

Département de géomatique appliquée  
Faculté des lettres et sciences humaines  
Université de Sherbrooke

**Caractérisation de la bande riveraine en milieu agricole à l'aide de réseaux de neurones convolutifs profonds multi-vues (MVDCNN) et d'images satellites**

Samuel de la Sablonnière

Mémoire présenté pour l'obtention du grade de Maître ès sciences (M.Sc.),  
Cheminement en télédétection

Avril 2022

© Samuel de la Sablonnière



**Directeur de recherche :**

Pr. Yacine Bouroubi

Département de Géomatique appliquée, Université de Sherbrooke, Sherbrooke (QC)

**Codirecteurs de recherche :**

Pr. Samuel Foucher

Département de Géomatique appliquée, Université de Sherbrooke, Sherbrooke (QC)

Dr. Etienne Lord

Centre de recherche et de développement de St-Jean-sur-Richelieu, Agriculture et Agroalimentaire Canada, St-Jean-sur-Richelieu (QC)

**Membres du jury**

Pr. Mickael Germain

Université de Sherbrooke, Sherbrooke (QC)

Membre interne du jury

Dr. Pierre-Luc St-Charles

MILA, Montréal (QC)

Membre externe du jury

## Résumé

En raison des fonctions écologiques qu'elles remplissent, les bandes riveraines végétalisées sont une solution fréquemment employée pour faire face aux problèmes de pollution diffuse en milieu agricole. Or, elles subissent des pressions importantes de nature anthropique qui justifient un suivi de leur état. La caractérisation des bandes riveraines est le plus souvent réalisée par des mesures *in situ* d'indices de qualité, ce qui mobilise des ressources importantes lorsqu'un vaste territoire est à couvrir et limite la récurrence du suivi. L'objectif global de cette recherche était donc de proposer une approche pour la caractérisation des bandes riveraines en milieu agricole qui est accessible et facilement applicable à l'échelle régionale. Elle s'inspire des récentes avancées en classification d'images apportées par le développement des réseaux de neurones convolutifs profonds (DCNN) et vise à évaluer la capacité de cette technologie pour la détermination de l'indice de qualité des bandes riveraines (IQBR) directement à partir d'images satellites. Pour ce faire, la méthode proposée fait appel aux images multispectrales à très haute résolution spatiale de la constellation Pléiades.

De par leur forme étroite et irrégulière, les bandes riveraines ne peuvent être entièrement couvertes par une seule image sans que cette dernière image y intègre des éléments trop éloignés de la bande riveraine. Conséquemment, une architecture de DCNN multi-vues (MVDCNN) modifiée permettant l'utilisation de plusieurs images en entrée de réseau a été entraînée pour établir une corrélation entre l'IQBR mesuré sur le terrain et un ensemble d'images représentant la bande riveraine. Spécifiquement, différents nombres d'images en entrée du réseau multi-vues, sept combinaisons de bandes spectrales des images Pléiades et deux modes d'entraînement ont été testés, soit l'entraînement à partir de zéro et le *fine tuning* d'un réseau préentraîné. Les résultats démontrent qu'avec l'utilisation des bandes spectrales RGB, l'architecture MVDCNN incorporant un RESNET-18 préentraîné parvient à établir la meilleure corrélation ( $R^2$  de 0,932) entre les images de la bande riveraine et son IQBR. De plus, l'emploi de plusieurs images en entrée de réseau lors de l'entraînement a amélioré les résultats par rapport à l'utilisation d'une seule image.

**Mots-clés :** Indice de qualité des bandes riveraines, télédétection, réseaux de neurones convolutifs, architecture multi-vues, agriculture, Pléiades.

## Table des matières

Liste des figures.....	iii
Liste des tableaux.....	iv
Liste des acronymes .....	v
Avant-propos .....	vi
Remerciements .....	vii
<b>1. Introduction.....</b>	<b>1</b>
<b>1.1 Contexte.....</b>	<b>1</b>
<b>1.2 Problématique.....</b>	<b>2</b>
<b>2. Revue de littérature .....</b>	<b>3</b>
<b>2.1 L'indice de qualité des bandes riveraines .....</b>	<b>3</b>
<b>2.2 Utilisation de la télédétection pour la caractérisation de la bande riveraine.....</b>	<b>5</b>
<b>2.3 Réseaux de neurones convolutifs profonds .....</b>	<b>8</b>
<i>2.3.1 Apprentissage par transfert.....</i>	<i>8</i>
<i>2.3.2 Architecture VGG (Visual Geometry Group) .....</i>	<i>10</i>
<i>2.3.3 Architecture RESNET.....</i>	<i>11</i>
<i>2.3.4 Architectures multi-entrées de réseau de neurones .....</i>	<i>12</i>
<b>3. Objectifs et hypothèses de recherche .....</b>	<b>15</b>
<b>4. Article: Riparian Strip Characterization in Agricultural Environments Using Multi-View Deep Convolutional Neural Network (MVDCNN) and Satellite Imagery .....</b>	<b>17</b>
<b>4.1 Introduction .....</b>	<b>18</b>
<b>4.2 Materials and Methods .....</b>	<b>21</b>
<i>4.2.1 Study area.....</i>	<i>21</i>
<i>4.2.2 Riparian Strip Quality Index (RSQI).....</i>	<i>23</i>
<i>4.2.3 Field data .....</i>	<i>24</i>
<i>4.2.4 Satellite images .....</i>	<i>24</i>
<i>4.2.5 Automated dataset creation.....</i>	<i>25</i>
<i>4.2.6 MVDCNN architecture.....</i>	<i>26</i>
<i>4.2.7 Network training .....</i>	<i>27</i>

4.2.8 Network testing.....	29
4.2.9 Object-based classification .....	30
<b>4.3 Results.....</b>	<b>31</b>
4.3.1. Determining the optimal number of input images.....	31
4.3.2 Effect of training strategy and choice of spectral bands.....	32
4.3.3 Object-based classification .....	35
4.3.4 Testing the model on simulated riparian strips.....	37
<b>4.4 Discussion .....</b>	<b>40</b>
<b>4.5 Conclusions .....</b>	<b>43</b>
<b>References .....</b>	<b>45</b>
<b>Supplementary .....</b>	<b>50</b>
<b>4.6 Discussion (en supplément de l'article).....</b>	<b>51</b>
<b>5. Conclusion .....</b>	<b>52</b>
5.1 Contributions .....	52
5.2 Limites et perspectives .....	54
<b>Références .....</b>	<b>57</b>

## Liste des figures

<b>Figure 1.</b> Exemple pratique de l'évaluation des superficies relatives des 9 composantes de la bande riveraine entrant dans le calcul de l'IQBR.....	4
<b>Figure 2.</b> Un bloc convolutif résiduel dans un RESNET .....	11
<b>Figure 3.</b> L'architecture du RESNET-18.....	12
<b>Figure 4.</b> La fusion de caractéristiques en entrée de réseau.....	13
<b>Figure 5.</b> L'architecture MVDCNN.....	15
<b>Figure 6.</b> Pleiades mosaic covering the L'Acadie River Watershed, Quebec (Canada), and riparian buffer strips in vector format. ....	22
<b>Figure 7.</b> Examples of riparian strips in a Pleiades image for different values and classes of RSQI. ....	24
<b>Figure 8.</b> Distribution of RSQI values in the entire dataset .....	26
<b>Figure 9.</b> MVDCNN architecture that was used for the determination of RSQI.....	27
<b>Figure 10.</b> Flowchart used in preparing the image dataset for network cross-validation training. ....	29
<b>Figure 11.</b> $R^2$ results of RSQI prediction on the test set as a function of the number of input images .....	31
<b>Figure 12.</b> RMSE (Root-Mean-Square Error) and regression slope (b) obtained on the test set as a function of the number of simultaneous input images.....	32
<b>Figure 13.</b> RSQI predictions of the best MVDCNN based model (a) and the results obtained from the OBC (b) as a function of the index values that were measured in the field.....	35
<b>Figure 14.</b> Representative object-based classification and RSQI predicted values .....	37
<b>Figure 15.</b> Model predictions when 4 images of two riparian strips are mixed.....	38
<b>Figure 16.</b> Example of a dataset that is used to evaluate a gradient of RSQI.....	39
<b>Figure 17.</b> Five predictions of RSQI that are made by the model as a function of the percentage of forest that was found in each of the images to which it was presented .....	39

**Liste des tableaux**

<b>Tableau 1.</b> L'architecture du réseau VGG.....	10
<b>Tableau 2.</b> From the five cross-validation trainings, average RMSE and standard deviation of RSQI predictions on the validation, balanced validation and test datasets for the two training modes and 7 spectral band combinations.....	34
<b>Tableau 3.</b> From the five cross-validation trainings, average $R^2$ values and standard deviation of RSQI predictions on the validation, balanced validation and test datasets for the two training modes and 7 spectral band combinations.....	34



## Liste des acronymes

B:	Bande spectrale Bleu d'une image satellitaire
DCNN :	Réseau de neurones convolutif profond, de l'anglais <i>Deep Convolutional Neural Network</i>
G :	Bande spectrale Verte ( <i>Green</i> en anglais) d'une image satellitaire
IBI :	Indice d'intégrité biologique, de l'anglais <i>Index for Biological Integrity</i>
IQBR :	Indice de qualité des bandes riveraines
LiDAR :	<i>Light Detection And Ranging</i> , en anglais
LSTM:	<i>Long Short-Term Memory</i> , en anglais
MELCC :	Ministère de l'Environnement et de la Lutte contre les Changements Climatiques
MVDCNN :	Réseau de neurones convolutif profond multi-vue, de l'anglais <i>Multi-View Deep Convolutional Neural Network</i>
N :	Bande proche-infrarouge, <i>Near-infrared</i> en anglais
OBV :	Organisme de Bassin Versant
PolSAR :	Radar polarimétrique à synthèse d'ouverture, de l'anglais <i>Polarimetric Synthetic Aperture Radar</i>
QBR :	Qualité de la forêt riveraine, du catalan <i>Qualitat del Bosc de Ribera</i>
R :	Bande spectrale Rouge d'une image satellitaire
RMSE :	La racine carrée de l'erreur quadratique moyenne, de l'anglais <i>Root Mean Square Error</i>
RQI :	Indice de qualité riveraine, de l'anglais <i>Riparian Quality Index</i>
SVM :	Machine à vecteurs de support, de l'anglais <i>Support Vector Machine</i>
THR :	Très Haute Résolution spatiale
TL :	Apprentissage par transfert, de l'anglais <i>Transfer Learning</i>

## **Avant-propos**

Ce document est présenté sous la forme d'un mémoire par article composé de 5 chapitres. Le premier introduit le contexte du projet ainsi que la problématique. Les deux chapitres suivants s'attardent respectivement à la revue de littérature aux objectifs de recherche. Le quatrième chapitre est l'article scientifique soumis à la revue *Remote Sensing*, le 2 novembre 2021 pour révision et dont je suis le premier auteur. Il est présenté en anglais et dans un format adapté pour faciliter la lecture du présent document. L'article intitulé *Riparian Strip Characterization in Agricultural Environments Using Multi-View Deep Convolutional Neural Network (MVDCNN) and Satellite Imagery* a été co-écrit par Samuel Foucher du département de géomatique appliquée de l'Université de Sherbrooke, Étienne Lord d'Agriculture et Agroalimentaire Canada (AAC), Philippe Vigneault d'AAC et Yacine Bouroubi du département de géomatique appliquée de l'Université de Sherbrooke. L'article comporte, entre autres, la démarche méthodologique ainsi que les résultats du projet. Aussi, des résultats préliminaires de cette recherche ont été présentés le 2 juin 2021 lors du congrès de l'Association Québécoise de Télédétection (AQT) ainsi que le 23 juin 2021 lors d'une session d'affiche au congrès de la Société Canadienne de Télédétection (SCT 2021). Le cinquième et dernier chapitre conclut ce mémoire.

## **Remerciements**

Je tiens tout d'abord à remercier mon directeur de recherche Yacine Bouroubi pour sa grande disponibilité, son soutien, ses idées et pour m'avoir fait grandement confiance tout au long de ce projet. Merci à mon co-directeur Samuel Foucher pour son aide technique précieuse, sa grande expertise en intelligence artificielle et son apport constant de nouvelles idées. Un grand merci également à mon co-directeur Étienne Lord pour son support en préparation des conférences, sa vision précise de l'ensemble du projet et pour m'avoir guidé à de nombreuses reprises depuis l'amorce de cette aventure. Je veux également remercier Louis Longchamps et Philippe Vigneault d'Agriculture et Agroalimentaire Canada qui ont été les instigateurs de ce projet de recherche.

Je me sens choyé d'avoir été aussi bien entouré pendant ces deux dernières années. Ce projet de maîtrise n'aurait pu aussi bien se dérouler sans l'appui de toute cette équipe et une bonne part de la qualité retrouvée dans ce document lui revient.

Ce projet de maîtrise a été rendu possible grâce à la contribution financière du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG), Mitacs, Effigis Geo Solutions©, le Centre de Recherche Informatique de Montréal (CRIM) et Agriculture et Agroalimentaire Canada (AAC). D'autre part, les données de terrain sur lesquelles repose ce projet ont été gracieusement fournies par Marcel Comiré de l'organisme de bassin versant COVABAR. Sans cette généreuse contribution et la grande ouverture de M. Comiré, ce projet dans sa forme actuelle n'aurait pu exister.

Enfin, je dédie cette maîtrise à ma famille pour son immense support et sa patience, et ce malgré mes nombreuses absences causées par le voyage Sherbrooke – Montérégie pendant ces cinq dernières années. Un merci tout spécial à ma conjointe Mélanie sans qui ce long retour aux études n'aurait pu être possible. Je t'en serai éternellement reconnaissant.



## 1. Introduction

### 1.1 Contexte

En milieu agricole, l'utilisation intensive de nutriments tels que l'azote et le phosphore dans la production des cultures fragilise les écosystèmes aquatiques (Carpenter *et al.*, 1998). À titre d'exemple, en 2003, 69 % de la charge en phosphore du bassin versant de la rivière aux Brochets, Québec, provenait des cultures annuelles, alors que celles-ci ne représentent que 22 % du territoire (Gangbazo *et al.*, 2006). Ces apports d'engrais aux cours d'eau, causés principalement par le ruissellement de surface, contribuent à la prolifération de plantes et d'algues nuisibles ce qui a un impact direct sur les écosystèmes aquatiques et les populations riveraines. En trop grande importance, ces apports peuvent également compromettre l'approvisionnement en eau potable et nécessiter l'utilisation de traitements coûteux pour atteindre les normes de consommation (Carluer *et al.*, 2017).

Les principaux acteurs du milieu (c.-à-d. les gouvernements fédéral, provincial et municipaux, les agronomes, les agriculteurs, les citoyens riverains, etc.) sont de plus en plus conscientisés à ce phénomène et mettent de l'avant des pratiques visant à réduire les impacts des activités agricoles sur l'environnement. Par exemple, dans le domaine de la production des cultures, des pratiques de gestion bénéfiques (PGB) comme la rotation des cultures, la conservation des résidus végétaux suite aux récoltes et l'implantation de rigoles gazonnées permettent de réduire les sources de pollution diffuses (Stewart *et al.*, 2012). L'aménagement de bandes riveraines, lisières de végétation naturelle et permanente en bordure d'un cours d'eau (Doyon, 2015) est quant à elle une formule privilégiée pour diminuer l'impact des activités agricoles sur les écosystèmes (de Sosa *et al.*, 2018a; de Sosa *et al.*, 2018b; Dosskey *et al.*, 2005; Stutter *et al.*, 2012; White et Arnold, 2009). Dans une optique de protection de l'environnement, la bande riveraine est régie au Québec par la *Politique de protection des rives, du littoral et des plaines inondables* (PPRLPI, 2020) du Gouvernement du Québec, qui stipule qu'elle doit couvrir au moins 3 mètres de largeur parallèlement au cours d'eau dans les secteurs voués à l'agriculture.

La bande riveraine constitue le milieu de transition entre les écosystèmes terrestres et aquatiques. De par son positionnement névralgique, elle est touchée par de perturbations naturelles récurrentes, ce qui en fait un milieu très dynamique (Gagnon et Gangbazo, 2007). Il s'agit d'un écosystème de grande importance en milieu agricole en raison de l'effet tampon qu'elle procure entre les milieux

anthropisés (c.-à-d. les terres en culture) et les cours d'eau. En effet, elle constitue le dernier rempart physique contre toutes pertes provenant des sols à proximité. En plus des rôles de capteur de nutriments et de rétenteur de sédiments, elle remplit les fonctions de stabilisation des berges, de thermorégulation et de maintien de la biodiversité (de Sosa, *et al.*, 2018; Saint-Jacques et Richard, 1998). Les bandes riveraines sont cependant sujettes à d'importantes pressions causées par l'activité humaine. On compte parmi elles l'étalement urbain, l'expansion des terres agricoles et la construction de digues et de barrages (Poff *et al.*, 2011). Un suivi de leur état est donc primordial afin d'avoir une évaluation juste de leur capacité à remplir leurs fonctions écologiques (ministère de l'Environnement et de la Lutte contre les changements climatiques [MELCC], 2020).

## 1.2 Problématique

Pour permettre la surveillance des écosystèmes riverains et afin de cibler les secteurs requérant la restauration ou l'implantation de bandes riveraines végétalisées, un portrait global d'une région donnée est requis par les décideurs. Les approches existantes pour la caractérisation des bandes riveraines, nécessaire à la réalisation de ce portrait, reposent généralement sur la mesure d'indices comme l'IBI (*Index for biological Integrity*, en anglais) (Miller *et al.*, 2006), le RQI (*Riparian quality index*, en anglais) (Gonzalez del Tanago et Garcia De Jalon, 2011) ou le QBR (*Qualitat del bosc de ribera*, en catalan) (Munné *et al.*, 2003) spécialement développés à cette fin et adaptés aux attributs de l'environnement visé. Ces indices intègrent un ou plusieurs indicateurs de la santé des écosystèmes riverains, dont les propriétés de la végétation riveraine et l'hydrogéomorphologie (Allostry *et al.*, 2020). Au Québec (Canada), c'est l'indice de qualité des bandes riveraines (IQBR, voir section 3.1 pour la description détaillée) (Saint-Jacques et Richard, 1998) qui est utilisé pour faire ce suivi. En résumé, celui-ci est une somme pondérée des superficies occupées par les différentes composantes d'occupation du sol de la bande riveraine, comme la forêt, les herbacées, les cultures et les surfaces bâties.

Le défi que pose les protocoles de mesures d'indices pour la caractérisation des bandes riveraines est la difficulté d'application lorsque des superficies considérables sont à couvrir. Au Québec, les organismes de bassins versants (OBV) sont responsables de l'évaluation de l'état des écosystèmes riverains pour leur territoire respectif. Lors des campagnes de caractérisation, les OBV déploient des équipes sur le terrain pour recueillir les données nécessaires au calcul de l'indice, spécifiquement l'IQBR. À titre d'exemple, un OBV a nécessité une saison complète et cinq

évaluateurs experts pour couvrir 14 000 hectares d'un bassin versant et ses 204 kilomètres de rive (COVABAR, 2016). Par conséquent, ces campagnes de terrain sont coûteuses en temps, en ressources humaines et en ressources financières. En raison de ces contraintes, un suivi récurrent de l'état des bandes riveraines par ce type de méthode est rarement envisageable. Cette recherche vise donc le développement d'une méthode de suivi qui est applicable à l'échelle régionale, tout en bénéficiant des bases scientifiques de caractérisation des bandes riveraines sur lesquels reposent les indices déjà développés. Elle se veut en conjonction avec le protocole de caractérisation actuellement utilisé sur le territoire québécois, soit la mesure de l'IQBR. Pour faciliter la lecture du présent document, les objectifs principal et spécifiques sont présentés suite à la revue de littérature.

## 2. Revue de littérature

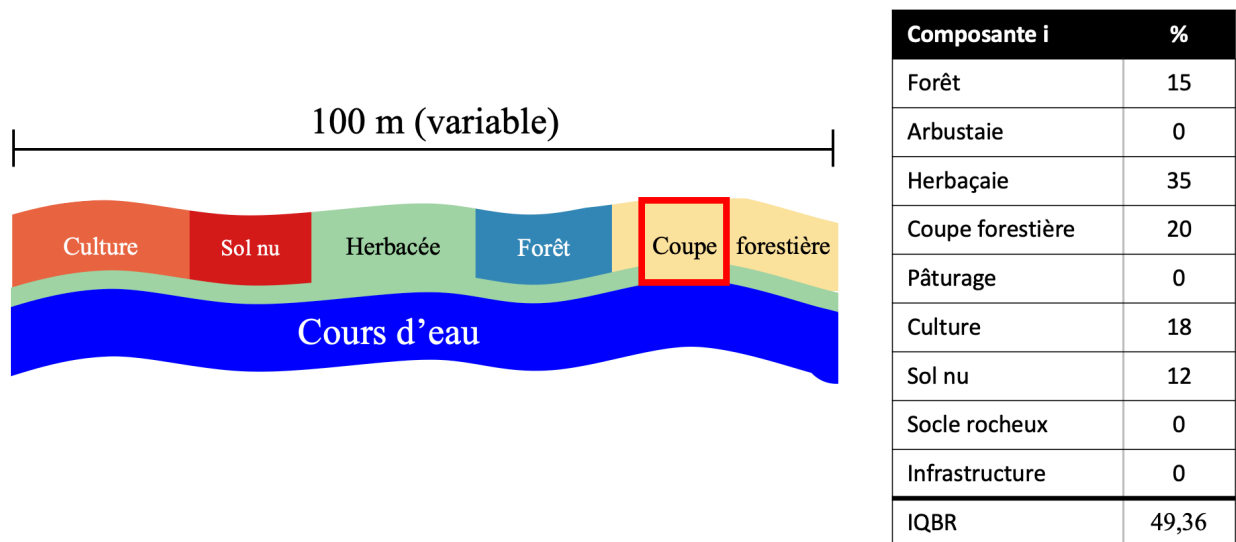
La présente section détaille trois parties nécessaires à la compréhension de ce projet. L'IQBR est l'indice sur lequel l'évaluation de la qualité de la bande riveraine repose dans le cadre de cette recherche. Sa description détaillée ainsi que son protocole de mesure sont présentés dans la première partie. Afin de situer l'approche proposée parmi les méthodes existantes, la seconde partie est un survol de l'état de l'art concernant l'utilisation de la télédétection pour la caractérisation de la bande riveraine. La troisième partie s'intéresse aux réseaux de neurones convolutifs profonds (DCNN de l'anglais *Deep Convolutional Neural Network*). Elle s'attarde plus spécifiquement aux techniques d'apprentissage par transfert, à deux architectures dont il est mentionné dans l'article (soit VGG et RESNET) et enfin aux architectures de DCNN multi-entrées, éléments centraux de l'approche présentée au chapitre 4.

