sélection

Lors de ses prédictions de boîtes englobantes, le réseau de neurone fournit un score de confiance. Celui-ci retranscrit la pertinence de cette prédiction. Une valeur-seuil, paramétrable, permet de décider quel est le score minimum nécessaire pour que la prédiction soit conservée. Nous avons fait varier cet indicateur pour voir à quel moment le réseau donnait la plus grande similarité de résultats avec la référence. L'évolution de la F-mesure est disponible en Figure 7.

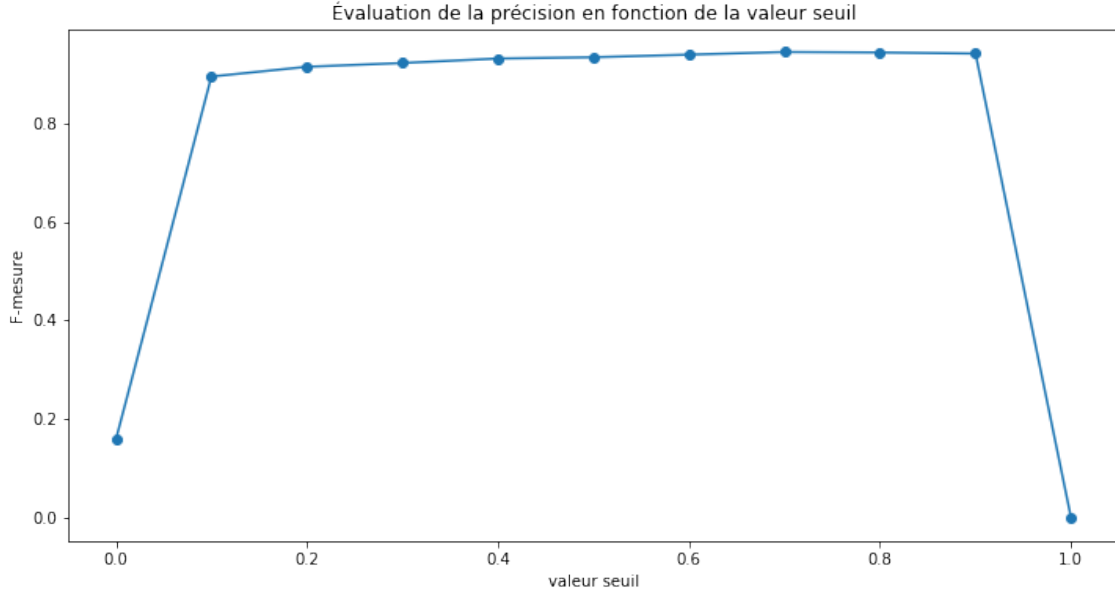


FIGURE 7 – Variation de la F-mesure en fonction du seuil de détection

En excluant les deux valeurs extrêmes (0 et 1.0) on constate que la valeur seuil impacte peu la F-mesure, avec un maximum à 0.9443 lorsque le seuil vaut 0,7. Ce type de paramétrage est plutôt utile pour le paramétrage fin du réseau. La valeur par défaut étant de 0.7, elle paraît bien adaptée à notre jeu de données car correspond au maximum de notre expérimentation.

3.3.3 Non-Maximum Suppression

Le NMS permet de configurer la gestion des occlusions. En effet, lors de ses prédictions, le réseau de neurones propose de multiples emplacements d'objets. Ceux-ci peuvent parfois entourer le même objet et se recouper. C'est à ce niveau là que le NMS intervient. Il définit la surface maximum de recouvrement entre deux objets (calcul de l'intersection sur l'union : $\frac{|S_A \cap S_B|}{|S_A \cup S_B|} \leq nms$). Si les deux boîtes englobantes se recoupent plus que ce seuil, elles sont considérées comme représentant le même objet et une seule est conservée. La variation de la F-mesure est présentée en Figure 8.

