

Descripteurs profonds et classification à une classe avec des données non supervisées pour la détection des mauvaises herbes.

Mamadou Dian BAH¹, Adel HAFIANE², Raphael CANALS¹

¹Laboratoire PRISME

Univ. Orleans, PRISME, EA 4229, F45072, Orleans, France

²Laboratoire PRISME

INSA-CVL, PRISME, EA 4229, F18020, Bourges, France

mamadou-dian.bah@univ-orleans.fr, adel.hafiane@insa-cvl.fr

raphael.canals@univ-orleans.fr

Résumé – Aujourd’hui, la détection automatique des mauvaises herbes est un grand défi pour l’agriculture de précision. L’apprentissage automatique est l’une des approches utilisées pour la détection des mauvaises herbes. Le principal point commun entre les algorithmes d’apprentissage automatique est le besoin de données d’entraînement. Dans cet article, nous proposons d’utiliser des caractéristiques profondes et une classification à une classe sur des données non supervisées pour la détection des mauvaises herbes dans les images de drones. Les résultats montrent que la classification à une classe peut être comparable à la littérature et également à un modèle d’apprentissage profond entraîné avec un étiquetage supervisé des données d’entraînement. Les résultats obtenus sur les jeux de données de test peuvent atteindre 90 %.

Abstract – Today, automatic weed detection is one of the most challenging problems for precision agriculture. Machine learning is one of the approaches used for weed detection. The main commonality among machine learning algorithms is the need for training data. In this paper, we propose to use deep features and one-class classification on unsupervised data for weed detection in drone images. The results show that the one-class classification can be comparable to the literature and also to a deep learning model trained with supervised labeling of the training data. The results obtained on the test datasets can reach 90%.

1 Introduction

L’apprentissage automatique est l’une des approches utilisées pour la détection des mauvaises herbes. Bakhshipour et Jafari [1] ont évalué la détection des mauvaises herbes avec une machine à vecteur de support (SVM) et des réseaux de neurones artificiels dans quatre espèces de mauvaises herbes communes dans les champs de betteraves à sucre en utilisant des caractéristiques de forme. Dans [2], une procédure semi-automatique d’analyse d’image basée sur les objets (OBIA) a été développée avec des forêts aléatoires (RF) combinées à des techniques de sélection de caractéristiques pour classer le sol, les mauvaises herbes et le maïs. Avec ces articles, nous pouvons remarquer que les caractéristiques sélectionnées changent en général d’un type de culture à un autre ou d’un type de données à un autre. C’est pourquoi certains auteurs ont proposé d’effectuer un apprentissage profond [3]. Dans [4], les auteurs ont entraîné une architecture profonde d’encodeur-décodeur convolutif pour un étiquetage sémantique pixellisé robuste (SegNet) afin d’identifier 3 classes : arrière-plan, culture et mauvaise herbe. Milioto et al. [5] ont fourni une classification précise des mauvaises herbes dans des champs de betteraves à sucre avec des robots agricoles mobiles. Bah et al. [6] ont appliqué AlexNet pour la détection des mauvaises herbes dans différents champs de

cultures tels que la betterave, les épinards et les haricots dans des images de drones.

Le principal point commun entre les algorithmes d’apprentissage automatique est la nécessité de disposer de données d’entraînement. Mais la création de grands ensembles de données agricoles avec des annotations au niveau du pixel est une tâche extrêmement longue. Peu de tentatives ont été faites pour développer des systèmes entièrement automatiques pour l’apprentissage et l’identification des mauvaises herbes dans les champs agricoles. Di Cicco et al. [7] ont suggéré l’utilisation de jeux de données d’entraînement synthétiques. Cependant, cette technique nécessite une modélisation précise en termes de texture, de modèles 3D et de conditions de lumière. Actuellement, la plupart des méthodes utilisées pour la collecte non supervisée de données détectent d’abord les rangées de cultures, puis étiquettent les plantes à l’intérieur des rangées comme des cultures et celles entre les rangées comme des mauvaises herbes (inter-row weed) [8–10]. Mais ces méthodes dépendent fortement de la présence de mauvaises herbes dans l’inter-rang. Par conséquent, les données étiquetées seront très déséquilibrées dans les cas où le champ a moins de mauvaises herbes entre les rangs de culture. Cela peut réduire l’efficacité des modèles formés et favoriser le surajustement. Dans [11], un SVM à une classe est utilisé pour la classification des mauvaises herbes. Pour leur expérience, 200

images de mauvaises herbes sont utilisées. Comparé au SVM binaire avec un noyau de fonction de base radiale (RBF), les résultats montrent que le SVM à une classe a amélioré la classification des mauvaises herbes.