### 2.1 L'indice de qualité des bandes riveraines

L'IQBR permet d'évaluer « la condition écologique de l'habitat riverain et son impact sur l'intégrité du milieu aquatique » (Saint-Jacques et Richard, 1998). Il est une somme pondérée des superficies occupées par les composantes d'occupation du sol de la bande riveraine où les poids les plus importants, selon les auteurs, sont attribués aux types de surface contribuant le plus à la santé du cours d'eau. Il est défini selon l'équation suivante:

$$IQBR = \frac{\sum(\%_i \times P_i)}{10} \quad (1)$$

Où  $\%i$  est le pourcentage de couverture du sol de la  $i$  ième composante de la bande riveraine, et  $Pi$ , le facteur de pondération de cette composante. Une largeur de 10 m de bande riveraine est considérée dans le calcul des superficies sur une longueur dépendant des segments de rives déterminés. Plus précisément, si un changement important d'occupation du sol est observé, par exemple en passant d'un milieu urbain à un milieu forestier, un nouveau segment est créé. Les facteurs de pondération sont les suivants : forêt (10,0), arbustaie (8,7), herbaçaie (5,8), coupe forestière (4,3), sol rocheux (3,8), pâturage (3,0), culture (1,9), infrastructure (1,9) et sol nu (1,7) (Saint-Jacques et Richard, 1998). Les valeurs de l'indice se situent donc entre 17 et 100, la plus élevée représentant la meilleure par rapport à la capacité de la bande riveraine à remplir ses fonctions écologiques. Ces valeurs sont classées de la façon suivante par l'OBV COVABAR: très faible (17-21), faible (21-40), moyen (40-60), bon (60-80), excellent (80-100). Un exemple pratique de l'évaluation des superficies relatives des 9 composantes de la bande riveraine entrant dans le calcul de l'IQBR est présenté à la **Figure 1**. En appliquant l'équation 1 aux valeurs déterminées à la **Figure 1**, on obtient un IQBR de 49,36  $\left(\frac{(15 \times 10) + (35 \times 5,8) + (20 \times 4,3) + (18 \times 1,9) + (12 \times 1,7)}{10}\right)$ . Ce protocole a largement été adopté par les OBV, entités responsables partout dans la province de l'établissement d'un plan directeur de l'eau dont l'évaluation de l'état des bandes riveraines est partie prenante.



**Figure 1.** Exemple pratique de l'évaluation des superficies relatives des 9 composantes de la BR entrant dans le calcul de l'IQBR. En appliquant l'équation 1 à ces données, on obtient un IQBR de 49,36. Le carré rouge représente une zone de 100 m<sup>2</sup>.



## 2.2 Utilisation de la télédétection pour la caractérisation de la bande riveraine

Comme démontré par Huylenbroeck *et al.* (2020), l'utilisation de la télédétection pour la caractérisation des couvertures végétales des secteurs riverains a grandement gagné en importance depuis les dix dernières années. Ceci est en bonne partie grâce au coût réduit d'acquisition des images d'observation de la Terre, de leur disponibilité accrue et de l'amélioration des résolutions spatiale et spectrale. De plus, les nouvelles méthodes d'acquisition d'images par drone et l'accessibilité aux données LiDAR (de l'anglais *Light Detection and Ranging*) ont multiplié les possibilités d'application. En ce sens, en fonction des paramètres de la végétation visés, les chercheurs ont recours à des produits de la télédétection spécifiques. Les images multispectrales (aéroportées ou satellitaires) sont majoritairement utilisées pour la délimitation de la végétation riveraine, comme dans Bertoldi *et al.* (2011), ou pour la détermination de la composition des espèces végétales (Fernandes *et al.*, 2013; Macfarlane *et al.*, 2017) pour laquelle les résolutions spatiales entre 1 et 10 m sont les plus indiquées (Huylenbroeck *et al.*, 2020). D'autre part, les chercheurs font appel aux images radar et aux nuages de points LiDAR pour caractériser la structure de la végétation. À titre d'exemple, Dufour *et al.* (2013) ont utilisé les images radar, les nuages de points LiDAR et les images acquises par drones pour une caractérisation tridimensionnelle de la végétation riveraine, et ce dans un objectif de suivi des projets de restaurations à l'échelle locale. Michez *et al.* (2017) ont aussi utilisé des données LiDAR pour cartographier les paramètres des forêts riveraines aux échelles locale et régionale. Les méthodes employées et les indicateurs que l'on peut extraire de produits de la télédétection pour soutenir l'étude des végétations riveraines sont donc très nombreux et souvent spécifiques à la région visée.

Dans un objectif d'opérationnalité, certaines études cherchent à établir des liens plus directs entre les méthodes de caractérisation *in situ* des bandes riveraines (c.-à-d. la mesure d'indices) et les images de télédétection. Bien que des observations sur le terrain soient un prérequis absolu pour la mesure de certains indices, par exemple l'IBI qui exige un inventaire exhaustif de la superficie relative qu'occupe les espèces de plantes envahissantes, de plantes annuelles et des cryptogames vasculaires parmi plusieurs, une part des indicateurs intégrés dans ces indices est liée aux caractéristiques de l'occupation du sol telles l'identification des communautés végétales, la couverture forestière et la pente du chenal. Puisque ces indicateurs sont observables ou dérivables

à partir d'images de télédétection, une corrélation peut être envisagée. Quelques travaux ayant fait ces rapprochements sont ici présentés.

Dans le contexte des bandes riveraines des savanes tropicales australiennes, habitats au centre d'activités liées au développement du pays comme le pâturage, l'agriculture, le tourisme et l'urbanisation, Johansen *et al.* (2007) ont comparé la mesure de l'état des bandes riveraines selon une méthode d'évaluation rapide sur le terrain, le *Tropical Rapid Appraisal of Riparian Condition* (TRARC), et une méthode basée sur des images à haute résolution spatiale (capteur QuickBird, 2,4 m de résolution). La méthode TRARC évalue 23 indicateurs rassemblés en un score sur 100 reflétant l'état général de la bande riveraine. Cinq de ces indicateurs, dont le pourcentage de couverture du couvert végétal, la litière organique et la continuité du couvert végétal ont été dérivés à partir des images satellites à l'aide du calcul d'indices de végétation (NDVI, EVI, et SAVI) et d'une classification orientée objet. Ces indicateurs ont été corrélés à la méthode de mesure TRARC et le résultat est un  $R^2$  moyen de 0,75 pour les 16 sites d'étude. Ils ont aussi démontré que la méthode qui utilise les images de télédétection était plus appropriée pour le suivi des changements dans le temps et qu'elle était moins coûteuse en temps et ressources financière lorsqu'une longueur considérable de cours d'eau devait être caractérisée (> 200 km).

Dans Ivits *et al.* (2009), le statut écologique de la bande riveraine en milieu agricole est classé comme favorable ou défavorable selon le pourcentage de végétation permanente que l'on retrouve sur le territoire. Ce pourcentage provient du calcul du *Green Vegetation Fraction* (Gutman et Ignatov, 1998) sur une série d'images satellitaires de 1989 à 2004 à l'aide du *Advanced Very High Resolution Radiometer* ayant une résolution de 1 km. Bien que cette étude renseigne sur l'état des bandes riveraines à une échelle très vaste, l'utilisation faite des images de télédétection pour leur caractérisation est peu précise, autant au niveau de la définition des classes (favorable ou défavorable) que de la résolution spatiale des prédictions. Elle qualifie la zone riveraine au sens très large et, reposant sur des images datant de plusieurs années, ne prend pas en compte les changements annuels de l'occupation du territoire agricole.

Fernández *et al.* (2014) ont quant à eux modélisé la qualité des habitats riverains en se basant sur des données d'occupation du sol à différentes échelles, soit de 1:5 000, de 1:25 000 et de 1:100 000, et ce pour une évaluation à l'échelle régionale du nord de l'Espagne. Les données les plus précises (échelle 1:5000) proviennent de cartes produites par l'interprétation de photos aériennes effectuées

par des experts en botanique. Les auteurs utilisent des régressions linéaires multiples et l'algorithme *Random Forest* pour évaluer la capacité des données à bien classifier les différents habitats. Les données de référence proviennent quant à elles de l'indice RQI. Les auteurs ont obtenu les meilleurs résultats par régression linéaire avec un  $R^2$  de 0,6 en utilisant les données cartographiques à l'échelle 1:5000. Ils montrent ainsi la force de la corrélation entre l'indice mesuré et les données d'occupation du sol. Par contre, cette recherche repose sur une cartographie précise comprenant une annotation manuelle de la flore par des spécialistes, ce qui limite l'applicabilité de la méthode à d'autres régions et la récurrence de l'opération. De plus, la délimitation de la zone riveraine, nécessaire pour l'utilisation de ce modèle, requiert plusieurs étapes de prétraitements difficilement accessibles aux non-experts.

Dans Novoa *et al.* (2018), un nouvel indice intégrant l'IQBR et prenant en compte la capacité de drainage de la bande riveraine, le *Riparian Strip Efficiency Index* (RSEI), a été développé. Dans cette étude, la détermination de l'IQBR a été réalisée à partir d'images WorldView-2 (50 cm de résolution pour la bande panchromatique, 2 m pour les 8 bandes multispectrales). Pour ce faire, les auteurs ont effectué une classification orientée objet des composantes de la bande riveraine. L'IQBR moyen obtenu à l'aide de cette classification sur les 80 sites de validation était de 56 contre 55 pour la moyenne des mesures *in situ* (aucun indice de corrélation n'est fourni). Bien que cette méthode fonctionne bien, les prétraitements effectués sur les images et le fait de recourir à la classification orientée objet nécessitent une bonne expertise, ce qui la rend moins accessible. D'autre part, des compétences en photo-interprétation sont toujours nécessaires pour effectuer une classification juste des différents types de végétations et classes d'occupation du sol. Malgré ces restrictions, les auteurs démontrent bien que des images satellitaires à très haute résolution spatiale (THR, i.e. < 1 m) contiennent l'information nécessaire pour la détermination de l'IQBR.

Puisque l'IQBR repose entièrement sur des éléments observables sur les images de télédétection THR, il est un candidat de choix pour envisager une bonne corrélation entre ces images et les mesures de l'indice sur le terrain. Dans le but de faciliter le suivi de l'état des bandes riveraines, c'est-à-dire en réduisant l'expertise et les prétraitements requis comme le calcul d'indices dérivés, la classification orientée objet ou la photo-interprétation des classes de végétation et d'occupation du sol, une approche faisant appel aux plus récentes avancées en matière de classification d'images pourrait s'avérer pertinente. Les réseaux de neurones convolutifs profonds, architectures

d'apprentissage profond, ont en effet l'avantage de générer de manière autonome, par le processus d'apprentissage, les caractéristiques nécessaires à l'établissement de la corrélation entre une image et sa classe d'appartenance (son étiquette), qui est dans le cas présent un indice numérique de caractérisation (l'IQBR) de la bande riveraine. Il n'existe pas, à notre connaissance, d'article dans la littérature qui explore cette approche.

Dès lors, l'intuition derrière l'amorce de cette recherche peut être formulée comme suit : Étant donné que les paramètres requis pour la mesure de l'IQBR sont observables sur des images de télédétection, un DCNN devrait être en mesure, avec un protocole d'entraînement adéquat, d'apprendre à établir une bonne corrélation entre les images constituant une bande riveraine et la valeur d'IQBR associée.

## **2.3 Réseaux de neurones convolutifs profonds**

### *2.3.1 Apprentissage par transfert*

Depuis quelques années, l'utilisation de DCNN pour les tâches de classification (associer une image à une classe), de reconnaissance de scènes, de segmentation et de détection d'objets sur les images de télédétection est très populaire en raison des performances de ce type d'apprentissage machine dépassant souvent et très largement celles des techniques dites traditionnelles (Kussul *et al.*, 2017; Ma *et al.*, 2019a; Romero *et al.*, 2016). D'autre part, les DCNN ne nécessitent pas l'intervention d'experts pouvant décrire les caractéristiques des objets ou des classes à l'étude sur les images. Ils génèrent par eux-mêmes, lors du processus d'entraînement, les caractéristiques du plus bas niveau (textures, arêtes) jusqu'au plus haut (branche d'un arbre par exemple) (Nogueira *et al.*, 2017).

L'entraînement complet d'un réseau de neurones profond exige une quantité importante de données. De surcroît, dans le cas d'un apprentissage supervisé, ces données doivent passer par l'étape d'annotation, ce qui exige beaucoup de temps. Le problème est que ces deux prérequis sont rarement atteints dans le domaine de l'observation de la Terre et que l'utilisation d'un jeu de données de taille insuffisante se solde le plus souvent en un important surapprentissage (Shorten et Khoshgoftaar, 2019). Ce phénomène se produit lorsque le réseau performe très bien sur le jeu de données d'entraînement, mais beaucoup moins sur les jeux de validation et de test. Le réseau apprend en quelque sorte les données « par cœur » et n'est pas en mesure de généraliser à d'autres cas. Des techniques d'augmentation de données (Shorten et Khoshgoftaar, 2019) et de

régularisation (Kukačka *et al.*, 2017) réduisent ce phénomène, mais ne sont pas suffisantes dans tous les cas. Il est par contre possible d'utiliser les caractéristiques apprises par des réseaux entraînés sur des jeux de données volumineux à d'autres fins qu'à celles pour lesquelles ils ont été conçus au départ. Cette technique se nomme l'apprentissage par transfert (TL, de l'anglais *transfer learning*, (Pan et Yang, 2010)) et plusieurs variantes de celle-ci existent. L'utilisation de DCNN pré-entraînés comme extracteur de caractéristiques consiste en le retrait de la dernière couche (couche supérieure) du réseau contenant les informations sur les classes d'origine afin de conserver les caractéristiques déjà apprises pour l'entraînement d'un nouveau classificateur. Pour le *fine tuning*, autre technique de TL, elle consiste en l'ajustement des poids du modèle à partir du jeu de données d'intérêt pour l'adapter à la tâche visée. Le modèle préentraîné sert alors de point de départ. Cette technique est par contre toujours sujette au surapprentissage puisque le réseau a toute la latitude pour se coller aux données qu'on lui montre. L'option qui s'offre pour éviter ce phénomène est alors de seulement ajuster une sélection des poids du modèle plutôt que l'intégralité. Bien entendu, les jeux de données de référence volumineux comme ImageNet (Deng *et al.*, 2009) et PASCAL-VOC (Everingham *et al.*, 2010), largement utilisés dans la littérature, sont composés d'images aux caractéristiques très différentes de celles des images satellitaires. Malgré ce fait, dans la mesure où l'on se limite à l'utilisation de trois bandes (ImageNet contient des images RGB seulement), il a été démontré qu'une architecture pré-entraînée sur ces jeux de données est adaptable aux images de télédétection (Marmanis *et al.*, 2016; Penatti *et al.*, 2015). En effet, les caractéristiques générales présentes dans les premières couches tendent à se ressembler d'un réseau à un autre et d'un type d'images à un autre. Ce sont ces caractéristiques peu profondes, les plus génériques et souvent analogues à des filtres de détection d'arêtes ou de couleurs qui sont facilement transférables. Ce n'est pas le cas pour les caractéristiques plus profondes qui sont beaucoup plus complexes et adaptées au jeu de données d'origine. De fait, une stratégie souvent utilisée est de garder intactes ces caractéristiques peu profondes et d'effectuer le *fine tuning* seulement sur les couches supérieures (Hu *et al.*, 2015; Mahdianpari *et al.*, 2018; Nogueira *et al.*, 2017; Penatti *et al.*, 2015). C'est cette stratégie qui a permis à Nogueira et al. (2017) d'obtenir les meilleurs résultats sur les jeux de données d'images satellitaires UCMerced (Yang et Newsam, 2010) et RS19 (Xia *et al.*, 2010).

### 2.3.2 Architecture VGG (Visual Geometry Group)

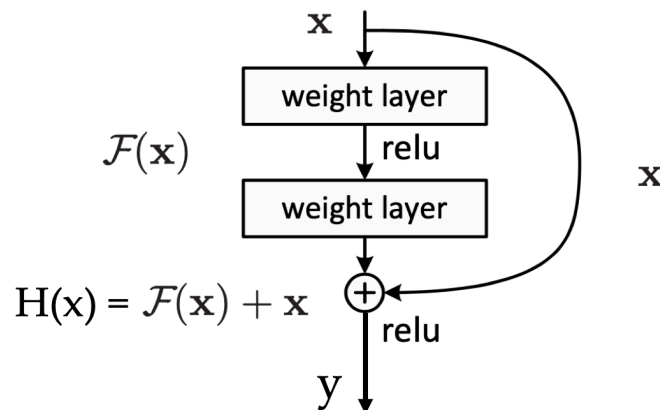
L'architecture *Visual Geometry Group* d'Oxford (VGG, Simonyan et Zisserman, 2015) a été marquante dans le développement des DCNN (plus de 11800 citations selon Scopus en novembre 2021) en raison de ses performances obtenues lors du concours *ImageNet Large Scale Visual Recognition Challenge 2014* (ILSVRC2014). VGG a innové en utilisant une succession de filtres convolutifs de petite taille (3 x 3 pixels) plutôt que les filtres de 11 x 11 pixels utilisés par le prédécesseur duquel il s'inspire, AlexNet (Krizhevsky *et al.*, 2012). Une succession de filtres de faible taille a l'avantage de réduire le nombre de paramètres dans le réseau et d'augmenter le nombre de filtres tout en conservant le même champ récepteur (Luo *et al.*, 2017). Ce nombre réduit de paramètres facilite l'entraînement du réseau en termes de temps de calcul et lui permet de converger plus rapidement. C'est aussi cette particularité (un nombre de paramètres plus bas) qui a permis aux chercheurs d'apporter leur innovation principale, soit l'augmentation de la profondeur du réseau. Ils ont ainsi testé des configurations de 11 couches jusqu'à 19 couches de profondeur (**Tableau 1**) plutôt que les 8 proposées par AlexNet. Ils ont ainsi démontré une amélioration des performances jusqu'à 16 couches (VGG-16, colonnes C et D dans le **Tableau 1**), mais une détérioration une fois dépassé les 19 couches. VGG-16 compte environ 138 millions de paramètres.

**Tableau 1.** L'architecture du réseau VGG (tableau tiré de Simonyan et Zisserman, 2015). La profondeur du réseau augmente de gauche à droite (A vers E). Les paramètres de la couche convolutive sont notés comme "conv[taille de la fenêtre]-[nombre de filtres]". Les couches de profondeur ajoutées de A vers E sont indiquées en gras. La nomenclature du réseau (e.g. VGG-16) fait référence au nombre de couches (e.g. 16 *weight layers*).

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

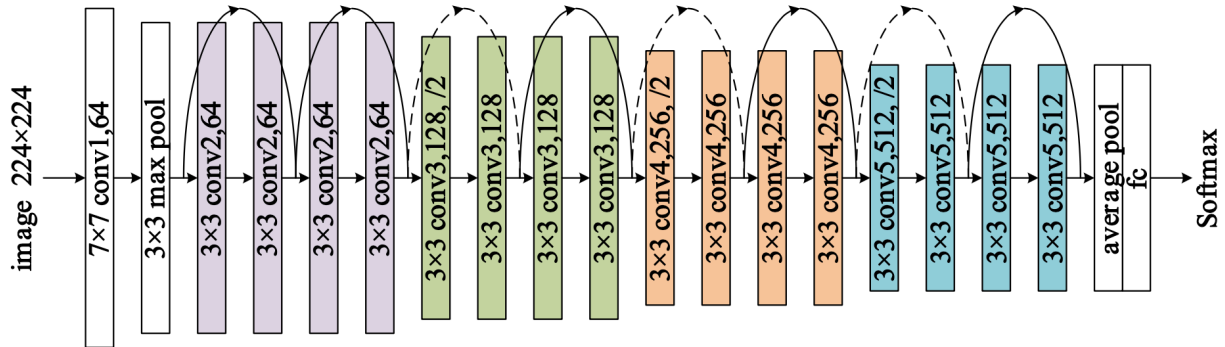
### 2.3.3 Architecture RESNET

Comme mentionné à la section 2.3.2, les performances du DCNN VGG cesse de s'améliorer lorsque la profondeur du réseau excède 19 couches, là où ses capacités de généralisation commencent à se détériorer. Une des raisons qui explique ce phénomène est qu'il est difficile pour un réseau très profond d'apprendre la fonction de représentation  $F(x)=y$  permettant de relier l'entrée de réseau  $x$  (une image) à la sortie  $y$  (une étiquette) en raison du problème du gradient qui s'éteint (*vanishing gradient problem* en anglais) lors du processus de rétro-propagation. He *et al.* (2015) ont mis en lumière ce problème de dégradation de performance lorsque la profondeur du réseau dépasse un certain seuil et s'y sont attaqués en introduisant dans les DCNN l'apprentissage résiduel. De manière générale, cette approche permet au réseau d'approximer graduellement, couche par couche, la fonction  $F(x)$  plutôt que directement sur toute la profondeur du réseau. En supposant une fonction de représentation sous-jacente  $H(x)$  (*underlying mapping* en anglais), cette technique permet au réseau d'apprendre la fonction résiduelle  $F(x) = H(x) - x$  plutôt que directement  $H(x)$ . Pour ce faire, un saut de connexion permet à l'entrée  $x$  d'être directement ajoutée à  $F(x)$  à la fin du bloc convolutif (**Figure 2**). Nous avons donc  $y = F(x) + x$ . L'intuition derrière cette innovation est que pour permettre au DCNN d'atteindre son objectif, c'est-à-dire que l'entrée égale la sortie ( $x = y$ ), celui-ci n'a qu'à apprendre la représentation résiduelle  $F(x) = 0$  ce qui est plus facile à optimiser que la représentation originale  $H(x)$ .