On peut voir une avec un maximum de 0,9562 lorsque le NMS vaut 0,2 et un minimum valant 0,2188 (pour $nms = 1,0$). Ce paramètre possède une variabilité importante (77% d'incertitude relative). Il est donc important de le régler au mieux afin d'optimiser les résultats obtenus par le réseau de neurone.

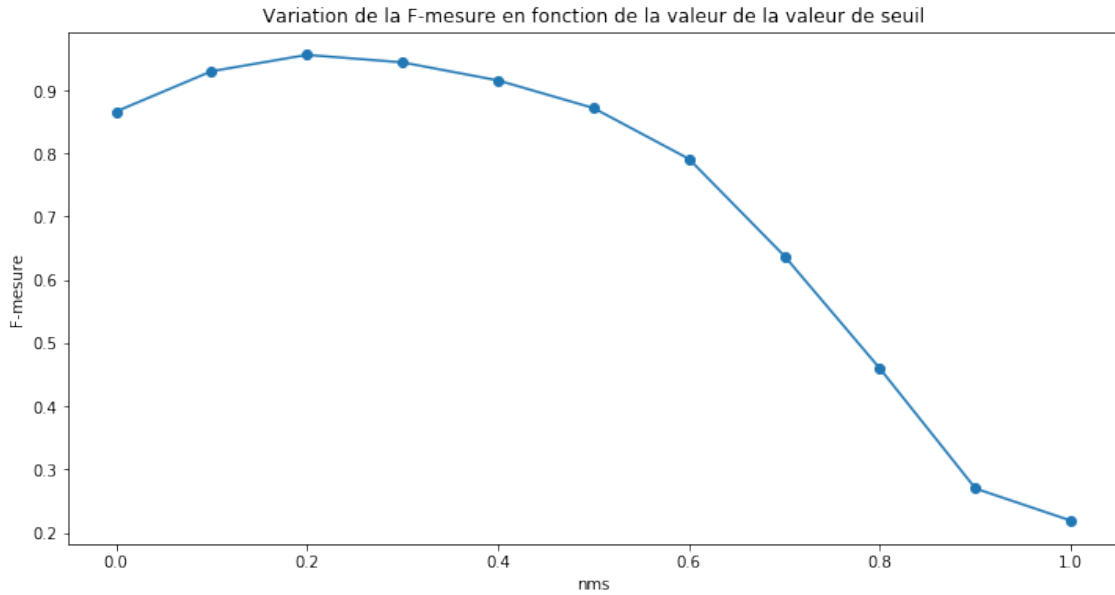


FIGURE 8 – Variation de la F-mesure en fonction de la modification du NMS

Nous avons vu que les différents paramètres ne possédaient pas la même influence sur les résultats. Le premier paramètre vérifié, le nombre d'itérations, n'a pas montré d'intérêt particulier. Le deuxième, le seuil de sélection, a montré un effet faible mais existant sur le résultat. Le dernier, le NMS, est le plus important des trois. C'est celui-ci qui faut plus spécifiquement étudier car un mauvais réglage peut amener à des résultats beaucoup moins pertinents.

4 Conclusion

Dans ce rapport, nous avons vu la spécialisation et une évaluation d'un réseau de neurones : le modèle *py-FASTER-RCNN* dans l'environnement *Caffe*.

Dans la première partie, nous nous sommes attardés sur les projets déjà existants ainsi que sur le concept de spécialisation d'un réseau de neurones. Celui-ci est un élément clé de la plupart des projets de *deep learning* actuels, permettant un entraînement et un déploiement beaucoup plus rapide et améliorant les résultats.

Dans un deuxième temps, nous avons évalué l'effet de différents facteurs sur les résultats obtenus par le réseau de neurone. Nous avons vu que la « vérité terrain » était assez subjective, avec une variabilité d'annotation entre les différents experts humains. A ce même niveau, nous avons vu alors que notre réseau de neurone était plus reproductible que la plupart des annotateurs humains. Le réseau 80img s'avérait présenter des résultats beaucoup plus proche de la référence (Julien) que la plupart des autres évaluateurs lorsqu'il était entraîné sur toutes les images annotées. Le réseau de neurones montre des résultats comparables aux annotateurs humains, cependant celui-ci traite les images environ 1200 fois plus vite (3 minutes par image³ contre 300 millisecondes pour le réseau de neurones). Ce gain de temps permet d'imaginer des analyses plus régulières sur les exploitations, et donc un suivi plus fin.