Dans cet article, nous proposons d'utiliser des caractéristiques profondes et une classification à une classe sur des données non supervisées pour la détection des mauvaises herbes dans les images de drones. Par rapport à notre travail précédent [10], seules les plantes dans les rangées sont utilisées pour l'entraînement du classificateur à une classe. Des études antérieures ont montré la complexité de la sélection des meilleures caractéristiques pour la classification des mauvaises herbes et des cultures en raison de la grande variabilité des plantes et des champs agricoles. Dans cet article, afin d'extraire des caractéristiques robustes, nous avons utilisé des caractéristiques profondes car elles ont prouvé leur efficacité dans plusieurs études [12, 13].

Cet article est divisé en quatre parties. Dans la section 2, nous présentons la méthode proposée. Les résultats expérimentaux sont commentés et discutés dans la section 3. La section 4 conclut l'article.

1.1 Méthode proposée

La méthode proposée consiste à collecter des jeux de données de cultures de manière non supervisée et à utiliser un classificateur à une classe pour entraîner un modèle. Pour cela, nous avons appliqué la méthode de détection des rangs de culture proposée dans [14]. La transformée de Hough est appliquée sur le résultat d'une squeletisation pour détecter les rangs. Une fois les rangs de culture identifiés, l'image est segmentée en régions ou superpixels. Tous les superpixels dans les rangées de cultures sont collectés pour former un ensemble de données d'entraînement. Ensuite, l'extraction de caractéristiques est effectuée sur cet ensemble de données d'entraînement. Le modèle SVM (Support Vector Machine) à une classe a été choisi pour entraîner le modèle. Le SVM à une classe est un algorithme qui apprend une fonction de décision pour la détection de la nouveauté, en classant les nouvelles données comme similaires ou différentes de l'ensemble de données d'apprentissage [15]. Cette fonction est utile pour la détection des anomalies, car c'est la rareté des exemples d'entraînement qui définit les anomalies. Conceptuellement, cet algorithme tente de construire un hyperplan séparant l'aberration du jeu de données d'entraînement.

1.2 Création des régions

En observant les champs agricoles, il est facile de voir que les cultures sont espacées à des distances presque constantes et alignées dans la direction du semis, et que la présence de mauvaises herbes perturbe cet alignement. A partir de cette information, nous avons décidé d'utiliser la ligne de partage des eaux pour segmenter les plantes en superpixels en fonction de l'irrégularité de l'alignement des plantes.

La ligne de partage des eaux (watershed) est une méthode de

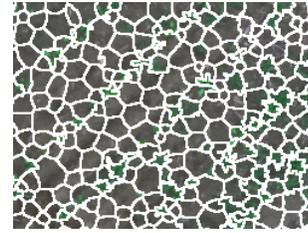


FIGURE 1 – Exemple de segmentation avec le watershed.

segmentation d'image qui considère l'image en niveaux de gris comme une surface topographique, plus la valeur d'un pixel est grande, plus le pixel est élevé (montagnes, ...) et inversement les plus petits représentent des dépressions dont la profondeur dépend de la valeur du pixel. En supposant que l'on décide d'inonder le relief en remplissant d'eau toutes les cuvettes, les barrières qui seront construites pour éviter que les eaux des cuvettes ne se rencontrent représentent les lignes de partage des eaux. Dans ce cas, les pixels blancs représentent les sommets et les noirs les creux. Avec les lignes de partage des eaux, les rangs de culture sont segmentés en superpixels en utilisant les creux présents au niveau des contours (Fig.1). Ainsi, les superpixels résultants ont soit une seule plante, soit un groupe de plantes (*crop – crop*, *crop – weed* ou *weed – weed*) et leurs formes dépendent de la disposition de ces plantes. Les superpixels avec un groupe de plantes *crop – weed* sont étiquetés comme des mauvaises herbes car, pour les agriculteurs, il est préférable d'avoir une surdétection que de manquer une mauvaise herbe. De plus, en raison de la résolution spatiale de l'imagerie par drone, il est parfois préférable de se concentrer sur un groupe de plantes pour localiser une mauvaise herbe que sur une seule plante, même pour un expert.