**Figure 2.** Un bloc convolutif résiduel dans un RESNET (figure adaptée de He *et al.*, 2015). L'entrée  $x$  est directement ajoutée à  $F(x)$  par l'intermédiaire d'un saut de connexion (flèche courbe).

Plusieurs versions du RESNET existent et leur nomenclature fait référence au nombre de couches comprises dans le réseau (e.g. RESNET-34 compte 34 couches). C'est le RESNET-152 (He *et al.*, 2016), vainqueur du ILSVRC2015 et du *Microsoft Common Objects in COntext* 2015 (MS-COCO, [Lin *et al.*, 2014]), qui a popularisé ce type d'architecture qui est encore une référence à ce jour (Wightman *et al.*, 2021). Le RESNET-18, qui compte environ 11 millions de paramètres, est présenté à la **Figure 3**.



**Figure 3.** Architecture du RESNET-18 tirée de Ou *et al.* (2019). Les sauts de connexion sont représentés par les flèches courbes.

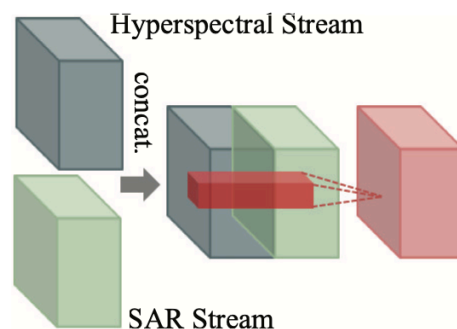
### 2.3.4 Architectures multi-entrées de réseau de neurones

Une particularité de l'IQBR a des conséquences importantes sur la façon dont on peut aborder le développement d'un DCNN qui en fait la détermination. En effet, cet indice est calculé selon la superficie relative occupée par les composantes de la BR *sur une longueur de rive variable*. En observant le carré de couleur rouge à la Figure 1, on peut comprendre qu'une image couvrant cette surface de 10 m par 10 m (100 m<sup>2</sup>) de bande riveraine ne représente pas la valeur calculée d'IQBR de 49,36, mais bien une valeur de 43,0 puisqu'elle est entièrement couverte d'une zone de coupe forestière (poids de 4,3 dans l'IQBR). Une seule image n'est donc pas suffisante pour représenter une valeur d'IQBR mesurée sur l'ensemble d'une bande riveraine. Un réseau de neurones doit être en mesure d'apprendre sur un ensemble d'images de la bande riveraine, ce qui nécessite une architecture capable de prendre plusieurs images en entrée pour une seule étiquette en sortie. Cette section sera donc consacrée à la description de ce type d'architecture.

Plusieurs architectures de DCNN multi-entrées sont décrites dans la littérature, chacune d'entre elles conçues pour des objectifs bien particuliers. Ce type d'architecture est souvent utilisé pour la fusion de caractéristiques (*feature fusion* en anglais) qui implique la combinaison de deux ou



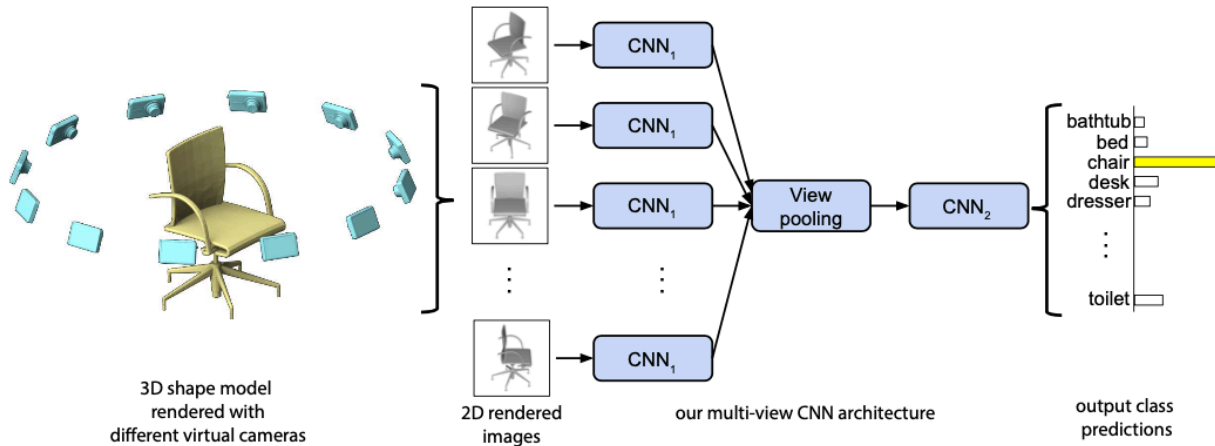
plusieurs caractéristiques extraites des images en un seul vecteur. Ce dernier contient plus d'information discriminative pour décrire la scène ou la classe que les vecteurs en entrée pris séparément (Chaib *et al.*, 2017). Dans (Penatti *et al.*, 2015), une simple concaténation des vecteurs de caractéristiques en sortie des architectures Caffe et OverFeat entraînées pour la classification du jeu de données UCMerced améliore considérablement les résultats. Xu *et al.* (2018) ont, quant à eux, proposé un modèle de DCNN à deux branches pour l'utilisation conjointe d'images hyperspectrales et de LiDAR pour des tâches de classification. La complexité du réseau est plus grande, mais le principe est semblable; les caractéristiques générées par les deux branches sont concaténées à l'intérieur du réseau, avant le classificateur. Hu *et al.* (2017) ont également utilisé une architecture à deux branches en entrée, mais cette fois pour la fusion d'images hyperspectrales (160 bandes spectrales) et de données PolSAR (radar polarimétrique à synthèse d'ouverture, de l'anglais *Polarimetric Synthetic Aperture Radar*) (12 bandes dérivées) pour une tâche de classification de scènes urbaines. En raison de la disparité du nombre de bandes utilisées pour ces deux types d'images, les auteurs ont jugé qu'une simple concaténation risquait d'occulter l'apport des données PolSAR. Ils ont donc utilisé une couche convolutive à la suite de la superposition des caractéristiques apprises par les deux branches, ce qui force le réseau à apprendre une correspondance entre ces différentes couches de caractéristiques (voir **Figure 4**). Ce type d'architecture à plusieurs branches ou par concaténation de réseaux est particulièrement approprié lorsque les données en entrée sont de provenances différentes, mais peut s'avérer lourd dans le cas contraire. Il faudrait effectivement entraîner huit réseaux en parallèle pour permettre une entrée simultanée de huit images, ce qui exigerait beaucoup de ressources de calcul.



**Figure 4.** La fusion de caractéristiques telle que décrite par Hu *et al.* (2017). Suite à la concaténation des caractéristiques générées par les branches traitant les données PolSAR et hyperspectrales indépendamment (en vert et gris), une couche convolutive (en rouge) apprend les correspondances entre ces couches.

Il existe d'autres méthodes pour intégrer les caractéristiques de plusieurs images à la fois dans un DCNN. Dans le domaine de la reconnaissance de forme en trois dimensions (3D), une architecture de DCNN multi-vues (MVDCNN, de l'anglais *Multi-View DCNN*) est proposée par Su *et al.* (2015). L'idée principale est de synthétiser l'information comprise dans plusieurs images en un seul descripteur compact de la forme 3D. L'architecture originale est constituée essentiellement du DCNN VGG-M (Chatfield *et al.*, 2014), une version moins évoluée du VGG présenté au **Tableau 1**, auquel les chercheurs ont ajouté une couche de mise en commun des vues (nommée *view pooling* dans Su *et al.* [2015]) juste après la dernière couche convolutive (la 5<sup>e</sup>). Pour l'entraînement du modèle, des images à tous les 30° (12 au total) sont prises par des caméras virtuelles situées autour d'un modèle 3D (voir **Figure 5**). Elles sont ensuite passées une par une dans la première partie du réseau (CNN<sub>1</sub> sur la **Figure 5**) et les descripteurs générés pour chacune sont superposés juste avant la couche de mise en commun des vues. Toutes les branches de la première partie du réseau se partagent donc les mêmes poids. Le principe de cette couche est très similaire au *max pooling* (Hutchison *et al.*, 2010) fréquemment utilisé dans les DCNN, excepté que l'opération n'est pas effectuée sur la même dimension. Une fois cette agrégation complétée, les caractéristiques (*features* en anglais) poursuivent leur chemin dans le reste du réseau (CNN<sub>2</sub> sur la **Figure 5**) afin de constituer le vecteur de caractéristiques compact de l'objet en 3D.

Cette même architecture a également été utilisée sur le jeu de données SketchClean composé d'environ 20 000 esquisses dans 160 catégories. L'intention ici était de tester les performances du MVDCNN en utilisant des transformations d'images (6 dans le cas présent), similaire à de l'augmentation de données, à titre de vues multiples. Les résultats sont légèrement supérieurs à l'utilisation seule de VGG-M avec des techniques d'augmentation de données standard (rotations et renversements par exemple). Ces résultats montrent que la méthode d'agrégation de caractéristiques via le *view pooling* dans le MVDCNN fonctionne aussi très bien en dehors du contexte des modèles 3D et que la représentation compacte qui en découle est riche en information descriptive. L'avantage principal de cette architecture est sa flexibilité puisque le nombre d'images en entrée de réseau ne fait pas partie de sa structure; il peut varier selon le choix de l'utilisateur. D'autre part, il demeure léger puisqu'un seul DCNN est utilisé de manière séquentielle.



**Figure 5.** L'architecture MVDCNN proposée par Su *et al.* (2015). À gauche, le système de prise de vues sur l'image en 3D. Les images passent une à une dans le CNN<sub>1</sub> et les descripteurs générés sont agrégés par la couche de *view pooling*. Le reste passe dans le CNN<sub>2</sub> afin de constituer le vecteur de caractéristiques compact.

### 3. Objectifs et hypothèses de recherche

Ce projet vise le développement d'une nouvelle méthode pour le suivi de l'état des bandes riveraines en milieu agricole qui est accessible et applicable à l'échelle régionale. Pour répondre à la problématique précédemment abordée, l'objectif principal de cette recherche est d'évaluer la capacité des DCNN pour la détermination de l'IQBR directement à partir d'images satellites THR. Les objectifs spécifiques sont les suivants:

1. Le développement d'un modèle de détermination de l'IQBR reposant sur un MVDCNN adapté;
2. L'évaluation de l'impact du choix des bandes spectrales et de la stratégie d'entraînement, c'est-à-dire le *fine tuning* d'un réseau préentraîné et l'entraînement à partir de zéro, sur les résultats de la régression.

Les hypothèses émises pour cette recherche sont :

- A. Il est possible de déterminer l'IQBR d'une bande riveraine en milieu agricole à l'aide d'un DCNN et d'une image satellite THR, avec une précision satisfaisante (RMSE inférieur à 16\*);
- B. L'utilisation simultanée de plusieurs images le long de la bande riveraine en entrée de réseau améliore les résultats par rapport à l'utilisation d'un DCNN standard à une seule image en entrée;

C. Le choix des bandes spectrales et de la stratégie d'entraînement a un impact important sur les résultats de la régression.

\*Le niveau de précision souhaité à l'hypothèse A repose sur l'objectif que l'écart moyen de prédiction de l'IQBR ne dépasse pas une classe d'IQBR (voir section 2.1). La moyenne (arrondie vers le bas) des plages de valeurs des classes d'IQBR est de 16 (soit les classes "très faible", "faible", "moyen", "bon" et "excellent" comprenant respectivement 4, 19, 20, 20 et 20 valeurs d'IQBR). Un RMSE de 16 a donc été considéré comme le seuil minimal acceptable pour une détermination sans erreur de l'IQBR dans la pratique.

#### 4. Article: Riparian Strip Characterization in Agricultural Environments Using Multi-View Deep Convolutional Neural Network (MVDCNN) and Satellite Imagery

L'article scientifique présentant la méthodologie et les résultats de ce projet de maîtrise a été soumis à la revue *Remote Sensing* le 2 novembre 2021. Les numéros de chapitre et de figures ont été adaptés à la forme du présent document. La preuve de soumission est disponible à l'annexe 1.

**Samuel de la Sablonnière**<sup>1,\*</sup>, **Samuel Foucher**<sup>1</sup>, **Yacine Bouroubi**<sup>1</sup>, **Philippe Vigneault**<sup>2</sup> and **Étienne Lord**<sup>2</sup>

<sup>1</sup> Department of Applied Geomatics, University of Sherbrooke, 2500 de l'Université Blvd., Sherbrooke, QC J1K 2R1, Canada; [samuel.de.la.sablonniere@usherbrooke.ca](mailto:samuel.de.la.sablonniere@usherbrooke.ca) (S.D.), [Yacine.bouroubi@usherbrooke.ca](mailto:Yacine.bouroubi@usherbrooke.ca) (Y.B.), [samuel.foucher@usherbrooke.ca](mailto:samuel.foucher@usherbrooke.ca) (S.F.)

<sup>2</sup> Saint-Jean-sur-Richelieu Research and Development Centre, Agriculture and Agri-Food Canada, 430 Goin Boulevard, Saint-Jean-sur-Richelieu, QC J3B 3E6, Canada; [etienne.lord@agr.gc.ca](mailto:etienne.lord@agr.gc.ca) (E.L.), [philippe.vigneault@agr.gc.ca](mailto:philippe.vigneault@agr.gc.ca) (P.V.)

\* Correspondence: [dels1605@usherbrooke.ca](mailto:dels1605@usherbrooke.ca); Tel.: +1-514-647-6037

**Abstract:** Given the ecological functions that they perform, vegetated riparian buffers are solutions for dealing with non-point source pollution in agricultural areas. Yet, they are subject to substantial anthropogenic pressures that justify monitoring their condition. Characterization of riparian buffers is frequently conducted using *in situ* measurements of quality indices, which mobilizes major resources when a vast territory must be covered and limits recurrence of monitoring. The main objective of this research, therefore, proposes an approach for characterizing riparian buffers in agricultural areas that is accessible and easily applicable at the regional scale. It is based upon recent advances in image classification accomplished through the development of deep convolutional neural networks (DCNN) and making use of very high spatial resolution, multi-spectral images from the Pleiades satellite constellation. Given their narrow and irregular shapes, riparian strips cannot be covered by a single image without incorporating external elements. Consequently, a modified multi-view DCNN architecture (MVDCNN) allowed the use of several images as input to the network, which was trained to correlate with the riparian strip quality index (RSQI). Different numbers of input images, seven spectral band combinations, and two training modes were evaluated; the latter included training from scratch and fine tuning of a pre-trained network. The MVDCNN architecture incorporating a pre-trained RESNET-18 provided the best correlation between riparian strip images and their RSQI, with an average RMSE = 7.35 and  $R^2 =$

0.932 using the RGB spectral bands. The use of four images as the network input during training also improved the average RMSE by 3.3% (7.35 vs. 7.6), compared to using a single image. The modified MVDCNN architecture could prove to be invaluable for remote sensing applications where different number of images needs to be taken into account for a classification task.

**Keywords:** riparian buffers; remote sensing; convolutional neural networks (CNNs); environmental assessment; agriculture; multi-view, Pleiades.

#### 4.1 Introduction

In agricultural environments, the intensive use of nutrients such as nitrogen and phosphorus in crop production compromises aquatic ecosystem structure and function [1]. Several beneficial management practices have been implemented to reduce these diffuse sources of pollution, including crop rotation, conservation of plant residues after harvesting, and the establishment of grassed swales. Riparian buffer strips, which are permanent natural vegetation strips along the edges of watercourses [2], are a frequently proposed solution to these problems [3,4]. Although variable in terms of its degree of effectiveness [5], the riparian buffer is the last physical barrier (i.e., “last line of defense”) against downslope losses from nearby soils to the receiving waterbodies. In addition to retaining nutrients and trapping sediment, this buffer establishment provides stabilization of the banks along the watercourse, together with its thermoregulation, as well as biodiversity ecoservices [6–9]. Worldwide, riparian buffer strips are subject to anthropogenic disturbances, including the adverse effects of urban sprawl, agricultural expansion, and the construction of levees and dams [10]. Characterization of these ecotones (ecosystem transition zones) and monitoring of their condition, therefore, is essential for providing an accurate assessment of their capacity to perform their ecological functions and to enable decision-makers to establish appropriate management and protection plans [11–13].

Existing approaches for characterizing riparian strips are generally based upon the measurement of quality indices, such as the IBI (Index of Biological Integrity) [14], MQI (Morphological Quality Index) [15] or QBR (*Qualitat del bosc de ribera*, in Catalan) [16], which have been developed specifically for this purpose. These indices include several indicators of ecosystem health [17], many of which are closely related to land-cover characteristics, such as the identification of plant communities, forest cover and channel slope. In Quebec (Canada), the Riparian Strip Quality Index (RSQI, described in section 2.2) [7] is used for this characterization. Overall, this index is a

weighted sum of the areas that are occupied by different land-use components of the riparian zone, such as forest, herbaceous plants, crops and built-up areas.

The measurement of these indices requires the collection of field data, expert aerial photo-interpretation or the production of detailed maps of the different plant species covering the territory. This is costly in terms of time, human resources and, therefore, financial resources. In addition, application at the regional scale requires local expertise, which makes the process not very accessible and limits the recurrence of the monitoring on an annual basis.

As demonstrated by [18], the use of remote sensing for characterizing riparian vegetation cover, which facilitates regional-scale applicability, has greatly increased in importance over the past decade. This is largely due to the reduced cost of acquiring Earth observation images, their increased availability, and improved spatial and spectral resolutions. Various remote sensing products like LiDAR (Light Detection and Ranging) data, radar and multi-spectral images are used for delineating riparian vegetation [19], determining plant species composition [20], and characterizing vegetation structure [11], among others. However, for operational purposes, some studies seek to establish more direct links between *in situ* characterization methods of riparian vegetation (i.e. measurement of indices) and remote sensing images.

In that sense and to allow an assessment of riparian buffer strips on a regional scale, researchers in Andalusia (Spain) [21], for example, correlated the Green Vegetation Fraction [22], which is calculated from Advanced Very High Resolution Radiometer (AVHRR) images, against the ecological status of the riparian buffer (favorable versus unfavorable) in an agricultural environment. Yet, the results that were obtained are not very precise due to the spatial resolution of the images (1 km) and the use of only two classes. In the Cantabria region of Northern Spain [23], the Riparian Quality Index (RQI) [24] was correlated with land-use maps that had been produced from aerial photographs that had been interpreted by vegetation experts. The best result had an  $R^2 = 0.6$ , but the method also required a high level of expertise. In the Portneuf region of Quebec [25], determination of RSQI was performed using object-based classification (OBC) of WorldView-2 images (50 cm resolution for the panchromatic band, 2 m for the 8 multispectral bands). The RSQI that was obtained by their classifier was an average value of 56 against 55 for the 80 *in situ* measurements. The authors of this study demonstrated that satellite images with very

high spatial resolution ( $< 1$  m) contain the necessary information for the determination of an index such as the RSQI.

In recent years, Deep Convolutional Neural Networks (DCNN) have been applied to classification, scene recognition, segmentation and object detection tasks on remote sensing images, given the excellent performance of this type of machine learning [26–28]. DCNNs do not require the intervention of experts to define the attributes of objects or classes that are found in an image. They generate by themselves, during the learning process, from the low-level features (e.g., textures, edges) to the highest levels (e.g., branches of a tree) [29]. In contrast, the complete training of a neural network requires a large amount of data [30]. This prerequisite is rarely met in the field of Earth observation; the use of a dataset that is not sufficiently large results in model overfitting [31]. Transfer Learning (TL) techniques [32] are generally used to mitigate this effect by using large datasets such as ImageNet [33] and PASCAL-VOC [34] as a training base.

In most cases, a DCNN learns from a label that is associated with an image, either a category (e.g., corn or wheat) or a numerical value. Characterization indices, on the other hand, are measured over strips of land varying in length from a few tens of meters to more than one km. Thus, a single image is not sufficient to represent the riparian strip and its index (the label).