3. La durée moyenne d'annotation par les humains a été obtenue par la division du nombre d'image par le temps passé à annoter la totalité des images.

Cependant, l'indicateur sélectionné, la F-mesure, montre la convergence de deux résultats. On pourrait alors réfléchir à différentes pistes pour améliorer notre vérité terrain. Une des pistes à envisager serait de pratiquer une moyenne pondérée des résultats des différents experts humains afin d'obtenir un consensus servant de référence. On pourrait alors étudier l'efficacité du réseau de neurones lorsqu'il est entraîné sur ces données là.

Ensuite, nous avons abordé les conditions d'acquisitions et les paramétrages du réseau. Il en ressort que les paramètres les plus importants sont la luminosité ainsi que la valeur du NMS. Le premier pourrait mener à des recommandations sur les prises de vue si on arrive à faire un lien entre luminosité sur la prise de vue et la luminosité sur l'image. Cependant, cette relation semble difficile à obtenir.

Lors de ces analyses, nous avons considéré les mangues comme une unique classe. Cependant, pour les besoins d'évaluation du tonnage, il serait utile de pouvoir différencier les mangues mures des mangues non mures. Deux méthodes ont déjà été envisagées afin de répondre à cette problématique. La première consisterait à enrichir notre jeu d'entraînement, et apprendre au réseau à différencier deux classes : des mangues mures et des mangues non-mures. La deuxième méthode consisterait à récupérer le résultat du réseau actuel, pour ensuite faire de l'analyse de l'histogramme dans les boîtes englobantes à la recherche d'une signature colorimétrique spécifique des mangue mures.

Nous avons eu recours à une référence externe (dans notre cas, Julien) afin de pouvoir comparer nos résultats. Le réseau nous fournit pour chaque résultat un score, que nous avons évoqué dans le cadre du seuil et du NMS. Un des objectifs à atteindre serait de pouvoir comprendre la signification de celui-ci pour pouvoir éventuellement s'en servir pour évaluer le réseau lors de ses prédictions.

De plus, à partir de ce score de précision, le réseau calcule une mean Average Precision (mAP) : c'est la moyenne des précisions moyennes obtenues pour la totalité des images traitées. Mieux comprendre son évolution permettrait de mettre en place des stratégies d'évaluation de prédiction plus efficaces.