1.3 Extraction de descripteurs profonds

Les descripteurs profonds ou deep features correspondent à des descripteurs extraits à différentes couches d'un réseau de neurones convolutifs (CNN) [3]. L'apprentissage par transfert est l'une des approches les plus populaires de l'apprentissage profond ; elle consiste à utiliser des modèles pré-entraînés sur un jeu de données tel qu'ImageNet comme point de départ pour résoudre un autre problème de vision par ordinateur. L'extraction de deep features s'inspire de l'idée que les humains peuvent intuitivement utiliser leur expérience acquise précédemment pour définir et résoudre de nouveaux problèmes. Dans cet article, les descripteurs sont extraits du réseau résiduel pré-entraîné avec 18 couches (ResNet18). Cette architecture de réseau a été introduite en 2015 [16]. Elle a remporté le défi de reconnaissance d'images sur la base d'images ImageNet 2015 avec sa version de 152 couches. La Fig.2 présente l'architecture ResNet18 et l'endroit où les deep features sont extraits. Nous avons extrait des features à différents niveaux du réseau pour évaluer l'avantage d'utiliser des couches trop profondes ou peu profondes. Chaque descripteur a une dimension égale au nombre de filtres présents sur la couche de convolution, où chaque valeur correspond à la moyenne de la sortie de chaque filtre. Ainsi,

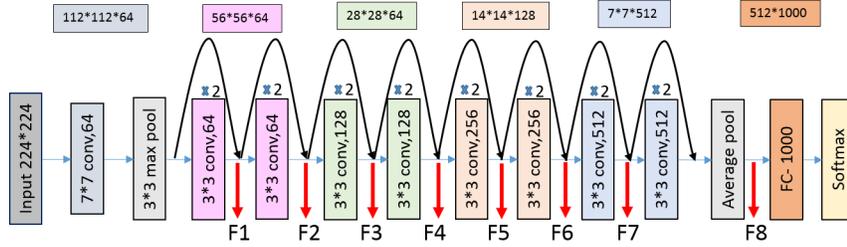


FIGURE 2 – Architecture de ResNet18 avec extraction de caractéristiques. Les flèches rouges montrent où les caractéristiques profondes ont été extraites. Les rectangles horizontaux montrent la taille à la sortie des couches qui ont la même couleur.

les dimensions de F1, F2, F3, F4, F5, F6, F7 et F8 sont respectivement de 64, 64, 128, 128, 256, 256, 512 et 512.

2 Expériences et résultats

Les expériences ont été menées sur le jeu de données public publié dans [4]. Les images utilisées ont été acquises dans un champ de betteraves à une altitude de 10 m avec une caméra multispectrale RedEdge-M. Le jeu de données contient cinq images orthomosaïques différentes (000, 001, 002, 003 et 004) et leurs vérités-terrain correspondantes. Chaque image orthomosaïque représente une partie du champ avec un certain chevauchement avec les autres. Il y a cinq bandes brutes par image : les trois bandes visibles (R, G, B) et deux autres qui sont la bande Red-Edge et le proche infrarouge (NIR). Deux types d'images à bande composite sont disponibles, notamment des images RGB et des images infrarouges couleur (CIR) par empilement des canaux R, G, B pour une image RGB et R, G, et NIR pour une image CIR. Dans la figure 3 nous présentons un échantillon d'image RGB.

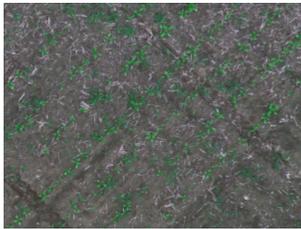


FIGURE 3 – Exemple d'échantillon.