Several methods are available for simultaneously integrating features from multiple images into a DCNN at once. This type of architecture is often used for feature fusion, which involves combining two or more features that are extracted from the images into a single feature vector. These architectures can also make use of multi-branch networks, such as for the joint use of hyperspectral and LiDAR images [35] or PolSAR [36]. They can also use concatenation of entire networks, as in [37]. These architectures have the advantage of being able to integrate multi-source data. Yet, in the case where images from the same source are used, such as from the same satellite sensor, the training of several networks in parallel would seem superfluous in terms of resources.

In the field of three-dimensional (3D) shape recognition, a multi-view DCNN-based architecture (MVDCNN) has been proposed [38]. Its main principle is to synthesize the information included in multiple images into a single compact descriptor. The architecture consists essentially of the VGG-M DCNN [39], to which the researchers have added a view- pooling layer immediately following the last convolutional layer. During training, different images of the same object pass through an encoding network sequentially and the descriptors that are generated for each of these



images are accumulated up to the view-pooling layer. The principle of this layer is very similar to a max pooling layer that is frequently used in DCNN, except that the operation is not performed on the same dimension. The main advantage of this architecture is its flexibility. Since the number of images at the network input is not part of its structure, it can vary according to the user's choice. It also remains computationally light in resources since only one DCNN is used sequentially.

This research, therefore, proposes a novel approach for the characterization of riparian buffers that is applicable over large areas. It is based on recent advances in image classification that have been brought about by the development of DCNN and takes advantage of the accessibility of satellite images that are of very high spatial resolution ( $< 1$  m). For this purpose, an MVDCNN architecture originally developed for 3D object recognition was adapted to correlate images of a riparian strip that were extracted from satellite imagery with the RSQI. The specific objectives of this study were: (1) to adapt the MVDCNN architecture for the determination of the RSQI; (2) to test the performance of the network as a function of the spectral bands used; and finally, (3) to compare the performance of the different modes of training DCNN. The results will also be compared to an object-based land cover classification and RSQI subsequent calculation. To our knowledge, this is the first direct application of DCNNs to the problem of riparian buffer (RB) characterization.

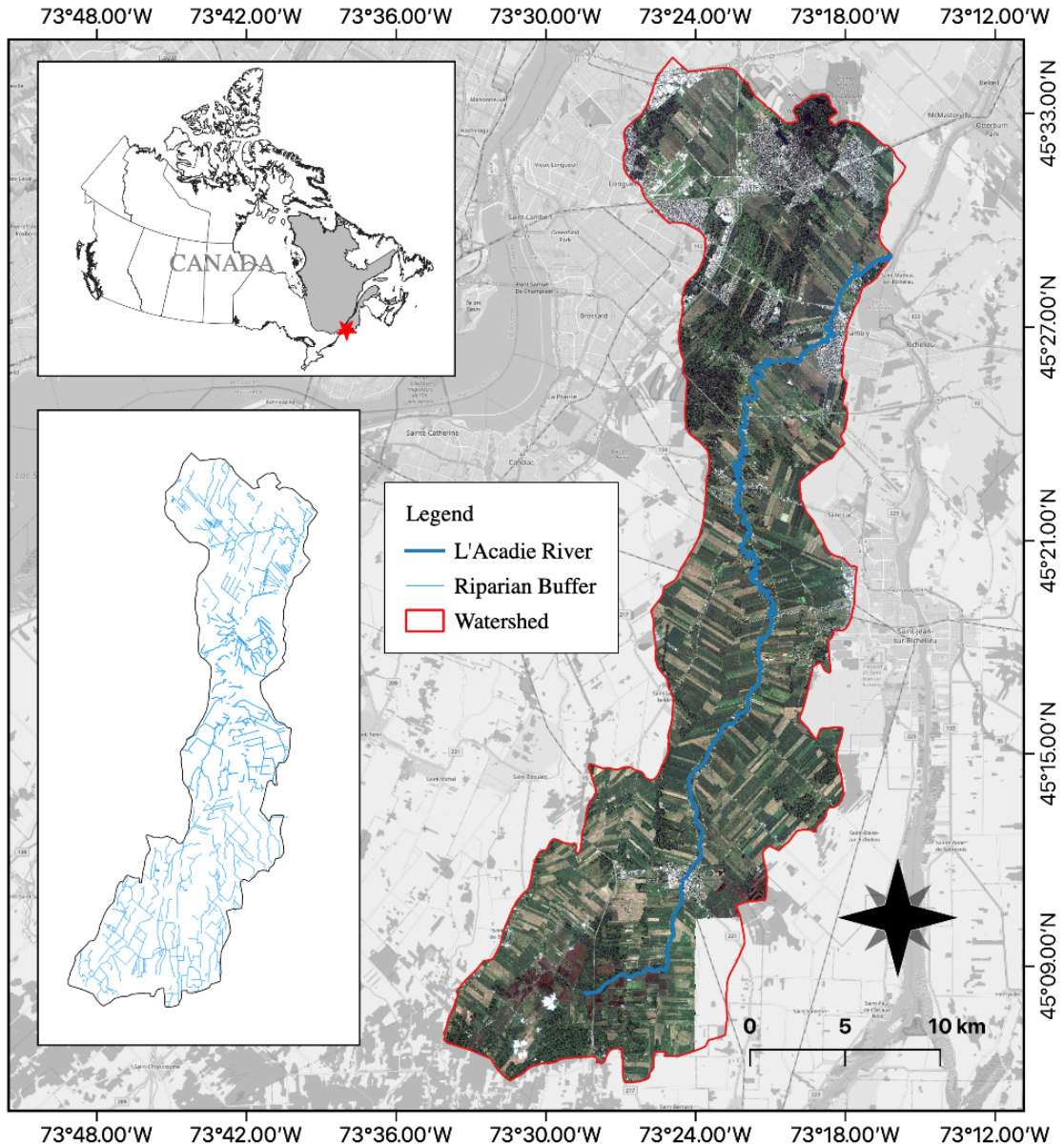
## **4.2 Materials and Methods**

### *4.2.1 Study area*

The area that was studied is the watershed of the L'Acadie River, which is a sub-basin of the Richelieu River, and is located in southern Quebec (Canada) (**Figure 6**). Its surface area of 562.5 km<sup>2</sup> makes it the most important sub-basin of the Richelieu River. The L'Acadie River is 82 km long, with its source near the border with the United States (New York State), south of the village of Hemmingford, QC. The river ends its course as a tributary merging with the Richelieu River, north of the Chambly Basin.

The Acadie River watershed is largely dominated by agriculture, which accounts for 67% of its surface area [40]. The main crops grown in the watershed are maize (*Zea mays*) and soybean (*Glycine max*), which account for about 75% of the agricultural land allocated to them. The wooded and urban sectors are few in number and account for 17% and 10% of the territory, respectively. The terrain is almost flat, except in the extreme north of the watershed, where Mont-Saint-Bruno rises to 218 m.

The methodology used relies on the ability of DCNN to discriminate riparian habitats according to their land-cover composition and to estimate their RSQI, in order to facilitate the monitoring of the state of these habitats. Two datasets are required for training the DCNN: 1) a ground truth dataset and 2) satellite imagery. These are presented following the detailed description of the RSQI.



**Figure 6.** Pleiades mosaic covering the L'Acadie River Watershed, Quebec (Canada), and riparian buffer strips in vector format.

#### 4.2.2 Riparian Strip Quality Index (RSQI)

The RSQI is a weighted sum of the areas occupied by the land-use components of the RS (riparian buffer strips) where the greatest weights are assigned to the surface types contributing most to the health of the watercourse. It assesses "the ecological condition of the riparian habitat and its impact on the integrity of the aquatic environment" [7]. It is defined according to the following equation:

$$\text{RSQI} = \sum (\%i \times P_i) / 10, \quad (1)$$

where  $\%i$  is the percent ground cover of the  $i^{\text{th}}$  component of the riparian buffer, and  $P_i$  is the weighting factor for that component. A width of 10 m of riparian buffer is considered in the calculation of the areas over a length that depends upon the bank sections being determined. More specifically, if a significant change in land use is observed, e.g., from urban to forest, a new section is delineated. The weighting factors are: forest (10.0); shrub (8.7); herbaceous (5.8); forest cut (4.3); rocky soil (3.8); pasture/agricultural grassland (3.0); crop (1.9); infrastructure (1.9); and bare soil (1.7). The RSQI index values, therefore, range from 17 to 100, with the highest value representing the best condition in terms of the capacity of the riparian buffer to fulfil its ecological functions. The RSQI can be classified as follows: very poor (17-21); poor (21-40); average (40-60); good (60-80); and excellent (90-100) (COVABAR, 2015).

In agricultural areas, a riparian buffer with a very low RSQI is characterized by a virtual absence of natural vegetation along the watercourse (**Figure 7a**), while a certain width of vegetated buffer would be present for a "poor" class (**Figure 7b**). In the "average" class, trees and natural vegetation are present to a greater extent along the RB, but these are often sparse and accompanied by built-up areas (**Figure 7c**). High values represent riparian buffers where natural forb and forest cover dominate (**Figure 7d, e**). This assessment protocol has been widely adopted by the entities responsible for assessing the state of riparian ecosystems in the Province of Quebec (Canada). Although it is measurable through the use of photo-interpretation, the RSQI is measured directly, in practice, in the field by experts.



**Figure 7.** Examples of riparian strips in a Pleiades image for different values and classes of RSQI: (a) 19 (very poor), (b) 45 (poor), (c) 65 (average), (d) 79 (good), and (e) 100 (excellent). The footprint of the images that was extracted from the Pleiades mosaic is represented by the white squares. Each square image is centered 10 m from the watercourse, and has dimension of 23 m / 46 pixels for each side.

#### 4.2.3 Field data

The data from the characterization of the tributaries of the L'Acadie River, which were collected in the field between 2012 and 2015, were provided by a watershed protection organization. They represent 697 km of watercourse that was travelled (see riparian strips in vector format in **Figure 6**). The riparian buffer strips are divided into homogeneous sections of varying length, based upon compositional changes that were observed in the field. The parameters needed to calculate the RSQI were estimated by the evaluators for each of the riparian areas. The vector data, which designate the watercourses, are based on public mapping that is provided by the governments of Quebec and Canada.

#### 4.2.4 Satellite images

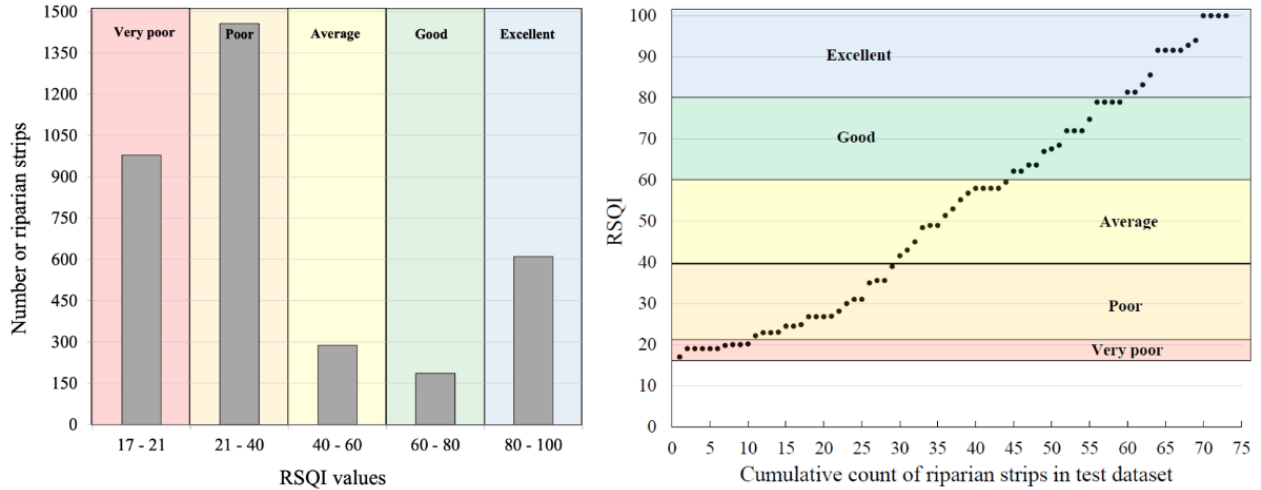
Three Pleiades satellite images were used to provide coverage of the area, i.e., two from 6 July 2013 and one from 8 July 2015. The spectral bands that were used are Blue (B, 430-550 nm), Green (G, 490-610 nm), Red (R, 600-720 nm) and Near infrared (N, 750-950 nm) with a spatial resolution of 2 m, together with a panchromatic band (480-830 nm) with a resolution of 0.5 m. The images were resolution enhanced by the PCI Geomatics Catalyst (version 2222.0.4) PANSHARP

algorithm and then orthorectified using a 1-m resolution digital elevation model (DEM) that was derived from LiDAR data. The radiometric resolution was reduced from 12-bit to 8-bit to facilitate training of the neural networks. Following pre-processing, a mosaic of the study region in RGB-NIR format at 0.5-m resolution was produced (see **Figure 6**).

#### *4.2.5 Automated dataset creation*

Manual verification of the dataset and images resulted in the removal of 11.4% of the RSQI values, which were found to be erroneous due to discrepancies that were observed between the images and the RSQI estimation. Following this validation step, training and test datasets were created using Python scripts. The extracted images have a size of 46 pixels square (23 m) with a centre distance of 10 m from the stream (see **Figure 7**). The annotation was performed automatically by retaining the corresponding RSQI value from the vector data. The size and location parameters of the images were determined empirically to provide sufficient contextual information to the DCNN, while avoiding the inclusion of objects that were too far from the riparian strip. In total, 46,209 images were sampled, which were grouped into 3,489 riparian strips. **Figure 8a** shows the distribution of the RSQI indices that were evaluated in the field dataset.

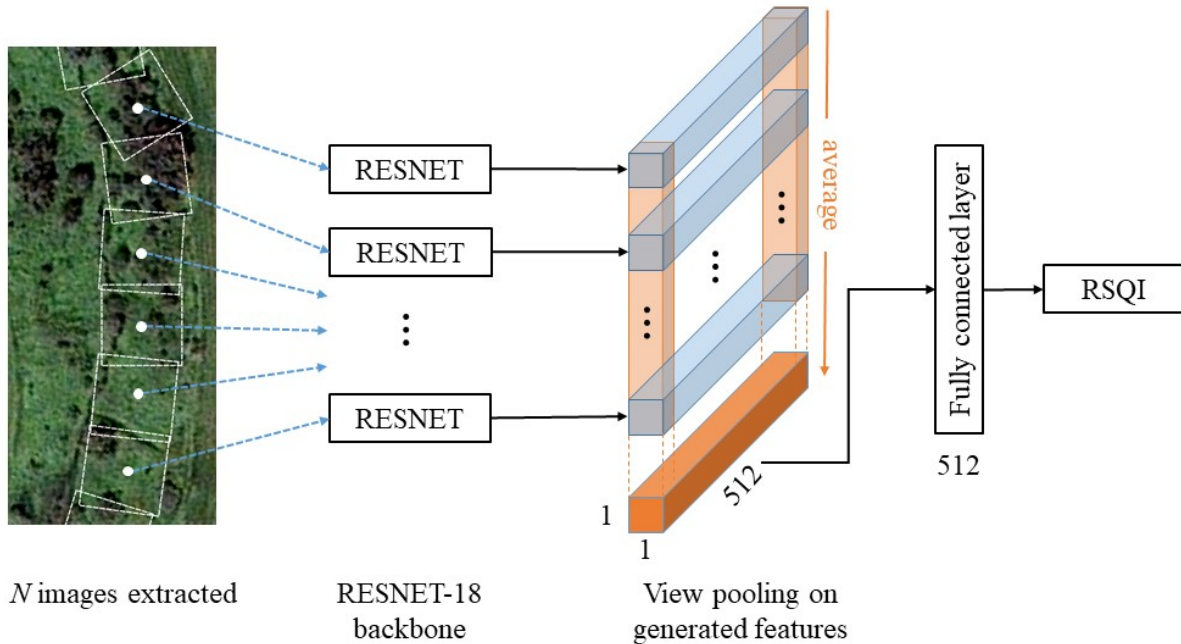
To provide a balanced and curated test dataset on which to evaluate the models, 75 riparian strips were randomly selected from across the mosaic, ensuring that the full range of RSQI values was covered and removed from the training and validation steps (see **Figure 8b**). Particular attention was paid to the reliability of RSQI, with respect to what was observable on the image. Furthermore, to properly isolate the test set from the training set, the riparian buffers of both river banks were retained in the test set when they were of similar land-cover composition. The relatively small size of this test dataset is due to the importance of keeping the test set balanced in terms of RSQI values and retaining a sufficient number of riparian strips for the training phase in each RSQI class (see **Figure 8**). As detailed in Section 2.7, another 20% is used during validation, leaving approximately 70% riparian strips available for the training step.



**Figure 8.** Distribution of RSQI values in the entire dataset ( $N = 3,489$  RS), in a) according to quality class and in b) distribution of the riparian strips composing the test dataset ( $N=75$  RS).

#### 4.2.6 MVDCNN architecture

In this paper, an architecture based on MVDCNN [38] was repurposed and is presented in **Figure 9**. In our case, the VGG-M network was replaced by the RESNET-18 architecture [41] due to the latter lower number of parameters and superior classification performance [41]. Since RSQI is a weighted average, the view-pooling layer operator that was used is the average rather than the maximum value used in the originally described MVDCNN. A dropout layer (50%) [42] was also added between the view pooling and the fully connected layer to reduce overfitting. During training, the RS images are, in no particular order, sequentially entered into the RESNET-18 backbone, and averaged over all of these features by the pooling layer (**Figure 9**). The generated features are then passed into the fully connected layer, composed of 512 fully connected nodes, from which the RSQI prediction is the output.



**Figure 9.** MVDCNN architecture that was used for the determination of RSQI, as inspired by [33]. A number ( $N$ ) of RGB images, that were extracted from the Pleiades mosaic, are used as input to the MVDCNN. The averaging operation is performed in the view-pooling layer. The output is a RSQI predicted value .

#### 4.2.7 Network training

For the MVDCNN network training, specific numbers of images of the riparian buffer strips, i.e., 1, 2, 4, 8, 12 and 16 were evaluated as input to the network to predict the RSQI value. The same training set was therefore used but each of its riparian strips, which may consist of several dozen images, is represented by a specific number of images (1-16). For the purposes of comparison, note that the use of a single input image corresponds to the training of a standard RESNET-18, i.e., without a view-pooling layer. Training was performed for this evaluation using a MVDCNN incorporating a pre-trained RESNET-18 with an image dataset composed of the RGB spectral bands. Once determined, this optimal number of input images was used for all subsequent training.

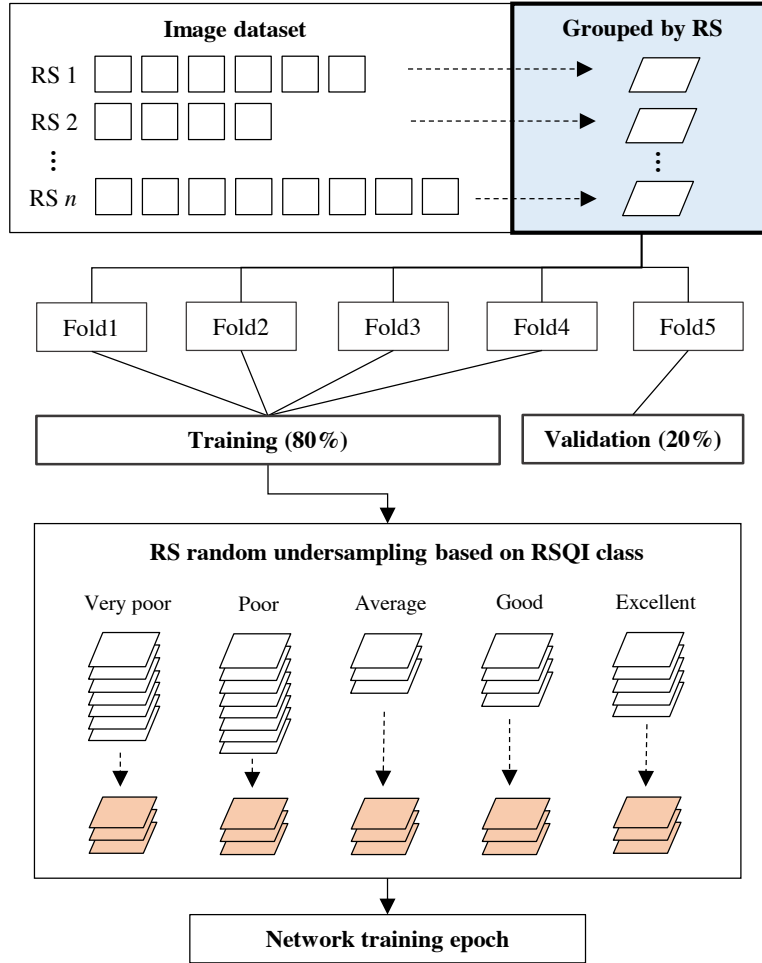
Training was performed with a 5-fold cross-validation protocol (see **Figure 10**). The dataset was randomly split, therefore, into five folds containing between a total of 599 and 634 RS. For each of the five training cycles, four folds (80%) were reserved for training and one fold for validation (20%). Random undersampling was employed to address the problem of imbalance in the distribution of the RSQIs (see **Figure 8a**). To take advantage of the entire dataset, random sampling of RS was also employed at each training epoch, as well as sampling of images from each of the

RS (i.e. the 1-16 images per riparian strip used). For all network training, data augmentation techniques such as image rotation ( $0-45^\circ$ ) and image flipping were employed. **Figure 10** illustrates the preparation of the image dataset for network training.