Annexes

A Exemple de resultat de segmentation d'image

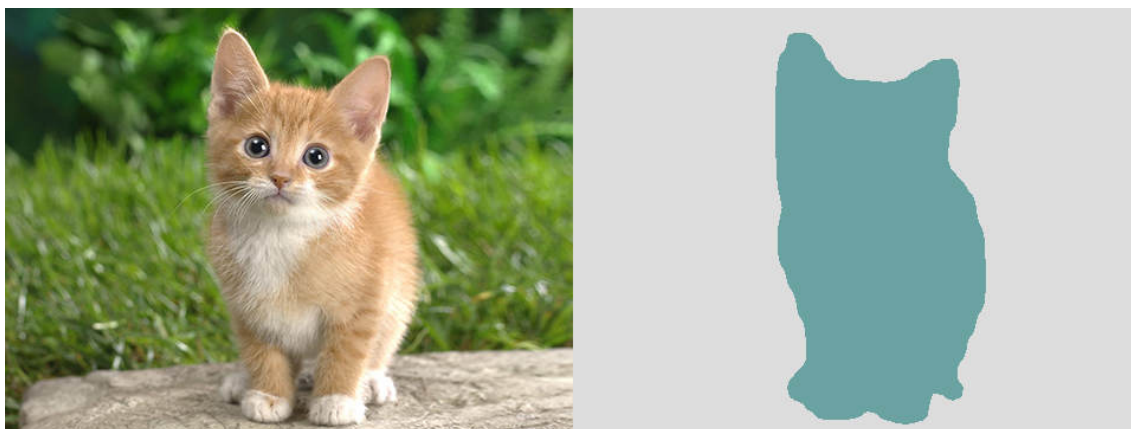


FIGURE 9 – Exemple de segmentation d'image obtenue grâce au modèle *Fully Convolutional Neural Network*[5]. La photo originale se situe à gauche, tandis que le résultat se trouve sur la partie droite. On retrouve en bleu foncé la classe « chat » et en bleu clair la classe « background ». Extrait de <http://nicolovaligi.com/convert-deep-learning-model-caffe-keras.html>.

B Exemple de fichier d'annotations utilisé par notre réseau de neurones

```
1 <?xml version="1.0" ?>
2 <annotation>
3   <filename>75715583495548</filename>
4   <size>
5     <width>500</width>
6     <height>500</height>
7     <depth>3</depth>
8   </size>
9   <object>
10    <name>mango</name>
11    <pose>Unspecified</pose>
12    <truncated>0</truncated>
13    <difficult>0</difficult>
14    <bndbox>
15      <xmin>116</xmin>
16      <ymin>78</ymin>
17      <xmax>212</xmax>
18      <ymax>188</ymax>
19    </bndbox>
20  </object>
21 </annotation>
```