Nous avons collecté deux ensembles de données différents pour chaque image à bande composite (RGB et CIR). Le premier jeu de données a été créé de manière non supervisée (US000, US001, US002, US003 et US004) et le second a été créé en uti-



FIGURE 4 – Exemple d'échantillons collectés. Les rectangles verts et rouges sont respectivement des échantillons de culture et de mauvaise herbe.

TABLE 1 – Nombre d'échantillons collectés pour l'entraînement et le test. Sup. et Unsup. désignent respectivement les données collectées de manière supervisée et non supervisée.

Image	000	001	002	003	004
Unsup. Crop	742	1791	1501	2260	680
Sup. Crop	4317	1972	7216	5620	1735
Sup. Weed	7144	5538	2888	2461	867

lisant la vérité terrain disponible avec les images orthomosaïques (S000, S001, S002, S003 et S004). Dans la suite du document, les ensembles de données non supervisées et supervisées signifient respectivement ensemble de données étiquetées de manière non supervisée et supervisée.

Seuls les superpixels dans les rangées de cultures ont été sélectionnés pour construire le jeu de données non supervisé. Un échantillon est considéré comme une mauvaise herbe si les superpixels contiennent au moins 10 % de pixels étiquetés comme mauvaises herbes dans la vérité terrain. En effet, pour les agriculteurs, il est préférable d'avoir une surdétection que de manquer une mauvaise herbe. En outre, en raison de la résolution spatiale des images de drone, il est parfois préférable de se concentrer sur un groupe de plantes pour localiser une mauvaise herbe plutôt que sur une seule plante, même pour un expert. Seuls les échantillons qui ne contiennent aucun pixel de mauvaise herbe sont sélectionnés pour la classe de culture. Le nombre total d'échantillons est présenté dans le tableau 1.

Nous avons combiné les huit descripteurs pour obtenir un nouvel descripteur de 1920 éléments appelé F1-8. La méthode d'analyse en composantes principales a été calculée pour réduire la taille de F1-8 à 200 composantes (PCA200). Le nombre de composantes principales retenues représente environ 98% de l'information. Nous avons également construit un nouveau jeu de données d'entraînement en rassemblant tous les jeux de données non supervisés (US0-4). En résumé, le nombre de caractéristiques (Nfeatures) est de 10 et le nombre de jeux de données d'apprentissage (Ntrain) est de 6. Les jeux de données supervisés sont utilisés pour les tests. Comme pour les données non supervisées, un nouveau jeu de données de test (S0-4) a été construit avec tous les jeux de données supervisés. La fonction Matlab *fitcsvm* a été appliquée avec un noyau gaussien pour détecter les valeurs aberrantes avec le SVM et l'apprentissage à une classe. Cette fonction trouve des limites séparant les va-

leurs aberrantes du reste de l'ensemble de données en fonction d'un taux d'aberration donné. Nous supposons que 10 % des données sont aberrantes, en raison de la présence de mauvaises herbes dans les rangées dans les jeux de données non supervisés. Avec le SVM à une classe, $N_{\text{features}} \times N_{\text{train}}$ modèles ont été formés.

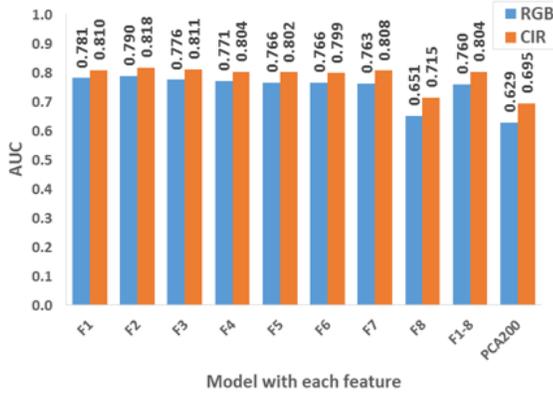


FIGURE 5 – Aire sous la courbe obtenue (AUC) avec chaque caractéristique en utilisant US0-4 comme ensemble de données d'entraînement et S0-4 comme ensemble de données de test.