The validation dataset was used to verify the accuracy of the selected hyperparameters and to evaluate the performance of the MVDCNN during training. In order to reduce the randomness of the validation phase (performed at each training epoch), all images of each of the riparian strips (as opposed to the value determined in training) are used as network inputs. This meant that the MVDCNN made predictions on a variable number of input images, depending upon the length of the riparian strip. The network that performs best, according to the training epoch, was selected for the test phase. No image augmentation techniques were used for this last phase.

Different combinations of spectral bands were evaluated: RGB, RGN and RGBN for training from scratch; and RGB and RGN for fine tuning the MVDCNN incorporating a pre-trained RESNET-18 on ImageNet. Tests on repeated monochrome bands (e.g., R-R-R) were also performed for both training techniques, which allowed us to evaluate the information that the MVDCNN was able to derive from each of the spectral bands. The loss function used for gradient backpropagation was the mean square error (MSE) and the optimizer was Adam [43]. The learning rate used for training from scratch was 0.001 and 0.0001 for fine tuning and the number of training epochs was set to 60 for both training strategies. All experiments were performed using the the Pytorch 1.5.0 library using python v3.7. All simulations and training was done on a computer with a 3.5 GHz Intel Core i7-5930k processor, 64 GB of RAM and a NVIDIA GeForce GTX 1080 graphics card.





**Figure 10.** Flowchart used in preparing the image dataset for network cross-validation training. First, the images are grouped by riparian strip (RS). Second, the set of riparian strips is split into 5-folds and the preliminary training set is formed. Third, undersampling of RS is performed according to each RSQI class, which constitutes the training dataset upon which the network learns.

#### 4.2.8 Network testing

The test dataset consisting of 75 randomly distributed riparian strips was used to evaluate the performance of the networks. The performance metrics that were employed are the root-mean-square error (RMSE) and the coefficient of determination  $R^2$ . The RMSE is the average error between values predicted by the model and the values observed while the  $R^2$  coefficient is the proportion of the variance in the dependent variable that is explained by the independent variables in the regression model. For each of these metrics, the mean and standard deviation of five runs of cross-validation are presented. Results on the validation fold are also presented, in addition to results on a balanced variant of this same fold (in terms of RSQI classes), which allows a more accurate assessment of the model performance in validation. To further evaluate the behaviour of

the trained model when presented with a variety of situations, two additional tests were imposed, namely: 1) the prediction of the RSQI for a randomly generated RS from images contained in the test set, and 2) the prediction of the RSQI as a function of the starting position of the extracted images relative to a simulated stream. In the first case, inferences are made on a random selection of distinct riparian buffer strip images from the test dataset. This test assesses the ability of the network to account for inter-image variability. For the second test, sensitivity to intra-image variability is assessed by modulating the percentage of forest and agricultural land found in each of the images composing a riparian buffer.

#### *4.2.9 Object-based classification*

In order to compare the performance of the developed model with a more traditional approach, we computed the RSQI of the riparian strips in our dataset using a method relying on object-based land cover classification similar to that used by Novoa *et al.* [25] To do so, we used the Object Analyst module from PCI Geomatics Catalyst (version 2222.0.4). For the segmentation step, a scale parameter of 30 was used while the shape and compactness parameters were 0.2. These parameters allow for the successful delineation of elongated objects, such as the herbaceous edges present along the riparian strips. In total, 186,936 objects were segmented within a 50 m buffer around the water streams in the field dataset.

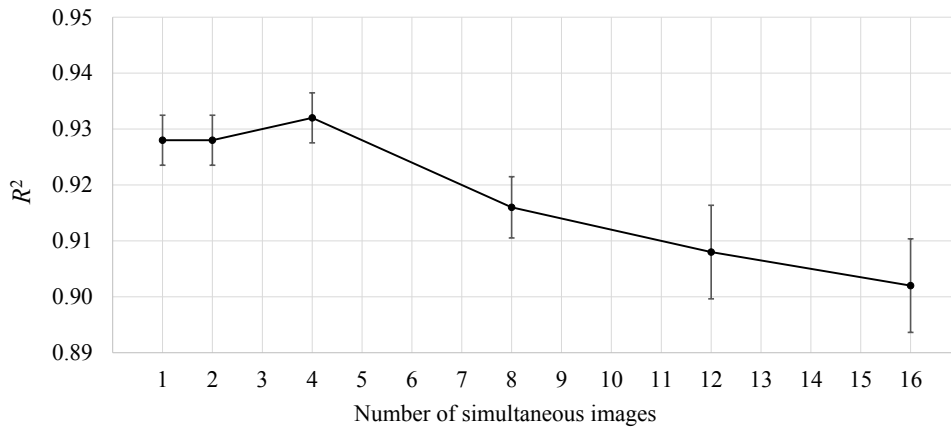
Following this segmentation step, an attribute calculation was performed in order to extract the statistical information of each of the objects created. Twenty-two attributes were retained for this step, namely the min/max values and the standard deviation of the 4 spectral bands of the images, in addition to the contrast and entropy layers generated for this purpose. The geometric attributes of elongation, circularity, compactness and convexity were also considered. For the classification step itself, objects were manually selected as training and validation sites for each of the land cover classes considered in the RSQI (see section 2.2). Since the "forest cut" and "rocky soil" classes were absent from the study site, they were not retained. A "other" class was however added to represent water and the inland shoreline. Finally, a support-vector machine (SVM) classifier was used to classify the segmented objects. A confusion matrix and the kappa coefficient ( $k$ ) were calculated to evaluate the agreement between the ground truth and the prediction of the classification on the validation objects.

For the calculation of the RSQI, a 10 m wide riparian buffer was considered (Section 2.2). For every riparian buffer strip of the dataset, the number of pixels corresponding to each of the land cover classes within this buffer was determined and the RSQI value was calculated using Equation 1. Note that pixels in the “other” class were not considered in the calculation. These RSQI values were compared to those predicted by the proposed model.

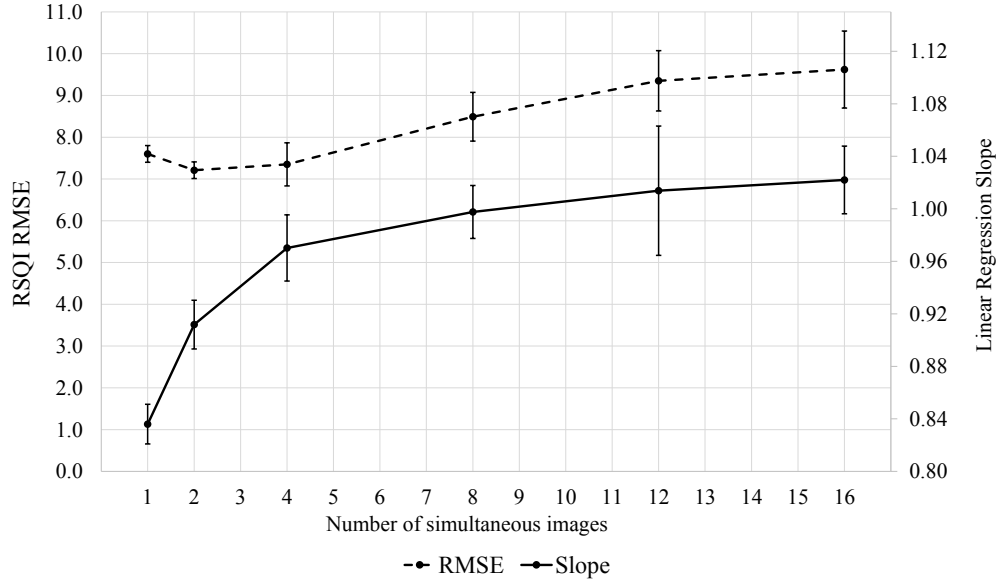
## 4.3 Results

### 4.3.1. Determining the optimal number of input images

During training, different numbers of simultaneous input images to the network were evaluated to determine the optimal value. In other words, we varied the number of images that represent each of the riparian strips in the dataset before sending them together as input to the network. The results are shown in **Figure 11** and **Figure 12**.



**Figure 11.**  $R^2$  results of RSQI prediction on the test set as a function of the number of input images (1 to 16). Best results were obtained with a set of 4 images per riparian strip. Training was performed with the RGB spectral bands with 5-fold cross-validation. Error bars represent standard deviations of five independent trainings of the MVDCNN incorporating a RESNET-18 that was pre-trained on ImageNet.



**Figure 12.** RMSE (Root-Mean-Square Error) and regression slope ( $b$ ) obtained on the test set as a function of the number of simultaneous input images. The most balanced results (low RMSE and regression slope  $b$  near 1.0) was obtained with 4 input images. Training was performed with the the RGB spectral bands. Error bars represent standard deviations of 5 independent trainings of the MVDCNN incorporating a RESNET-18 that was pre-trained on ImageNet.

When comparing RMSE of the predicted RSQIs that were obtained as a function of the number of input images, we noted that the latter varied from  $7.21 \pm 0.20$  for 2 input images to  $9.62 \pm 0.92$  for 16 input images, with respective  $R^2$ -values of 0.928 and 0.902. RMSE and  $R^2$  that were obtained using a single input image are respectively  $7.6 \pm 0.20$  and 0.928. In contrast, its regression slope is the furthest from the perfect 1:1 slope (i.e.,  $b = 1.0$ ) among all of the results, with  $b = 0.836$  for a single image versus  $b = 0.970$  for the use of four images. This would indicate that predictions of low RSQI values (17 to 21) would be overestimated, while high RSQI values (85 to 100) are underestimated. The most balanced tradeoffs for RMSE and slope are obtained with 4 input images, i.e., with estimates of  $7.35 \pm 0.52$  and  $b = 0.970$ , respectively. Moreover, this configuration also yields both the best average  $R^2$  with 0.932 and the best RMSE with 6.95 among the 5 cross-validation runs. We retained this 4-image configuration for all further experiments

#### 4.3.2 Effect of training strategy and choice of spectral bands

Two training strategies were evaluated to validate the use of transfer learning, namely: 1) training from scratch an MVDCNN incorporating RESNET-18 with randomly initialized weights and 2) the fine-tuning of the top two layers of the RESNET-18 pre-trained on ImageNet. In addition, 7

spectral band combinations were tested. RMSE and  $R^2$  that were obtained on the validation fold ("as is" and balanced) and the test dataset are shown in Table 2 and Table 3, respectively.

As can be seen in Table 2, all configurations for training from scratch have obtained very similar results on the test dataset when comparing the average RMSEs that are attained, which vary between  $8.18 \pm 0.31$  for the GGG bands and  $8.94 \pm 0.58$  for the RGBN bands. The results obtained with a single repeated spectral band (RRR, GGG, BBB and NNN) are superior to those using compound spectral bands (RGB, RGN, RGBN). In general, these results tend to show that the network does not depend upon diverse spectral information to converge. The same observation can be made in Table 3 where the best  $R^2$  is obtained with the GGG bands (0.908) and the worst with the joint use of the four RGBN spectral bands (0.892). Yet, the use of RGB spectral bands obtains a better  $R^2$  (0.90) than with the RGN bands (0.892), which was not expected, given that the B-band is generally not the most informative about the vegetation characteristics. Besides the above-mentioned details, it can be said that while trained from scratch the network obtains good results regardless of the spectral bands used.

With respect to fine-tuning of a pre-trained RESNET-18, the RMSEs that were obtained on the test dataset range from  $7.35 \pm 0.52$  and  $7.76 \pm 0.24$  for the respective RGB and RGN bands to  $9.10 \pm 0.20$  for the NNN bands. Results obtained with the use of the RGB and RGN spectral bands outperform all configurations that use repeated spectral bands. We also obtained the best average  $R^2$ -values (0.932 and 0.924, respectively; Table 3) with these same spectral bands. These results are also superior to those obtained while training from scratch. In contrast, the RMSE of  $8.83 \pm 0.29$  for training from scratch with the RGB bands resulted in an improvement of 16.8%. When compared to the best results that are obtained in training from scratch in all configurations (RMSE = 8.18, GGG bands), this is an improvement of about 10%. The features that were already learned by the pre-trained RESNET-18, therefore, are useful for the RSQI prediction task and perform better than training from scratch. Use of a pre-trained network improves the generalizability of the network for our experimental setup. Lastly, the same thing can be said here as with training from scratch, where good results were obtained regardless of the spectral bands used. On the other hand, the results obtained with the RGB and RGN bands turn out to be the overall best and those with the NNN bands the worst. For all subsequent tests in this paper, this fine-tuned trained network with the RGB spectral bands and 4-image input was used.

**Table 2.** From the five cross-validation trainings, average RMSE and standard deviation of RSQI predictions on the validation, balanced validation and test datasets for the two training modes and 7 spectral band combinations. Results from the RSQI calculation using the OBC are also presented. Best results for both training methods are shown in bold.

Spectral bands	From scratch			Fine-tuned <sup>2</sup>		
	Validation	Balanced validation	Test	Validation	Balanced validation	Test
RRR	11.49 ± 0.68	12.71 ± 0.61	8.56 ± 0.28	11.95 ± 0.52	12.64 ± 1.11	8.02 ± 0.36
GGG	11.58 ± 0.52	13.02 ± 0.80	<b>8.18 ± 0.31</b>	11.72 ± 0.61	13.03 ± 0.64	8.27 ± 0.30
BBB	11.22 ± 0.37	12.69 ± 0.61	8.35 ± 0.26	11.67 ± 0.44	12.82 ± 1.06	8.04 ± 0.37
NNN	11.79 ± 0.63	12.93 ± 1.13	8.37 ± 0.22	11.73 ± 0.45	12.70 ± 0.93	9.10 ± 0.20
RGB	11.68 ± 0.77	12.66 ± 0.29	8.83 ± 0.29	11.69 ± 0.75	12.56 ± 1.23	<b>7.35 ± 0.52</b>
RGN	11.63 ± 0.85	12.65 ± 0.61	8.86 ± 0.52	11.39 ± 0.59	12.57 ± 0.75	7.76 ± 0.24
RGBN	11.28 ± 0.82	12.68 ± 0.31	8.94 ± 0.58			
OBC classifier	13.65 ± 0.53	15.11 ± 0.64	11.25 <sup>1</sup>			

<sup>1</sup>In the case of the OBC on the test set, only the best model was used for the RSQI determination. <sup>2</sup>The fine-tuned network was only used with combinations of 3 spectral-bands.

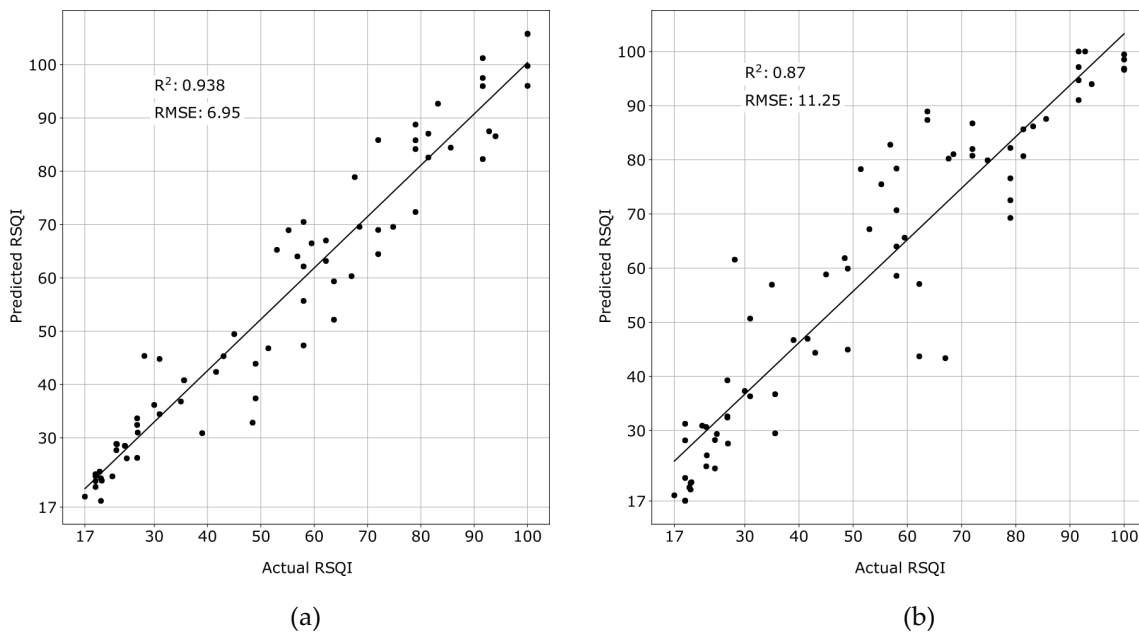
**Table 3.** From the five cross-validation trainings, average R2 values and standard deviation of RSQI predictions on the validation, balanced validation and test datasets for the two training modes and 7 spectral band combinations. Results from the RSQI calculation using the OBC are also presented. Best 1In the case of the OBC on the test set, only the best model was used for the RSQI determination. 2The fine-tuned network was only used with combinations of 3 spectral-bands.

Spectral bands	From scratch			Fine-tuned <sup>2</sup>		
	Validation	Balanced validation	Test	Validation	Balanced validation	Test
RRR	0.846 ± 0.018	0.798 ± 0.021	0.898 ± 0.008	0.840 ± 0.012	0.795 ± 0.030	0.916 ± 0.009
GGG	0.836 ± 0.018	0.790 ± 0.023	<b>0.908 ± 0.008</b>	0.838 ± 0.013	0.791 ± 0.012	0.908 ± 0.005
BBB	0.848 ± 0.015	0.801 ± 0.016	0.902 ± 0.008	0.842 ± 0.016	0.797 ± 0.027	0.914 ± 0.009
NNN	0.836 ± 0.021	0.795 ± 0.028	0.906 ± 0.006	0.838 ± 0.016	0.799 ± 0.025	0.890 ± 0.007

RGB	$0.840 \pm 0.019$	$0.804 \pm 0.011$	$0.900 \pm 0.007$	$0.842 \pm 0.022$	$0.804 \pm 0.033$	<b><math>0.932 \pm 0.005</math></b>
RGN	$0.842 \pm 0.022$	$0.803 \pm 0.017$	$0.892 \pm 0.013$	$0.844 \pm 0.015$	$0.805 \pm 0.018$	$0.924 \pm 0.006$
RGBN	$0.848 \pm 0.020$	$0.803 \pm 0.015$	$0.892 \pm 0.008$			
OBC	$0.795 \pm 0.008$	$0.732 \pm 0.027$	$0.871^1$			

<sup>1</sup>In the case of the OBC on the test set, only the best model was used for the RSQI determination. <sup>2</sup>The fine-tuned network was only used with combinations of 3 spectral-bands.

**Figure 13a** shows the result of the predictions made by the best model, i.e., the best iteration in cross-validation with the RGB bands and 4 network input images, for each of the riparian buffers of the test dataset. RSQIs of less than 30 are estimated by the model with excellent accuracy, with RMSE = 5.26. Predictions for values between 30 and 80 are more variable, with an RMSE of 8.09. An RMSE of 6.11 is obtained for values from 80 to 100.



**Figure 13.** RSQI predictions of the best MVDCNN based model (a) and the results obtained from the OBC (b) as a function of the index values that were measured in the field. Each of the points represents a prediction of the RSQI from the set of images of a riparian strip of the test dataset ( $N=75$ ).

#### 4.3.3 Object-based classification

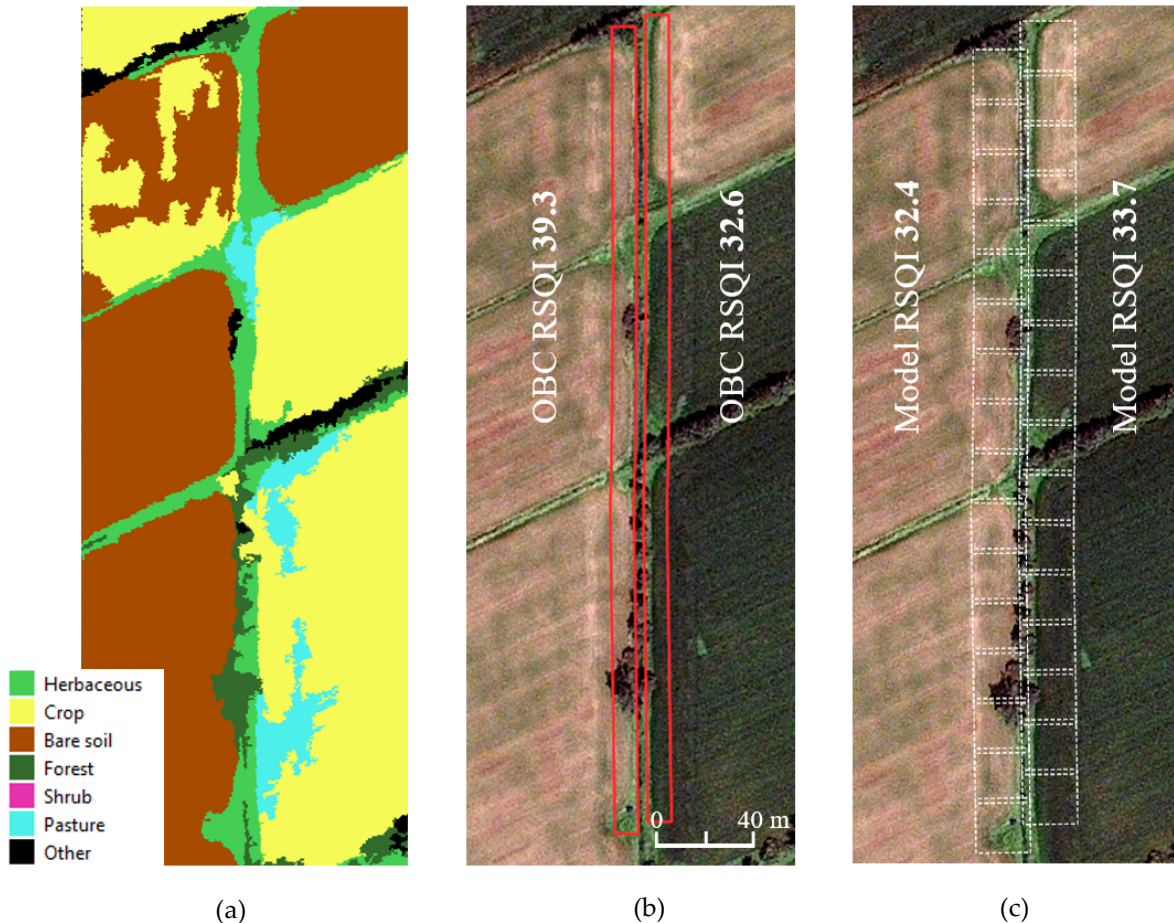
For the object-based classification, the overall accuracy obtained for the 8 land cover classes is 73% on 100 validation objects, with an overall k coefficient of 0.68 (supplementary Table 1). More

importantly, the forest, herbaceous, pasture and crop classes obtained k coefficients of 0.79, 0.82, 0.74 and 0.57 respectively. These are the classes most present in the first ten meters of riparian strip width, those considered in the calculation of the RSQI. We can also note the excellent producer's accuracy of 95 % and 93.3 % of the crop and forest classes. On the other hand, the “shrub” class was not correctly classified but is sparse in the dataset. It can also be pointed out that despite a k coefficient of 0.52, the “bare soil” class is most often mixed with infrastructure and crop classes, which have nearly identical weights in the RSQI (1.9, 1.7 and 1.9 respectively), which reduces the impact on RSQI determination.