C Exemple des modifications des conditions d'acquisition

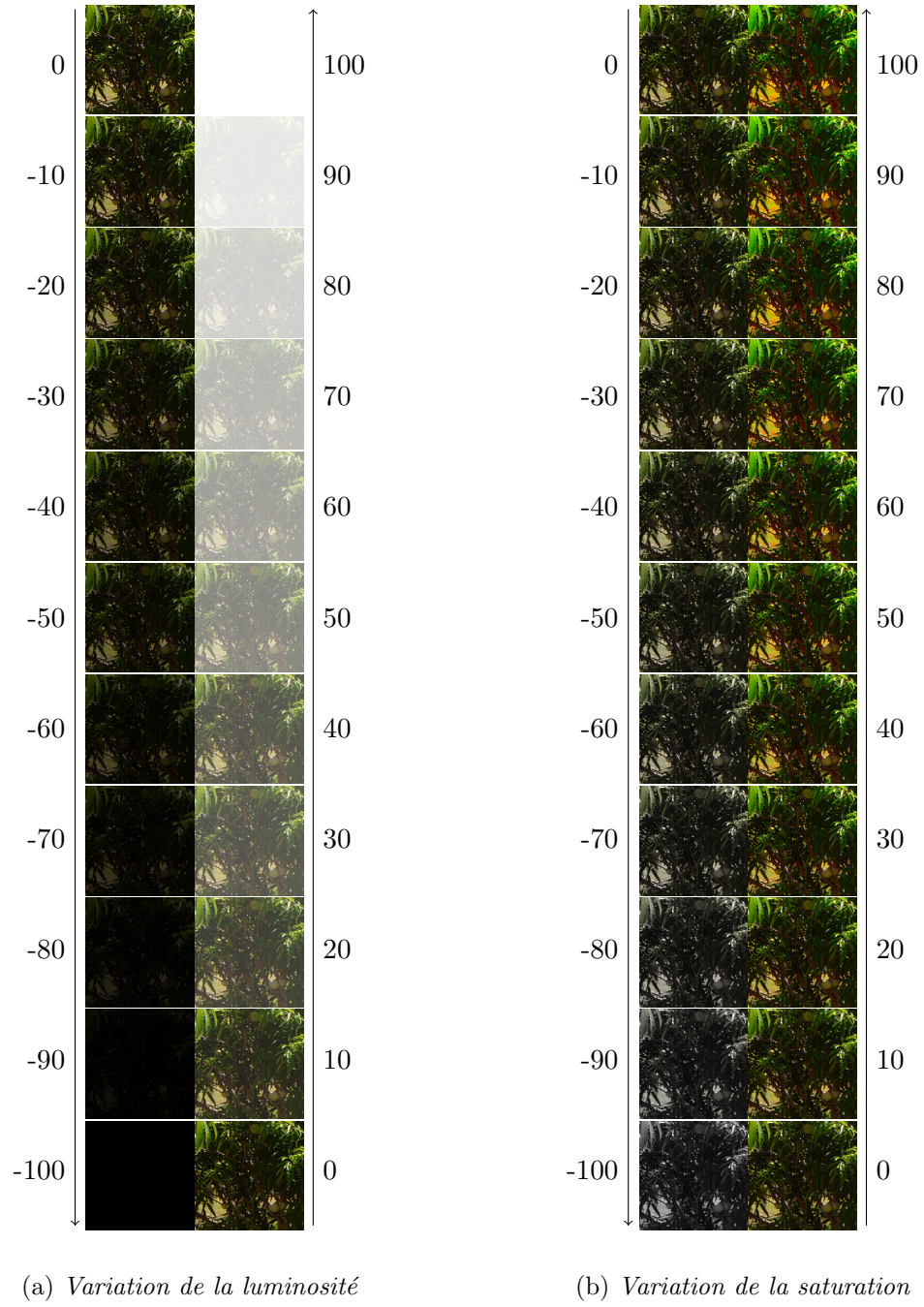


FIGURE 10 – *Variation des conditions de saturation et de luminosité : exemple sur une image*

Références

- [1] A. J. Challinor, J. Watson, D. B. Lobell, S. M. Howden, D. R. Smith, and N. Chhetri, “A meta-analysis of crop yield under climate change and adaptation,” *Nature Climate Change*, vol. 4, pp. 287–291, Mar. 2014.
- [2] H. Vannière, C. Didier, J.-Y. Rey, T. M. Diallo, S. K ?ita, and M. Sangar ?, “The mango in French-speaking West Africa : cropping systems and agronomical practices,” *Fruits*, vol. 62, pp. 187–201, May 2007.
- [3] A. Payne, K. Walsh, P. Subedi, and D. Jarvis, “Estimating mango crop yield using image analysis using fruit at ‘stone hardening’ stage and night time imaging,” *Comput. Electron. Agric.*, vol. 100, pp. 160–167, Jan. 2014.
- [4] R. Linker, “A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation,” *Precision Agriculture*, vol. 18, pp. 59–75, Feb. 2017.
- [5] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
- [6] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, “DeepFruits : A Fruit Detection System Using Deep Neural Networks,” *Sensors*, vol. 16, p. 1222, Aug. 2016.
- [7] S. Bargoti and J. Underwood, “Deep Fruit Detection in Orchards,” *arXiv preprint arXiv :1610.03677*, 2016.
- [8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe : Convolutional architecture for fast feature embedding,” *arXiv preprint arXiv :1408.5093*, 2014.
- [9] D. M. Powers, “Evaluation : from precision, recall and F-measure to ROC, informedness, markedness and correlation,” 2011.
- [10] W. S. Qureshi, A. Payne, K. B. Walsh, R. Linker, O. Cohen, and M. N. Dailey, “Machine vision for counting fruit on mango tree canopies,” *Precision Agriculture*, vol. 18, pp. 224–244, Apr. 2017.

Glossaire

entraînement Dans un réseau de neurone, c'est la phase où l'on apprend au réseau les instances qu'il doit classifier et à quelle classe elles appartiennent. 4, 5, 11, 14

moyenne harmonique Inverse de la moyenne arithmétique des inverses des valeurs de base. 7

Acronymes

FN faux négatif. 6–8

FP faux positif. 6–8

GPU Graphics Processing Unit. 5

mAP mean Average Precision. 14

NMS Non-Maximum Suppression. 2, 11–14

VP vrai positif. 6–8

VPP valeur prédictive positive. 6