L'aire sous la courbe obtenue (AUC) calculée à partir de la courbe ROC (receiver operating characteristic) a été utilisée pour évaluer les résultats. Le jeu de données US0-4 a été utilisé pour construire un modèle avec chaque fonction et S0-4 pour les tester. Dans la Fig.5, nous remarquons que F2 présente les meilleurs résultats, F8 et PCA200 ont les plus faibles. De plus, nous pouvons observer que les descripteurs les moins profonds sont plus efficaces. En général, les CNN apprennent à détecter les contours dans la première couche, puis à utiliser les contours pour détecter des formes simples dans la deuxième couche, et enfin à utiliser ces formes pour extraire des descripteurs de plus haut niveau dans les couches plus profondes. Selon cette hypothèse, nous pouvons dire que les formes sont les plus appropriées pour la détection des cultures et des mauvaises herbes dans ce champ de betteraves. Nous remarquons également que les jeux de données CIR donnent de meilleurs résultats que les jeux de données RGB. Ainsi, dans les parties suivantes, seuls les jeux de données CIR ont été utilisés pour l'entraînement et le test.

Avec la fonction F2, chacun des jeux de données d'apprentissage a été évalué sur tous les jeux de données de test. Pour rappel, les jeux de données supervisés (S000, S001, S002, S003, S004 et S0-4) sont utilisés pour les tests. Le tableau 2 montre que les modèles formés et évalués dans le même domaine peuvent donner de bons résultats (valeurs en gras). De plus, le Modèle002 et le Modèle003 sont les plus efficaces sur tous les jeux de données de test, avec des AUC supérieures à 0.90 dans le jeu de données de test S000. Dans S0-4, leurs AUC sont respectivement de 0,885 et 0,873. Le modèle001 a donné le résultat le plus faible et son meilleur résultat est obtenu dans S001 avec une AUC de 0.757. Ces résultats s'expliquent par le fait que

TABLE 2 – AUC des modèles SVM à une classe et de ResNet18 sur les données supervisées.

Models	AUC sur les données supervisées					
	S000	S001	S002	S003	S004	S0-4
Model000	0.812	0.810	0.680	0.721	0.684	0.719
Model001	0.703	0.757	0.619	0.665	0.647	0.609
Model002	0.914	0.869	0.850	0.850	0.776	0.885
Model003	0.902	0.868	0.826	0.848	0.763	0.873
Model004	0.827	0.781	0.744	0.738	0.715	0.805
Model0-4	0.835	0.820	0.778	0.802	0.738	0.810
ResNet18	-	0.929	0.926	0.929	0.873	0.957

les données d'apprentissage US002 et US003 ont été collectées dans les deux images les moins infestées par les mauvaises herbes et inversement pour les données US000 et US001. L'image 004 est moins infestée que les deux dernières, son résultat peut être justifié par la présence de flou sur sa partie inférieure.

Dans [4], une architecture profonde de codeur-décodeur convolutif (SegNet) a été entraînée sur les jeux de données S000, S001, S002, S004 et testée dans S003. Les meilleurs résultats ont été obtenus en combinant 9 bandes spectrales, telles que le bleu, la couleur-infrarouge (CIR), le vert, le NDVI, le proche infrarouge, le rouge et le bord rouge. La valeur AUC pour l'arrière-plan, les cultures et les mauvaises herbes est respectivement de 0,839, 0,863 et 0,782. Afin de comparer la méthode proposée avec [4], le Model003 a été appliqué sur tous les superpixels de l'image orthomosaique 003 pour créer une carte de scores. L'AUC obtenu est égal à 0,91 et il surpasse le résultat obtenu dans [4].

Par la suite, les résultats sont comparés à ceux qui ont été obtenus avec le ResNet18 sur deux classes différentes, culture et mauvaise herbe étiquetées de manière supervisée. Le jeu de données S000 a été utilisé pour créer un modèle supervisé avec ResNet18 (ResNet18-000). Comparé aux autres données supervisées, ce jeu de données est plus équilibré. Les résultats sont présentés dans le Tableau 2. ResNet18 a été entraîné avec 100 époques (epochs), une taille de mini-batch de 10 et un taux d'entraînement initial de 0,001.

ResNet18-000 a donné un meilleur résultat que le SVM à une classe unique dans tous les jeux de données de test. Son AUC sur le jeu de données S0-4 est de 0,957. Ce résultat est supérieur à ceux obtenus avec le Model002 et le Model003. Mais comme mentionné précédemment, la création de grands ensembles de données agricoles avec des annotations au niveau du pixel est une tâche extrêmement longue et pénible.