An example of classification can be seen in **Figure 14a**, where, for these two riparian strips of the test dataset, we can see that some of the agricultural land is incorrectly assigned to the “pasture” class (in cyan). In this specific case, this misclassification results in a small difference (32.6 instead of 30.6 if correctly classified) on the calculated RSQI value (see **Figure 14b**). Natural herbaceous plants, found very often in the form of long and thin objects on the images, are generally well identified (light green). Their discrimination from crops is crucial since the weighting factors are very different here (5.8 versus 1.9 for crops). In general, a visual analysis of the results showed a good classification quality, especially for the dominant land-cover classes in the dataset (i.e. forest, herbaceous, pasture and crop classes).

The results obtained from the RSQI calculation using the OBC are shown in Table 2 and Table 3. Results on the 75 riparian strips of the test set, an  $R^2$  of 0.87 and a RMSE of 11.25, are displayed in **Figure 13b**. **Figure 14** displays a representative example of RS from the test dataset where the OBC approach is compared to our MVDCNN model. In this example, the average RSQI obtained using OBC for the two riparian strips is 35.95. In contrast, our trained MVDCNN (**Figure 14c**) predicted an average RSQI of 33.05 which is closer to the validated field data which resulted in an RSQI of 26.8 for both riparian strips.

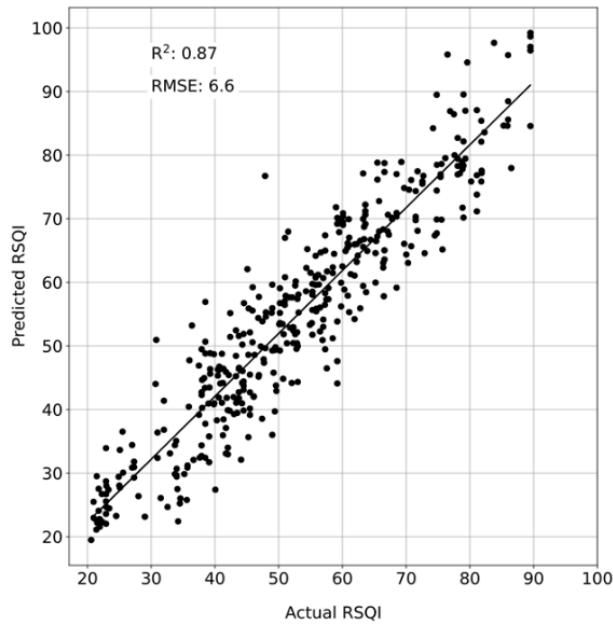




**Figure 14.** Representative object-based classification and RSQI predicted values. In (a) the result of the OBC on the Pleiades image, (b) the result of the RSQI calculation from the OBC and the 10 m buffer zone considered in this calculation (in red), and (c) result of RSQI predictions from the best MVDCNN based model and footprints of the images used to make those predictions (white squares). Field measured RSQI was determined as 26.8 for both stream banks.

#### 4.3.4 Testing the model on simulated riparian strips

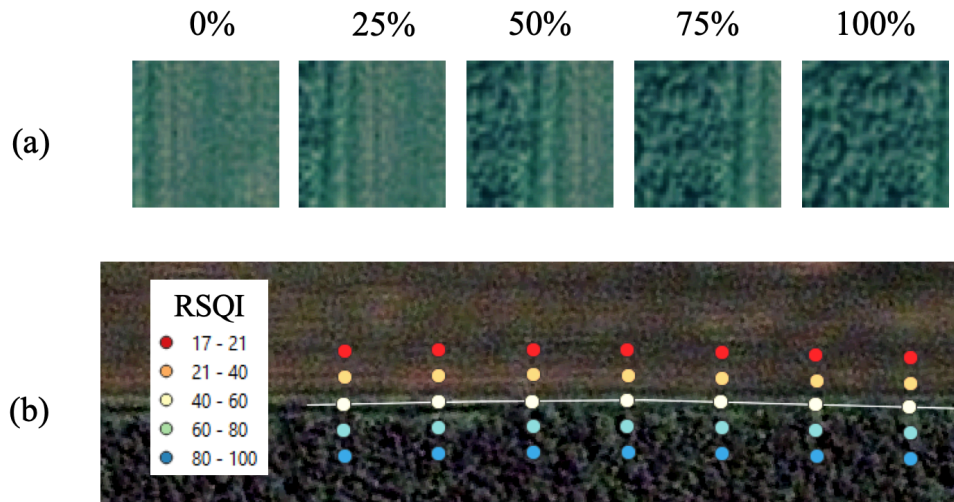
A series of cross-validation runs were performed to measure the model response to inter-image variability in land cover. For this simulation, four images of two different riparian strips from the test set, for a total of 8 images, were used as network inputs. The ground-truth RSQI value is the average of the values of these two riparian strips. The images constitute a randomly determined set using all combinations of RSQI classes, e.g., very poor-poor (RSQI 17-21, RSQI 21-40), poor-average (RSQI 17-21, RSQI 40-60), poor-good (RSQI 17-21, RSQI 60-80), and so on. The result of 400 simulations is shown in **Figure 15**.



**Figure 15.** Model predictions when 4 images of two riparian strips are mixed ( $N = 400$ ). The RSQI value used as the reference is the average of these two riparian strips.

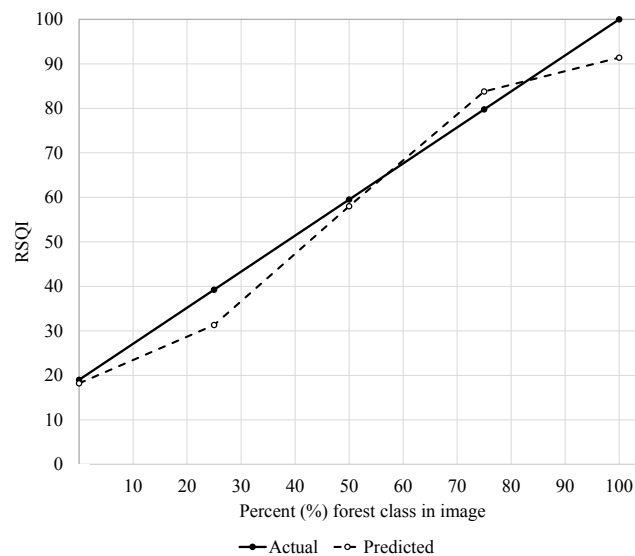
As shown, in **Figure 15**, the model can take into account the individual characteristics of a set of images to make its predictions, even with a series of images not from the same RS, with an  $R^2$ -value that is only 0.06 (0.87 obtained vs 0.93 on actual test dataset) lower, but an RMSE that is 0.35 (6.6 obtained vs 6.95 on actual test dataset) lower than the original selected test set.

The response of the model to intra-image variability in land cover was also assessed. For this purpose, a pseudo-riparian strip was delineated as a vector at the border of forest and agricultural land. The images that were extracted from the latter, whose center is represented by each of the dots in **Figure 16b**, are composed of a defined percent of land-use of the forest and agricultural land classes. An estimated RSQI value was calculated from Equation 1. The same exercise was conducted by shifting the location of the riparian buffer to one side in order to obtain images containing 100%, 75%, 50%, 25% and 0% forest area. The calculated land-use of the crop is equal to 100% minus the forest area. The presence of these two land-use categories in the images is representative of a large portion of the riparian strips in the dataset that was used (data not shown), which justified their choice for this test. An example of the resulting images is shown in **Figure 16a**; the result of the predictions is shown in **Figure 17**.



**Figure 16.** Example of a dataset that is used to evaluate a gradient of RSQI. In a) Images composed of different percentages of forest (%) and agricultural land (%). In b) The riparian buffer at the border of forest and crop is presented in vector format (white line). Each of the points represents the centre of an extracted image.

Again, the developed MVDCNN model is sensitive to intra-image variation of land cover that is represented by forest and agricultural land. The prediction errors of the RSQI (RMSE) vary from 8.6 for 100% forest to 0.8 for 0% forest, which remains close to the errors that were observed on the test set.



**Figure 17.** Five predictions of RSQI that are made by the model as a function of the percentage of forest that was found in each of the images to which it was presented. The solid line represents RSQI that was calculated from Equation 1.

#### 4.4 Discussion

A multi-input architecture was used in this study in order to obtain the best prediction of the RSQI using multiple images. We evaluated the input of 1 to 16 simultaneous images in the network during training to determine the optimal number. Results obtained with a single input image (see **Figure 11**), which is equivalent to using a standard RESNET-18, show that this DCNN architecture allows the network to learn appropriate features. This success, using a single image, can be explained by the homogeneous nature of riparian buffer strips in agricultural areas. Yet, the regression slope that is obtained with a single image is less accurate, and does not allow the model to cover the entire range of values of the RSQI as well as when several images are used. The MVDCNN architecture using four input images obtained the best results, with a lower RMSE by 0.25 (or 3.3%) on the test set and a regression slope that more closely follows the linear variation in the RSQI index range. Nevertheless, it would be interesting to analyze the results of a single-image training in heterogeneous environments, such as partially populated areas, with rocky ground surfaces or in the presence of extensive deforestation.

The training results using different spectral bands provided insight into the features considered by MVDCNN for this correlation task. For example, the use of the single spectral B (blue) or G (green) band, when training from scratch, performed better than with any other combination of spectral bands, even with the joint use of all four available bands (see Table 2 and 3). This result would indicate that texture, shape and pattern information are crucial. Indeed, both poor and good quality riparian strips can be covered with healthy vegetation (trees, crops, grasses), which implies similar spectral signatures. Since the differentiation of these plant types is of primary importance for the characterization of the riparian buffer, these elements of texture and shape would seem to be key. Very high spatial resolution provides this level of detail, thereby allowing the network to extract these features. The use of higher spectral resolution sensors, combined with a spatial resolution similar to that used in this research (50 cm), could potentially allow for better characterization of these vegetation features [44]. The spectral and spatial information that is needed to discern other land cover features from vegetation, such as bare soil, buildings, and roads, however, is present in all four Pleiades sensor bands [45].

The two training modes evaluated in this study (from scratch or using a pre-trained RESNET-18), showed the importance of the size and diversity of the training dataset that is being used. Although

the image categories in the ImageNet dataset do not resemble the image set that was used in this study, the features that were learned by RESNET-18 (when trained on the latter) are still relevant. Thus, an architecture that is pre-trained on these datasets is adaptable to remote sensing images [37,46]. Indeed, features that are generated in the first layers tend to be similar from one network to another and from one type of image to another. It is these shallow features, i.e., the most generic and often analogous to edge or color detection filters, which are easily transferable. This is not the case for deeper features, which are much more complex and representative of the original dataset. In fact, a strategy that is often used is to keep the shallow features intact and to perform learning only on the higher layers [29,37,47,48]. It is this strategy that allowed the researchers in [29] to obtain the best results on the UCMerced [49] and RS19 [50] satellite image datasets, compared to training their network architecture from scratch. We had similar results in this study, where fine-tuning of the pre-trained RESNET-18 achieved the best results for four of the six spectral band combinations that were tested (see Table 3), and by margins of 16.8% and 12.4%, respectively, when using the RGB and RGN bands. The fine-tuning of the top two layers of the pre-trained network leverages the features that are learned on a large dataset, while allowing the network to adapt to the present task. The tendency to overfit is also reduced and the network is forced to find a solution with fewer adjustable parameters. Although the dataset used for this study is of considerable size in terms of number of images, it is not large and diverse enough to allow the network trained from scratch to match the results obtained from a pre-trained network. Only results including the Near infrared (N) band are really superior to those obtained with fine-tuning (RMSE 8.4% lower), which can be explained by the absence of this spectral band in the ImageNet dataset. In light of these results, the use of a network pre-trained on a large Earth observation dataset, such as Eurosat [51] and GeoImageNet [52], could take advantage of learned features more closely related to the images used here (as opposed to ImageNet), especially with respect to the N-band.

Evaluation of the errors in the test riparian buffer strip images revealed some limitations of the trained models. First, riparian strips with an RSQI between 30 and 80 were more difficult to predict from the model (see **Figure 13a**). Such responses may be attributable to the smaller number of different examples that are seen by the network for these values (30 to 80) during training (see **Figure 8a** for the distribution of the data set). Yet, it is the riparian strips for which the land cover is the most heterogeneous, which is more complex to evaluate by both the *in situ* evaluator and the model. Second, the model in some cases tends to overestimate RSQI due to the presence of trees.

This is an inherent limitation of using satellite images, given their top-down (or bird's eye) perspective. Due to their canopies, large trees can hide a riparian strip that is composed almost entirely of grass or crops, which subsequently is not observable on the images. Because of their potential to penetrate the forest canopy, LiDAR and Radar data could be used to reduce the uncertainty of RSQI estimates in the presence of this type of vegetation.

A major impediment to the development of the model is the subjective nature of the *in situ* measurements of the RSQI upon which DCNN training is based [20]. The accuracy of estimate land-use for each component of the riparian buffer is effectively dependent upon the perspective of the assessor at the time of the survey campaign. For the large area that is covered in this study, field data had been collected over three separate years by several teams. The inherent subjectivity of the measurement method, combined with the use of different evaluators, brings an element of uncertainty to this dataset. A concrete example of this uncertainty is the categorization of pasture /agricultural grassland (weight of 3.0 in the RSQI) versus natural grassland (weight of 5.8). For nearly identical areas, these components were sometimes categorized as one class or the other, depending upon the evaluator. This confusion is expected since natural grasses are often an integral part of agricultural grasslands, which adds to the difficulty of interpretation. Context is therefore an important part of the evaluation. To allow a more replicable method and thus address this limitation, new characterization indices based solely on remote sensing products are being developed [17]. Another source of uncertainty in the training dataset relies in the difference between the years of image acquisition and the field campaigns (1 to 2 years apart) and between the time in the season of the field campaigns and the image acquisitions. In the end, all these sources of uncertainty induce a part of noise to the training dataset. Still, despite this limitation, the trained models were able to learn relevant features for the determination of the RSQI, which is evidenced by the results obtained on the curated test dataset being superior to those obtained on the validation dataset.

The RSQI calculation was also performed using an object-based classifier to compare the proposed model to a more traditional approach [25]. As we can see in Table 2 and Table 3, results obtained using this approach resulted in lower accuracy to those of the proposed method. Comparing the average RMSE achieved by the MVDCNN on the test dataset with those of the OBC, we note an average improvement of 3.16 (or 28.1%) while fine tuning and 2.66 (or 23.6%) while training from scratch. The same observation can be made on the validation folds, with average RMSE

improvements of around 15.5%. Despite its inferior performance, the OBC approach still achieves respectable results. However, the accuracy of its results depends upon the quality of object segmentation and on the quality of final land-cover class determinations. Moreover, this last prerequisite requires photo-interpretation skills. The RSQI calculation method, which strictly considers the defined buffer zone, is also not as flexible as the MVDCNN based method which takes into account a wider context.

The developed model was trained to predict a riparian buffer characterization index, the RSQI, directly from satellite images. Tests on simulated riparian strips (**Figure 15** and **Figure 17**) show that it could take into account variations in land cover to make its predictions and that it was particularly sensitive to the presence of trees. From a practical standpoint, the use of this model could facilitate the monitoring of riparian buffer status year after year, thereby improving the ability of decision-makers to target areas where interventions are required. A quick overview of the quality of riparian buffers in a sector that has yet to be covered by field campaigns would also be possible. It could provide an important decision support tool to assess the need for such campaigns, depending upon what is being observed. An important advantage of the developed model is the simplicity of its use. No advanced image processing or photo-interpretation skills are required, merely the basic pre-processing of satellite images. The only important manual task for the user is the determination of the riparian buffer sections. The MVDCNN architecture has this flexibility in terms of the number of input images, which does not limit its use to a specific river length. The novelty of this approach was demonstrated in this study and could therefore provide a better alternative to traditional methods requiring fieldwork.

Other avenues for the application of DCNN to the riparian zone problem could be considered, besides the use of our modified MVDCNN. An architecture using encoder-decoder architectures could be used for semantic segmentation [53,54]. Such architectures could improve the results of riparian vegetation cover classification in the method proposed by [20] or [25]. A methodology like the one used in [55], to generate riparian buffers indicators from LiDAR point clouds, could also benefit from deep learning classification methods designed for this type of data [56].

#### **4.5 Conclusions**

This research studied the use of a MVDCNN architecture for characterizing riparian buffer strips in agricultural areas for remote sensing monitoring at a regional scale. To our knowledge, this is

the first direct application of DCNN to this problem, which is the main contribution of this study. Results show that the MVDCNN architecture incorporating a RESNET-18 manages to establish good correlations between images of the riparian strip and its RSQI. The combination of spectral bands for training the networks did not prove to be crucial since all of them resulted in comparable results. However, the use of a pre-trained network was found to be superior to training from scratch. In addition to being more accessible in terms of required skills, this approach also achieved better correlation results than the traditional RSQI computation using object-based classification of satellite images.

Due to their narrow and irregular shapes, riparian strips cannot be covered by a single image without incorporating external elements. The choice of the MVDCNN architecture, therefore, was motivated by the need to use several input images with a single index value. This architecture and the use of four input images improved the average RMSE that was obtained on the test set, as opposed to the use of a single image. Furthermore, the MVDCNN remained flexible with respect to inferences that were being made, given that its architecture is not limited to a specific number of images. The second contribution of this study is the use of an architecture that was initially developed for 3D object classification in a new context, namely, the characterization of objects on remote sensing images.

In the future, this type of MVDCNN architecture could be compared to other types of image series in agriculture, in addition to the Long Short-Term Memory networks (LSTM) [57,58] that are used for temporal image analysis.

**Author Contributions:** Conceptualization, S.D., S.F., P.V. and E.L.; methodology, S.D. and S.F.; software, S.D.; formal analysis, S.D. and E.L.; supervision, Y.B., S.F. and E.L.; funding acquisition, Y.B. and E.L.; paper review and editing, Y.B., S.F., S.D. and P.V. All authors agreed to the published version of the manuscript.

**Funding:** This research was funded by Agriculture and Agri-Food Canada, Mitacs and the Natural Sciences and Engineering Research Council of Canada under Grant to S. de la Sablonnière. This project was financed in part by the Ministère de l'Économie et de l'Innovation (MEI) of the province of Québec.