3 Conclusion

Dans cet article, nous avons proposé d'utiliser des descripteurs profonds et un SVM à une classe sur des données non supervisées pour la détection de mauvaises herbes dans des images de drones. Les résultats ont montré qu'avec des jeux de données non supervisés collectés dans une image orthomosaique, nous sommes capables, avec un SVM à une classe,

de détecter des anomalies (mauvaises herbes) dans la même image avec une AUC allant de 0,715 (004) à 0,85 (002). L'aire sous la courbe (AUC) calculée sur tous les jeux de données de test est comprise entre 0,609 et 0,885. Comparé aux résultats obtenus dans la littérature avec SegNet dans le même jeu de données, la méthode proposée donne un meilleur résultat. Dans les perspectives nous souhaitons explorer des méthodes comme le few shot learning pour une classification efficace avec de petite base de données. Nous comptons également tester différentes valeurs du taux d'aberrations à appliquer pour la classification à classe unique.

Références

- [1] Adel Bakshpour and Abdolabbas Jafari. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Computers and Electronics in Agriculture*, 145 :153–160, feb 2018.
- [2] Junfeng Gao, Wenzhi Liao, David Nuyttens, Peter Lootens, Jürgen Vangeyte, Aleksandra Pižurica, Yong He, and Jan G. Pieters. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *International Journal of Applied Earth Observation and Geoinformation*, 67 :43–53, may 2018.
- [3] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2323, 1998.
- [4] Inkyu Sa, Marija Popović, Raghav Khanna, Zetao Chen, Philipp Lottes, Frank Liebisch, Juan Nieto, Cyrill Stachniss, Achim Walter, and Roland Siegwart. WeedMap : A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sensing*, 10(9), 2018.
- [5] Andres Milioto, Philipp Lottes, and Cyrill Stachniss. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W3 :41–48, aug 2017.
- [6] M Dian Bah, Eric Dericquebourg, Adel Hafiane, and Raphael Canals. Deep learning based classification system for identifying weeds using high-resolution UAV imagery. In *Computing Conference 2018*, 2018.
- [7] Maurilio Di Cicco, Ciro Potena, Giorgio Grisetti, and Alberto Pretto. Automatic model based dataset generation for fast and accurate crop and weeds detection. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5188–5195. IEEE, sep 2017.
- [8] Marine Louargant, Gawain Jones, Romain Faroux, Jean-Noël Paoli, Thibault Maillot, Christelle Gée, and Sylvain Villette. Unsupervised Classification Algorithm for Early Weed Detection in Row-Crops by Combining Spatial and Spectral Information. *Remote Sensing*, 10(5) :761, may 2018.
- [9] Ana I. de Castro, Jorge Torres-Sánchez, Jose M. Peña, Francisco M. Jiménez-Brenes, Ovidiu Csillik, and Francisca López-Granados. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sensing*, 10(2) :1–21, 2018.
- [10] M Dian Bah, Adel Hafiane, and Raphael Canals. Deep learning with unsupervised data labeling for weed detection in line crops in uav images. *Remote Sensing*, 10(11), 2018.
- [11] Shahrani Shahbudin, Masturah Zamri, Murizah Kassim, Syahrul Afzal Che Abdullah, and Saiful Izwan Suliman. Weed classification using one class support vector machine. In *2017 International Conference on Electrical, Electronics and System Engineering (ICEESE)*, pages 7–10. IEEE, nov 2017.
- [12] Prakruti Bhatt, Sanat Sarangi, and Srinivasu Pappula. Coarse clustering and classification of images with cnn features for participatory sensing in agriculture. In *ICPRAM*, pages 488–495, 2018.
- [13] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf : A deep convolutional activation feature for generic visual recognition. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 647–655, Beijing, China, 22–24 Jun 2014. PMLR.
- [14] M. Dian Bah, Adel Hafiane, and Raphael Canals. Weeds detection in UAV imagery using SLIC and the hough transform. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, nov 2017.
- [15] Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. In *Advances In Neural Information Processing Systems 12*, volume 12, pages 582–588. MIT Press, 2000.
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. IEEE, jun 2016.