**Acknowledgments:** The authors thank Marcel Comiré of COVABAR for generously providing the RSQI field data and to Effigis Geo-Solutions for providing the satellite imagery. We also thank Louis Longchamps now at Cornell University, for the initial conception of this research project.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Carpenter, S.R.; Caraco, N.F.; Correll, D.L.; Howarth, R.W.; Sharpley, A.N.; Smith, V.H. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* **1998**, *8*, 559–568, doi: 10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2.
2. Doyon, N. Fiche - La Bande Riveraine, Développement durable, Environnement et Parcs Québec, Available online: <https://www.cobali.org/wp-content/uploads/2016/11/La-bande-riveraine.pdf> (accessed on 22 October 2021).
3. de Sosa, L.L.; Glanville, H.C.; Marshall, M.R.; Abood, S.A.; Williams, A.P.; Jones, D.L. Delineating and mapping riparian areas for ecosystem service assessment. *Ecohydrology* **2018**, *11*, e1928, doi: 10.1002/eco.1928.
4. Stutter, M.I.; Chardon, W.J.; Kronvang, B. Riparian buffer strips as a multifunctional management tool in agricultural landscapes: Introduction. *Journal of Environmental Quality* **2012**, *41*, 297–303, doi: 10.2134/jeq2011.0439.
5. Gagnon, É.; Gangbazo, G. Efficacité des bandes riveraines : analyse de la documentation scientifique et perspectives. *ministère du Développement durable, de l'Environnement et des Parcs* **2007**, *Direction des politiques de l'eau*, 17.
6. de Sosa, L.L.; Glanville, H.C.; Marshall, M.R.; Pryor Williams, A.; Jones, D.L. Quantifying the contribution of riparian soils to the provision of ecosystem services. *Science of the Total Environment* **2018**, *624*, 807–819, doi: 10.1016/j.scitotenv.2017.12.179.
7. Saint-Jacques, N.; Richard, Y. Développement d'un indice de qualité de la bande riveraine : application à la rivière Chaudière et mise en relation avec l'intégrité biotique du milieu aquatique. In *Le bassin de la rivière Chaudière: l'état de l'écosystème aquatique*; Ministère de l'Environnement et de la Faune: Québec (Canada), **1998**; p. 6.1-6.41.
8. Schuft, M.J.; Moser, T.J.; Wigington, P.J.; Stevens, D.L.; Mcallister, L.S.; Chapman, S.S.; Ernst, T.L. Development of Landscape Metrics for Characterizing Riparian-Stream Networks. *Photogrammetric Engineering and Remote Sensing* **1999**, *65*, 1157–1167.
9. Naiman, R.J.; Décamps, H. The Ecology of Interfaces: Riparian Zones. *Annual Review of Ecology and Systematics* **1997**, *28*, 621–658, doi:10.1146/annurev.ecolsys.28.1.621.
10. Poff, B.; Koestner, K.A.; Neary, D.G.; Henderson, V. Threats to riparian ecosystems in western North America: An analysis of existing literature. *JAWRA Journal of the American Water Resources Association* **2011**, *47*, 1241–1254, doi: 10.1111/j.1752-1688.2011.00571.x.
11. Dufour, S.; Bernez, I.; Betbeder, J.; Corgne, S.; Hubert-Moy, L.; Nabucet, J.; Rapinel, S.; Sawtschuk, J.; Trollé, C. Monitoring restored riparian vegetation: How can recent developments in remote sensing sciences help? *Knowledge & Management of Aquatic Ecosystems* **2013**, *10*, doi: 10.1051/kmae/2013068.

12. Ministère de l'Environnement et de la Lutte contre les changements climatiques Suivi de La Qualité Des Bandes Riveraines Available online: [http://www.environnement.gouv.qc.ca/eau/eco\\_aqua/suivi\\_mil-aqua/bandes-riv.htm](http://www.environnement.gouv.qc.ca/eau/eco_aqua/suivi_mil-aqua/bandes-riv.htm) (accessed on 28 July 2020).
13. Ashraf, S.; Brabyn, L.; Hicks, B.J.; Collier, K. Satellite remote sensing for mapping vegetation in New Zealand freshwater environments: A review. *New Zealand Geographer* **2010**, *66*, 33–43, doi: 10.1111/j.1745-7939.2010.01168.x.
14. Miller, S.J.; Wardrop, D.H.; Mahaney, W.M.; Brooks, R.P. A plant-based Index of Biological Integrity (IBI) for headwater wetlands in Central Pennsylvania. *Ecological Indicators* **2006**, *6*, 290–312, doi: 10.1016/j.ecolind.2005.03.011.
15. Rinaldi, M.; Belletti, B.; Bussettini, M.; Comiti, F.; Golfieri, B.; Lastoria, B.; Marchese, E.; Nardi, L.; Surian, N. New tools for the hydromorphological assessment and monitoring of European streams. *Journal of Environmental Management* **2017**, *202*, 363–378, doi: 10.1016/j.jenvman.2016.11.036.
16. Munné, A.; Prat, N.; Solà, C.; Bonada, N.; Rieradevall, M. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR Index. *Aquatic Conservation: Marine and Freshwater Ecosystems* **2003**, *13*, 147–163, doi: 10.1002/aqc.529.
17. Allostry, J.; Varin, M.; Paranjape, M.; Théau, J.; Fournier, R. *Développement d'un indice de suivi de l'état des écosystèmes riverains (ISÉÉR) dans le Québec méridional – Revue de littérature – Définitions et indicateurs*; Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc. (CERFO) – Université de Sherbrooke, **2020**; p. 120.
18. Huylbroeck, L.; Laslier, M.; Dufour, S.; Georges, B.; Lejeune, P.; Michez, A. Using Remote Sensing to Characterize Riparian Vegetation: A Review of Available Tools and Perspectives for Managers. *Journal of Environmental Management* **2020**, *267*, 110652, doi:10.1016/j.jenvman.2020.110652.
19. Bertoldi, W.; Drake, N.A.; Gurnell, A.M. Interactions between River Flows and Colonizing Vegetation on a Braided River: Exploring Spatial and Temporal Dynamics in Riparian Vegetation Cover Using Satellite Data. *Earth Surface Processes and Landforms* **2011**, *36*, 1474–1486, doi:10.1002/esp.2166.
20. Macfarlane, W.W.; McGinty, C.M.; Laub, B.G.; Gifford, S.J. High-Resolution Riparian Vegetation Mapping to Prioritize Conservation and Restoration in an Impaired Desert River. *Restoration Ecology* **2017**, *25*, 333–341, doi:10.1111/rec.12425.
21. Ivits, E.; Cherlet, M.; Mehl, W.; Sommer, S. Estimating the ecological status and change of riparian zones in Andalusia assessed by Multi-Temporal AVHRR datasets. *Ecological Indicators* **2009**, *9*, 422–431, doi: 10.1016/j.ecolind.2008.05.013.
22. Gutman, G.; Ignatov, A. The derivation of the Green Vegetation Fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing* **1998**, *19*, 1533–1543, doi: 0.1080/014311698215333.
23. Fernández, D.; Barquín, J.; Álvarez-Cabria, M.; Peñas, F.J. Land-use coverage as an indicator of riparian quality. *Ecological Indicators* **2014**, *41*, 165–174, doi: 10.1016/j.ecolind.2014.02.008.
24. Gonzalez del Tanago, M.; Garcia De Jalon, D. Riparian Quality Index (RQI): A methodology for characterising and assessing the environmental conditions of riparian zones. *Limnetica* **2011**, *30*, 235–251.

25. Novoa, J.; Chokmani, K.; Lhissou, R. A novel index for assessment of riparian strip efficiency in agricultural landscapes using high spatial resolution satellite imagery. *Science of the Total Environment* **2018**, *644*, 1439–1451, doi: 10.1016/j.scitotenv.2018.07.069.
26. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 778–782, doi: 10.1109/LGRS.2017.2681128.
27. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *152*, 166–177, doi: 10.1016/j.isprsjprs.2019.04.015.
28. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2016**, *54*, 1349–1362, doi: 10.1109/TGRS.2015.2478379.
29. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting Convolutional Neural Networks for remote sensing scene classification. *Pattern Recognition* **2017**, *61*, 539–556, doi: 10.1016/j.patcog.2016.07.001.
30. Scott, G.J.; England, M.R.; Starns, W.A.; Marcum, R.A.; Davis, C.H. Training deep Convolutional Neural Networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 549–553, doi: 10.1109/LGRS.2017.2657778.
31. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Journal of Big Data* **2019**, *6*, 60, doi: 10.1186/s40537-019-0197-0.
32. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **2010**, *22*, 1345–1359, doi: 10.1109/TKDE.2009.191.
33. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Kai Li; Li Fei-Fei ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; Miami, FL, 20-25 June **2009**; pp. 248–255.
34. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* **2010**, *88*, 303–338, doi: 10.1007/s11263-009-0275-4.
35. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource remote sensing data classification based on Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, *56*, 937–949, doi: 10.1109/TGRS.2017.2756851.
36. Hu, J.; Mou, L.; Schmitt, A.; Zhu, X.X. FusioNet: A two-stream Convolutional Neural Network for urban scene classification using PolSAR and hyperspectral data. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE); Dubai, UAE, 6-8 March **2017**; pp. 1–4.
37. Penatti, O.A.B.; Nogueira, K.; dos Santos, J.A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Boston, MA, 7-12 June **2015**; pp. 44–51.
38. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-View Convolutional Neural Networks for 3D shape recognition. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); IEEE: Santiago, Chile, 7-13 December **2015**; pp. 945–953.
39. Raj, A.; Maturana, D.; Scherer, S. *Multi-Scale Convolutional Architecture for Semantic Segmentation*; Robotics Institute - Carnegie Mellon University: Pittsburgh, Pennsylvania, **2015**;

40. COVABAR *Rapport de Caractérisation Du Bassin Versant de La Rivière l'Acadie : Secteur Napierville*; Comité de concertation et de valorisation du bassin de la rivière Richelieu (COVABAR), Beloeil, Québec, Canada, **2015**;
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, *arXiv:1512.03385*.
42. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving Neural Networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]* **2012**.
43. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization, *arXiv:1412.6980 [cs]* **2017**.
44. Mariotto, I.; Thenkabail, P.S.; Huete, A.; Slonecker, E.T.; Platonov, A. Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HypSIRI Mission. *Remote Sensing of Environment* **2013**, *139*, 291–305, doi: 10.1016/j.rse.2013.08.002.
45. Li, M.; Stein, A.; Bijker, W.; Zhan, Q. Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *122*, 192–205, doi: 10.1016/j.isprsjprs.2016.10.007.
46. Marmanis, D.; Datcu, M.; Esch, T.; Stilla, U. Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters* **2016**, *13*, 105–109, doi:10.1109/LGRS.2015.2499239.
47. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing* **2015**, *7*, 14680–14707, doi:10.3390/rs71114680.
48. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing* **2018**, *10*, doi:10.3390/rs10071119.
49. Yang, Y.; Newsam, S. Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification. In Proceedings of the Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10; ACM Press: San Jose, California, 2010; p. 270.
50. Xia, G.-S.; Yang, W.; Delon, J.; Gousseau, Y.; Sun, H.; Maître, H. Structural High-Resolution Satellite Image Indexing. In Proceedings of the ISPRS TC VII Symposium-100 Years ISPRS; Vienna, Austria, July 5 **2010**; Vol. 38, pp. 298–303.
51. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2019**, *12*, 2217–2226, doi: 10.1109/JSTARS.2019.2918242.
52. Bouroubi, Y.; Chapdelaine, C.; Foucher, S.; Byrnes, D.; Beaulieu, M.; St-Charles, P.-L.; Germain, M.; Lauzier-Hudon, É.; Bugnet, P.; Sabo, N.; et al. GeoImageNet: A collaborative platform for deep learning application to very high resolution EO images. In Proceedings of the Canadian Symposium on Remote Sensing 2019; Fredericton, Canada, June **2019**.
53. Kemker, R.; Salvaggio, C.; Kanan, C. Algorithms for Semantic Segmentation of Multispectral Remote Sensing Imagery Using Deep Learning. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *145*, 60–77, doi:10.1016/j.isprsjprs.2018.04.014.
54. Yuan, X.; Shi, J.; Gu, L. A Review of Deep Learning Methods for Semantic Segmentation of Remote Sensing Imagery. *Expert Systems with Applications* **2021**, *169*, doi:10.1016/j.eswa.2020.114417.

55. Laslier, M.; Hubert-Moy, L.; Dufour, S. Mapping Riparian Vegetation Functions Using 3D Bispectral LiDAR Data. *Water* **2019**, *11*, 483, doi:10.3390/w11030483.
56. Xie, Y.; Tian, J.; Zhu, X.X. Linking Points with Labels in 3D: A Review of Point Cloud Semantic Segmentation. *IEEE Geoscience and Remote Sensing Magazine* **2020**, *8*, 38–59, doi:10.1109/MGRS.2019.2937630.
57. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735–1780, doi: 10.1162/neco.1997.9.8.1735.
58. Teimouri, N.; Dyrmann, M.; Jørgensen, R.N. A Novel Spatio-Temporal FCN-LSTM Network for recognizing various crop types using multi-temporal radar images. *Remote Sensing* **2019**, *11*, 990, doi: 10.3390/rs11080990.

## Supplementary

**Table 1.** Confusion matrix from the object-based land-cover classification (OBC)

	Bare soil	Herbaceous	Crop	Pasture	Forest	Infra- structure	Shrub	Other	<sup>1</sup> User's acc. (%)	<i>k</i>
Bare soil	<b>8</b>	1	0	0	0	3	0	2	57.14	0.519
Herbaceous	0	<b>11</b>	0	0	1	1	0	0	84.62	0.821
Crop	1	0	<b>19</b>	6	0	0	3	0	65.52	0.569
Pasture	0	1	1	<b>11</b>	0	0	1	0	78.57	0.742
Forest	0	1	0	0	<b>14</b>	0	2	0	82.35	0.792
Infrastructure	2	0	0	0	0	<b>7</b>	0	0	77.78	0.750
Shrub	0	0	0	0	0	0	<b>0</b>	1	0.00	-0.064
Other	0	0	0	0	0	0	0	<b>3</b>	100.0	1.00
<sup>1</sup> Producer's acc.	72.7%	78.6%	95.0%	64.7%	93.3%	63.6%	0.0 %	50.0%		
Overall accuracy	73.0 %									
Overall <i>k</i>	0.68									

<sup>1</sup>Producer's accuracy: number of correctly classified reference sites / total number of reference sites images; User's accuracy: number of correctly classified sites / total number of classified sites.

#### 4.6 Discussion (en supplément de l'article)

Puisque l'architecture proposée par Su *et al.* (2015) intègre un réseau qui n'est plus à la fine pointe en 2021 (soit VGG-M), nous l'avons remplacé par un DCNN qui a fait ses preuves et qui est encore une référence à ce jour (Wightman *et al.*, 2021), soit RESNET (He *et al.*, 2015). Pour ce faire, nous avons testé plusieurs niveaux de profondeurs de ce réseau, soit le RESNET-18, RESNET-34 et RESNET-50. Pour le même protocole d'entraînement, RESNET-34 et RESNET-50 ont obtenu sur le jeu de test respectivement des RMSE de 8,65 ( $R^2$  de 0,917) et 9,78 ( $R^2$  de 0,875) contre 7,0 ( $R^2$  de 0,931) pour RESNET-18. En plus d'augmenter le temps d'entraînement, aucune amélioration en performance n'a été obtenue avec l'emploi des réseaux plus profonds. Il faut dire que la tâche de classification dans cette étude est relativement simple puisqu'il s'agit d'images de petite taille (46 x 46 pixels) et que leur diversité est beaucoup plus faible que celle des images dans le jeu de données ImageNet par exemple (taille moyenne des images de 469 x 387 pixels). VGG-16 a aussi été utilisé lors de tests préliminaires et a obtenu des résultats similaires à RESNET-18, mais le temps requis pour l'entraînement est plus important en raison du nombre de paramètres plus élevé (138 millions contre 11 millions). Plusieurs autres architectures, comme EfficientNet (Tan et Le, 2019) et MobileNet (Howard *et al.*, 2019) seraient aussi pertinentes à évaluer pour leur intégration dans le MVDCNN.

D'un point de vue pratique, deux étapes sont nécessaires à l'utilisation du modèle développé. En premier lieu, l'échantillonnage des images des bandes riveraines, réalisé à l'aide des scripts développés en langage Python, est effectué automatiquement à partir des données vectorielles des cours d'eau et des images satellites fournies par l'utilisateur. La position géographique des images extraites (distance par rapport au cours d'eau et orientation) est déterminée à partir du segment vectoriel en question. Le regroupement des images par bande riveraine est aussi effectué automatiquement à cette étape. En ce sens, le segment vectoriel est considéré comme l'identifiant de la bande riveraine, c'est-à-dire qu'un groupe d'image est associé à un segment vectoriel donné. En deuxième lieu, le modèle effectue ses prédictions sur chacun des groupes d'images (un groupe par bande riveraine) et un fichier de type CSV contenant l'IQBR et l'identifiant de ces bandes riveraines est généré. Ce dernier fichier permet d'intégrer les résultats à une application cartographique comme ArcGis d'ESRI ou le logiciel libre QGIS. La seule tâche qui incombe à l'utilisateur est la détermination des segments vectoriels représentant les bandes riveraines. Pour

ce faire, l'utilisation des données hydrographiques libres d'accès du gouvernement fédéral ou provincial, est un bon point de départ. L'autre avenue possible est de ne pas utiliser de sections, mais d'effectuer une prédiction de l'IQBR pour chacune des images le long des cours d'eau, sans le regroupement des images selon le segment vectoriel. L'architecture MVDCNN a cette flexibilité en termes du nombre d'images en entrée, ce qui favorise son utilisation pour des longueurs de cours d'eau de quelques dizaines de mètres jusqu'à plusieurs kilomètres.

## **5. Conclusion**

### **5.1 Contributions**

Le survol des approches existantes pour la caractérisation des bandes riveraines présenté aux chapitres 1 et 2 montre qu'elles sont nombreuses et diversifiées. Ces méthodes sont développées pour des applications à petite ou grande échelle et adaptées aux enjeux de caractérisation locaux. L'utilisation de la télédétection pour ce faire est de plus en plus courante, mais s'attarde généralement qu'à des caractéristiques spécifiques de la bande riveraine comme la structure de sa végétation, l'identification des communautés végétales et la délimitation de la végétation riveraine. Dans ce projet, nous avons adopté une approche qui prend en compte la bande riveraine dans sa globalité et qui met à profit les techniques d'apprentissage profond. L'objectif global de cette recherche était donc de proposer une méthode pour la caractérisation des bandes riveraines et le suivi de leur état qui soit applicable à l'échelle régionale, et ce en lien direct avec les protocoles de caractérisation employés actuellement sur le territoire québécois, soit la mesure sur le terrain de l'IQBR.

L'évaluation de la capacité des DCNN pour la détermination de l'IQBR directement à partir d'images satellites THR, qui est l'objectif principal de ce projet, a été inspirée par les succès récents de ce type d'apprentissage profond pour la classification d'images, et ce dans plusieurs domaines d'application. Les meilleurs résultats ( $R^2$  de 0,932 et RMSE moyen de 7,35) montrent qu'avec l'utilisation d'images satellites THR et d'un MVDCNN incorporant un DCNN bien connu dans la littérature (le RESNET-18), leur potentiel pour cette tâche est excellent. L'hypothèse A est donc vérifiée, soit qu'il est possible de déterminer l'IQBR d'une bande riveraine en milieu agricole à l'aide d'un DCNN et d'une image satellite THR, avec une précision satisfaisante (RMSE inférieur à 16). Il s'agit, à notre connaissance, de la première application directe des DCNN pour la caractérisation des bandes riveraines ce qui constitue la contribution principale de ce projet.



En ce qui concerne le développement du modèle de détermination de l'IQBR reposant sur l'architecture MVDCNN adaptée, qui est le premier objectif spécifique de cette recherche, il a été motivé par la nécessité d'utiliser plusieurs images en entrée pour une seule annotation, soit la valeur d'IQBR. Cette architecture et l'emploi de 4 images de la bande riveraine en entrée de réseau lors de l'entraînement ont amélioré le RMSE moyen obtenu de 3,3% (7,35 contre 7,6) sur le jeu de test par opposition à l'utilisation d'une seule image, soit l'équivalent de l'utilisation d'un RESNET-18 standard. De plus, le MVDCNN demeure flexible lors des inférences, car il n'est pas limité à un nombre d'images précis. L'hypothèse B, qui stipule que l'utilisation simultanée de plusieurs images le long de la bande riveraine en entrée de réseau améliore les résultats par rapport à l'utilisation d'un DCNN standard à une seule image en entrée, est donc elle aussi vérifiée. Cette adaptation d'une architecture au départ conçue pour la classification de modèles 3D à la caractérisation d'objets sur des images de télédétection constitue la seconde contribution de cette recherche.

L'évaluation de l'impact du choix des bandes spectrales et de la stratégie d'entraînement sur les résultats de la régression est le deuxième objectif spécifique de ce projet. Les résultats ont montré que le choix des bandes spectrales utilisées n'est pas un facteur déterminant, réfutant ainsi la première partie de l'hypothèse C. En effet, avec une résolution spatiale de 50 cm, la texture, la forme et les patrons sont les caractéristiques qui semblent les plus importantes. Ce constat a été mis en lumière par les résultats obtenus avec l'utilisation d'une seule bande spectrale, peu importe laquelle, qui sont très proches des meilleurs. Ceci a l'avantage de ne pas limiter le choix du capteur pour des raisons de résolution spectrale. Tout de même, les images hyperspectrales, avec leur potentiel accru de discrimination des différents types de végétaux (Howard *et al.*, 2019), pourraient permettre au réseau de générer des caractéristiques plus pertinentes et ainsi améliorer les résultats.

Malgré la taille relativement importante de notre jeu de données d'images, le réseau préentraîné sur ImageNet s'est avéré plus performant que le réseau entraîné à partir de zéro, ce qui vérifie la seconde partie de l'hypothèse C. Les bandes spectrales ayant donné les meilleurs résultats en *fine tuning* sont également celles présentes dans ImageNet, soit les bandes R, G et B. Ce résultat met de l'avant l'importance de la taille et de la diversité des jeux de données pour l'apprentissage des DCNN. Il met aussi de l'avant la pertinence d'un jeu de données de référence en observation de la Terre comme GeoImageNet (Bouroubi *et al.*, 2019) (en développement) qui permettra de tirer

profit de caractéristiques apprises plus étroitement liées aux images utilisées ici (par opposition à ImageNet), en particulier en ce qui concerne la bande proche-infrarouge.

Le modèle développé dans cette étude permet une évaluation de la qualité de la bande riveraine en milieu agricole, et ce à une échelle régionale. Son utilisation pourrait faciliter le suivi de leur état année après année et permettre aux décideurs de mieux cibler les secteurs où des interventions sont requises. Un aperçu de la qualité de ces bandes végétalisées dans un secteur pas encore couvert par des campagnes de terrain serait également possible. Il pourrait d'ailleurs s'agir d'un outil d'aide à la décision important pour évaluer la nécessité de ces campagnes, selon ce qui est observé.

En bref, la méthode présentée innove en proposant une approche pratique d'application des DCNN pour la problématique des bandes riveraines en milieu agricole et constitue ainsi une contribution à l'état des connaissances.

## **5.2 Limites et perspectives**

Bien que les résultats présentés au chapitre 4 (l'article) soient encourageants quant aux performances du modèle développé, plusieurs pistes d'amélioration sont possibles.

Premièrement, le modèle a été entraîné avec des images acquises au mois de juillet uniquement, ce qui ne permet pas d'évaluer sa robustesse face aux changements de dates d'acquisition. Puisque la végétation change fortement d'aspect durant toute la saison de croissance, il est possible que les caractéristiques apprises par le réseau ne soient pas adaptées à la végétation du mois d'août par exemple. En ce sens, un réentraînement du réseau sur le nouveau jeu de données, éventuellement multidates, serait souhaitable. Deuxièmement, certaines classes d'occupation du sol comme les socles rocheux et les coupes forestières considérées dans la définition de l'IQBR sont absentes du jeu de données utilisé. Le modèle n'est par conséquent pas adapté à des régions où leur présence est importante. Troisièmement, ce modèle est particulièrement sensible à la présence de la canopée et peut pour cette raison surestimer la valeur réelle d'IQBR. Des données LiDAR ou radar ayant un potentiel de pénétration de la végétation important, telles qu'utilisées dans Dufour *et al.* (2013), pourrait potentiellement amoindrir cet effet.

Dès lors, l'utilisation de données multidates et/ou multisources (radar, LiDAR, optique) ouvre la porte à l'emploi de plusieurs réseaux en parallèle comme présenté à la section 2.3.4. Un MVDCNN (ou même un simple DCNN) distinct pourrait être entraîné sur des dates d'acquisition d'images

distinctes pour générer des caractéristiques pertinentes en fonction du moment dans la saison de croissance. Le même principe peut s'appliquer pour les données multisources, où chacun des réseaux exploiterait les propriétés de l'une ou l'autre de ces sources de données. Une technique de fusion de réseaux, comme celles développées par Hu *et al.* (2017) et Xu *et al.* (2018), pourrait ensuite être appliquée pour agréger l'information extraite. Avec cette information plus riche en termes de propriétés de la bande riveraine, tant au du point de vue spatial que temporel, de nouvelles avenues seraient à investiguer telle qu'une modélisation de la dynamique temporelle de la bande riveraine. Que ce soit d'année en année ou au cours d'une même saison, cette dynamique temporelle pourrait être caractérisée à l'aide de techniques d'apprentissage profond, plus spécifiquement grâce aux réseaux de type Mémoire à long terme (*Long Short Term Memory*; LSTM). Ces réseaux sont spécialement adaptés à l'apprentissage des dépendances d'ordre dans les séries temporelles grâce à leurs cellules mémoires. Une approche de ce type pourrait être informative quant à l'évolution de l'état des bandes riveraines d'un secteur donné. Un portrait dynamique plutôt que statique serait dès lors un autre outil d'aide à la décision important.

Enfin, l'entraînement du modèle présenté repose sur des mesures *in situ* d'IQBR difficilement reproductibles en raison de l'aspect subjectif inhérent à la méthode de mesure. Une part de bruit est par conséquent présente dans le jeu de données de référence, ce qui complique l'apprentissage du réseau et l'évaluation des résultats en test. La recherche pour la création d'un nouvel indice de caractérisation applicable au Québec méridional, mesurable à partir de produit de la télédétection et dont des répétitions dans les mesures sont réalisables est en cours (Allostry *et al.*, 2020). Il pourrait s'agir d'une avenue intéressante pour l'application de la méthode décrite dans le présent projet de recherche.

L'approche présentée dans ce mémoire se veut une méthode directe, en quelque sorte une réinterprétation de l'IQBR par intelligence artificielle d'un point de vue global (à haut niveau). Par contre, d'autres avenues pour l'application des DCNN à la problématique des bandes riveraines pourraient être envisagées. Cette technologie pourrait être employée à plus bas niveau en faisant appel à des architectures de type encodeur-décodeur utilisées pour la segmentation sémantique (Kemker *et al.*, 2018; Yuan *et al.*, 2021). Ces architectures pourraient par exemple améliorer les résultats de la classification des couvertures végétales de la bande riveraine dans la méthode de Macfarlane *et al.* (2017) ou de Novoa *et al.* (2018). Une méthodologie comme celle de Laslier *et*

*al.* (2019) utilisée pour générer des indicateurs issus de nuages de points LiDAR pourrait aussi bénéficier de méthodes de classification par intelligence artificielle conçues pour ce type de données (Xie *et al.*, 2020).

En conclusion, l'accessibilité et la diversité grandissante des produits de la télédétection offrent de nouvelles possibilités pour la caractérisation et le suivi des écosystèmes riverains. En plus des images provenant de multiples capteurs satellitaires et aéroportés, l'imagerie multispectrale acquise par drones (Jeziorska, 2019) est de plus en plus à la portée des différents organismes traitant de la problématique des bandes riveraines. Bien qu'adapté à des enjeux plus locaux que régionaux, ce type d'images dont la résolution spatiale est sub-métrique ouvre lui aussi la porte au développement de nouvelles approches de caractérisation. Les techniques reposant sur l'intelligence artificielle, encore peu nombreuses dans ce domaine, peuvent quant à elles mettre à profit l'abondance de ces données et permettre d'automatiser, de faciliter et d'améliorer certains processus liés à cette caractérisation. Les résultats de ce projet de maîtrise s'inscrivent dans cette lignée et ouvrent la porte à de nouvelles perspectives dans le suivi réglementaire des bandes riveraines mais aussi dans le suivi de leur rôle de protection de l'intégrité des terres agricoles et des cours d'eaux.

## Références

- Allostry, J., Varin, M., Paranjape, M., Théau, J. et Fournier, R. (2020) Développement d'un indice de suivi de l'état des écosystèmes riverains (ISÉÉR) dans le Québec méridional – Revue de littérature – Définitions et indicateurs (p. 120). Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc. (CERFO) – Université de Sherbrooke.
- Bertoldi, W., Drake, N. A. et Gurnell, A. M. (2011) Interactions between river flows and colonizing vegetation on a braided river: exploring spatial and temporal dynamics in riparian vegetation cover using satellite data. *Earth Surface Processes and Landforms*, vol. 36, n°11, p. 1474-1486.
- Bouroubi, Y., Chapdelaine, C., Foucher, S., Byrnes, D., Beaulieu, M., St-Charles, P.-L., Germain, M., Lauzier-Hudon, É., Bugnet, P., Sabo, N. et Gosselin, P. (2019) GeoImageNet: a Collaborative Platform for Deep Learning Application to Very High Resolution EO Images. Canadian Symposium on Remote Sensing 2019. Fredericton, Canada.
- Carluer, N., Lauvernet, C., Noll, D. et Munoz-Carpena, R. (2017) Defining context-specific scenarios to design vegetated buffer zones that limit pesticide transfer via surface runoff. *Science of the Total Environment*, vol. 575, p. 701-712.
- Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N. et Smith, V. H. (1998) Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, vol. 8, n°3, p. 559-568.
- Chaib, S., Liu, H., Gu, Y. et Yao, H. (2017) Deep Feature Fusion for VHR Remote Sensing Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, n°8, p. 4775-4784.
- Chatfield, K., Simonyan, K., Vedaldi, A. et Zisserman, A. (2014) Return of the devil in the details: Delving deep into convolutional nets.
- COVABAR (2016) Rapport de caractérisation du bassin versant de la rivière l'Acadie : secteur Acadie-Centre (p. 67). Québec, Canada.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li et Li Fei-Fei (2009) ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, p. 248-255.
- de Sosa, L. L., Glanville, H. C., Marshall, M. R., Abood, S. A., Williams, A. P. et Jones, D. L. (2018a) Delineating and mapping riparian areas for ecosystem service assessment. *Ecohydrology*, vol. 11, n°2, p. e1928.
- de Sosa, L. L., Glanville, H. C., Marshall, M. R., Prysor Williams, A. et Jones, D. L. (2018b) Quantifying the contribution of riparian soils to the provision of ecosystem services. *Science of The Total Environment*, vol. 624, p. 807-819.

- Dosskey, M. G., Eisenhauer, D. E. et Helmers, M. J. (2005) Establishing conservation buffers using precision information. *Journal of Soil and Water Conservation*, vol. 60, n°6, p. 349-354.
- Doyon, N. (2015) Fiche - La Bande Riveraine. in COBALI, Centre d'information, Bande riveraine, <https://www.cobali.org/wp-content/uploads/2016/11/La-bande-riveraine.pdf> (page consultée le 22 octobre 2021).
- Dufour, S., Bernez, I., Betbeder, J., Corgne, S., Hubert-Moy, L., Nabucet, J., Rapinel, S., Sawtschuk, J. et Trollé, C. (2013) Monitoring restored riparian vegetation: how can recent developments in remote sensing sciences help? *Knowledge and Management of Aquatic Ecosystems*, n°410, p. 10.
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. et Zisserman, A. (2010) The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, vol. 88, n°2, p. 303-338.
- Fernandes, M. R., Aguiar, F. C., Ferreira, M. T. et Pereira, J. M. C. (2013) Spectral separability of riparian forests from small and medium-sized rivers across a latitudinal gradient using multispectral imagery. *International Journal of Remote Sensing*, vol. 34, n°7, p. 2375-2401.
- Fernández, D., Barquín, J., Álvarez-Cabria, M. et Peñas, F. J. (2014) Land-use coverage as an indicator of riparian quality. *Ecological Indicators*, vol. 41, p. 165-174.
- Gagnon, É. et Gangbazo, G. (2007) Efficacité des bandes riveraines : analyse de la documentation scientifique et perspectives. ministère du Développement durable, de l'Environnement et des Parcs, vol. Direction des politiques de l'eau, p. 17.
- Gangbazo, G., Vallée, P., Émond, C., Roy, J., Beaulieu, R. et Gagnon, É. (2006) Contrôle de la pollution diffuse d'origine agricole : quelques réflexions basées sur la modélisation de scénarios de pratiques agricoles pour atteindre le critère du phosphore pour la prévention de l'eutrophisation dans la rivière aux Brochets. Ministère du Développement durable, de l'Environnement et des Parcs, vol. Direction des politiques de l'eau, p. 10.
- Gonzalez del Tanago, M. et Garcia De Jalon, D. (2011) Riparian Quality Index (RQI): A methodology for characterising and assessing the environmental conditions of riparian zones. *Limnetica*, vol. 30, n°2, p. 235-251.
- Gutman, G. et Ignatov, A. (1998) The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, vol. 19, n°8, p. 1533-1543.
- He, K., Zhang, X., Ren, S. et Sun, J. (2015) Deep Residual Learning for Image Recognition. arXiv, vol. arXiv:1512.03385.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 770-778.

- Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Le, Q. et Adam, H. (2019) Searching for mobileNetV3, vol. 2019-October, p. 1314-1324.
- Hu, F., Xia, G.-S., Hu, J. et Zhang, L. (2015) Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sensing*, vol. 7, n°11, p. 14680-14707.
- Hu, J., Mou, L., Schmitt, A. et Zhu, X. X. (2017) FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. 2017 Joint Urban Remote Sensing Event (JURSE), p. 1-4.
- Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Scherer, D., Müller, A. et Behnke, S. (2010) Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In K. Diamantaras, W. Duch, et L. S. Iliadis (dir.), *Artificial Neural Networks – ICANN 2010, Lecture Notes in Computer Science* (Vol. 6354, p. 92-101). Berlin, Heidelberg : Springer Berlin Heidelberg.
- Huylenbroeck, L., Laslier, M., Dufour, S., Georges, B., Lejeune, P. et Michez, A. (2020) Using remote sensing to characterize riparian vegetation: A review of available tools and perspectives for managers. *Journal of Environmental Management*, vol. 267, p. 110652.
- Ivits, E., Cherlet, M., Mehl, W. et Sommer, S. (2009) Estimating the ecological status and change of riparian zones in Andalusia assessed by multi-temporal AVHRR datasets. *Ecological Indicators*, vol. 9, n°3, p. 422-431.
- Jeziorska, J. (2019) UAS for Wetland Mapping and Hydrological Modeling. *Remote Sensing*, vol. 11, n°17, p. 1997.
- Johansen, K., Phinn, S., Dixon, I., Douglas, M. et Lowry, J. (2007) Comparison of image and rapid field assessments of riparian zone condition in Australian tropical savannas. *Forest Ecology and Management*, vol. 240, n°1-3, p. 42-60.
- Kemker, R., Salvaggio, C. et Kanan, C. (2018) Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, p. 60-77.
- Krizhevsky, A., Sutskever, I. et Hinton, G. E. (2012) ImageNet classification with deep convolutional neural networks, vol. 2, p. 1097-1105.
- Kukačka, J., Golkov, V. et Cremers, D. (2017) Regularization for Deep Learning: A Taxonomy. arXiv:1710.10686 [cs, stat].
- Kussul, N., Lavreniuk, M., Skakun, S. et Shelestov, A. (2017) Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, vol. 14, n°5, p. 778-782.

- Laslier, M., Hubert-Moy, L. et Dufour, S. (2019) Mapping Riparian Vegetation Functions Using 3D Bispectral LiDAR Data. *Water*, vol. 11, n°3, p. 483.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. et Zitnick, C. L. (2014) Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, n°PART 5, p. 740-755.
- Luo, W., Li, Y., Urtasun, R. et Zemel, R. (2017) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *arXiv:1701.04128 [cs]*.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. et Johnson, B. A. (2019) Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, p. 166-177.
- Macfarlane, W. W., McGinty, C. M., Laub, B. G. et Gifford, S. J. (2017) High-resolution riparian vegetation mapping to prioritize conservation and restoration in an impaired desert river. *Restoration Ecology*, vol. 25, n°3, p. 333-341.
- Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F. et Zhang, Y. (2018) Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, vol. 10, n°7.
- Marmanis, D., Datcu, M., Esch, T. et Stilla, U. (2016) Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters*, vol. 13, n°1, p. 105-109.
- Miao, X., Patil, R., Heaton, J. S. et Tracy, R. C. (2011) Detection and classification of invasive saltcedar through high spatial resolution airborne hyperspectral imagery. *International Journal of Remote Sensing*, vol. 32, n°8, p. 2131-2150.
- Michez, A., Piégay, H., Lejeune, P. et Claessens, H. (2017) Multi-temporal monitoring of a regional riparian buffer network (>12,000 km) with LiDAR and photogrammetric point clouds. *Journal of Environmental Management*, vol. 202, p. 424-436.
- Miller, S. J., Wardrop, D. H., Mahaney, W. M. et Brooks, R. P. (2006) A plant-based index of biological integrity (IBI) for headwater wetlands in central Pennsylvania. *Ecological Indicators*, vol. 6, n°2, p. 290-312.
- Ministère de l'Environnement et de la Lutte contre les changements climatiques (2020) Suivi de la qualité des bandes riveraines. *Eau*.
- Munné, A., Prat, N., Solà, C., Bonada, N. et Rieradevall, M. (2003) A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 13, n°2, p. 147-163.



- Nogueira, K., Penatti, O. A. B. et dos Santos, J. A. (2017) Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, vol. 61, p. 539-556.
- Novoa, J., Chokmani, K. et Lhissou, R. (2018) A novel index for assessment of riparian strip efficiency in agricultural landscapes using high spatial resolution satellite imagery. *Science of The Total Environment*, vol. 644, p. 1439-1451.
- Ou, X., Yan, P., Zhang, Y., Tu, B., Zhang, G., Wu, J. et Li, W. (2019) Moving Object Detection Method via ResNet-18 With Encoder–Decoder Structure in Complex Scenes. *IEEE Access*, vol. 7, p. 108152-108160.
- Pan, S. J. et Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, n°10, p. 1345-1359.
- Penatti, O. A. B., Nogueira, K. et dos Santos, J. A. (2015) Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), p. 44-51.
- Poff, B., Koestner, K. A., Neary, D. G. et Henderson, V. (2011) Threats to Riparian Ecosystems in Western North America: An Analysis of Existing Literature1. *JAWRA Journal of the American Water Resources Association*, vol. 47, n°6, p. 1241-1254.
- Politique de protection des rives, du littoral et des plaines inondables, PPRLP. 2020, Q-2, r.35.
- Romero, A., Gatta, C. et Camps-Valls, G. (2016) Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, n°3, p. 1349-1362.
- Saint-Jacques, N. et Richard, Y. (1998) Développement d'un indice de qualité de la bande riveraine : application à la rivière Chaudière et mise en relation avec l'intégrité biotique du milieu aquatique. *Le bassin de la rivière Chaudière: l'état de l'écosystème aquatique* (p. 6.1-6.41). Québec (Canada) : Ministère de l'Environnement et de la Faune.
- Shorten, C. et Khoshgoftaar, T. M. (2019) A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, vol. 6, n°1, p. 60.
- Simonyan, K. et Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*.
- Stewart, A., Canada, Agriculture et agroalimentaire Canada, Canada, et Direction générale des services agroenvironnementaux (2012) Manuel de conception des bandes tampons dans le Canada atlantique. Ottawa : Agriculture et agroalimentaire Canada.
- Stutter, M. I., Chardon, W. J. et Kronvang, B. (2012) Riparian buffer strips as a multifunctional management tool in agricultural landscapes: Introduction. *Journal of Environmental Quality*, vol. 41, n°2, p. 297-303.

- Su, H., Maji, S., Kalogerakis, E. et Learned-Miller, E. (2015) Multi-view Convolutional Neural Networks for 3D Shape Recognition. 2015 IEEE International Conference on Computer Vision (ICCV), p. 945-953. Santiago, Chile : IEEE.
- Tan, M. et Le, Q. V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946 [cs, stat].
- White, M. J. et Arnold, J. G. (2009) Development of a simplistic vegetative filter strip model for sediment and nutrient retention at the field scale. *Hydrological Processes*, vol. 23, n°11, p. 1602-1616.
- Wightman, R., Touvron, H. et Jégou, H. (2021) ResNet strikes back: An improved training procedure in timm. arXiv:2110.00476 [cs].
- Xia, G.-S., Yang, W., Delon, J., Gousseau, Y., Sun, H. et Maitre, H. (2010) STRUCTURAL HIGH-RESOLUTION SATELLITE IMAGE INDEXING, p. 7.
- Xie, Y., Tian, J. et Zhu, X. X. (2020) Linking Points with Labels in 3D: A Review of Point Cloud Semantic Segmentation. *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, n°4, p. 38-59.
- Xu, X., Li, W., Ran, Q., Du, Q., Gao, L. et Zhang, B. (2018) Multisource Remote Sensing Data Classification Based on Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, n°2, p. 937-949.
- Yang, Y. et Newsam, S. (2010) Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10*, p. 270. San Jose, California : ACM Press.
- Yuan, X., Shi, J. et Gu, L. (2021) A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, vol. 169.

## Annexe 1 – Preuve de soumission de l'article

De : susy@mdpi.com <susy@mdpi.com> De la part de Remote Sensing Editorial Office

Envoyé : 2 novembre 2021 11:51

À : Lord, Etienne <Etienne.lord@AGR.GC.CA>

Cc : Remote Sensing Editorial Office <remotesensing@mdpi.com>

Objet : [Remote Sensing] Manuscript ID: remotesensing-1469577 - Submission Received - Available Discount Vouchers

Dear Dr. Lord,

Thank you for submitting the following manuscript for publication in Remote Sensing:

Manuscript ID: remotesensing-1469577

Type of manuscript: Article

Title: Riparian Strip Characterization in Agricultural Environments Using Multi-View Deep Convolutional Networks (MVDCNN) and Satellite Images

Authors: Samuel de la Sablonnière, Samuel Foucher, Yacine Bouroubi, Philippe Vigneault, Etienne Lord \*

Received: 2 November 2021

E-mails: samuel.de.la.sablonniere@usherbrooke.ca, samuel.foucher@crim.ca,

Yacine.bouroubi@usherbrooke.ca, philippe.vigneault@agr.gc.ca, etienne.lord@agr.gc.ca Submitted to

section: Remote Sensing in Agriculture and Vegetation, <https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.mdpi.com%2Fjournal%2Fremotesensing%2Fsections%2FRSAV&data=04%7C01%7CSamuel.De.La.Sablonniere%40USherbrooke.ca%7Ce18aa01f273547a0cb3e08d99edb6d85%7C3a5a8744593545f99423b32c3a5de082%7C0%7C0%7C637715487531659097%7CUnknown%7CTWFpbGZsb3d8eyJWljoIMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCi6Mn0%3D%7C1000&data=P3cSRKqd pBxBsTFxNYwcH9BvHOeWIJZkAHkAqAAZm6U%3D&reserved=0>

<https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.mdpi.com%2Fjournal%2Fremotesensing%2Fsections%2FRSAV&data=04%7C01%7CSamuel.De.La.Sablonniere%40USherbrooke.ca%7Ce18aa01f273547a0cb3e08d99edb6d85%7C3a5a8744593545f99423b32c3a5de082%7C0%7C0%7C637715487531659097%7CUnknown%7CTWFpbGZsb3d8eyJWljoIMC4wLjAwMDAiLCJQIjoiV2luMzliLCJBTiI6IjEhaWwiLCJXVCi6Mn0%3D%7C1000&data=P3cSRKqd pBxBsTFxNYwcH9BvHOeWIJZkAHkAqAAZm6U%3D&reserved=0>