Picterra

Smart Data Analytics

FEASIBILITY STUDY - AREALSTATISTIK2020

REVIEW OF STATE-OF-THE-ART, TESTS ON DATA & RECOMMENDATIONS

Version 2.0

02.11.2017

Prepared for:

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS
Sektion Geoinformation

Espace de l'Europe 10,
CH-2010 Neuchâtel

Picterra Sàrl Avenue de Jurigoz 11, CH-1006 Lausanne www.picterra.ch

Contents

List of Abbreviations 3
1. Executive summary 4
1.1. Executive Summary 4
1.2. Résumé 6
1.3. Zusammenfassung 8
2. General 10
3. BFS land cover/use classes 11
4. Methodology 20
5. State-of-the-art 22
5.1. Machine learning and computer vision review 23
5.1.1. Learning strategies 23
5.1.1.1. Unsupervised learning 23
5.1.1.2. Supervised learning 24
5.1.1.3. Semi-supervised learning: 25
5.1.1.4. Data augmentation 25
5.1.1.5. Active learning 25
5.1.2. Feature extraction 26
5.1.2.1. Hand-crafted and unsupervised feature learning 26
5.1.2.2. Supervised feature learning 27
5.1.3. Classification models 28
5.1.3.1. Support Vector Machines 28
5.1.3.2. Decision trees and Random Forest 29
5.1.3.3. Convolutional Neural Networks (CNN) 30
5.1.3.4. General comparison of classification models 31
5.2. Land cover / land use classification review 32
5.3. Change detection review 42
5.4. Complementary data sources 48
5.4.1. Optical Earth Observation satellite data 48
5.4.2. High Altitude Platforms data 50
5.4.3. Crowdsourced Geospatial data 50
5.5. References 51
6. Comparative assessment 57
6.1. Comparative assessment of the methods reviewed 57
6.2. Comparative assessment in a Kiviat diagram 58
7. Test experiments 60

Picterra Sàrl 2017, all rights reserved

7.1. Introduction 60
7.2. Test data description 61
7.2.1. Aerial data 61
7.2.2. Auxiliary information 62
7.2.3. Ground truth information 62
7.3. Data pre-processing 66
7.3.1. Aerial data 66
7.4. Experiments and method description 66
7.4.1. General considerations 66
7.4.2. Classification 67
7.4.2.1. Approach A3: Deep feature-based classification 67
7.4.3. Change detection 68
7.4.3.1. Approach C1: pixel-based feature change detection 68
7.4.3.2. Approach C3: Deep feature-based change detection 69
7.5. Experiments results 69
7.5.1. Classification 69
7.5.2. Change detection 74
7.5.2.1. Approach C1: pixel-based feature change detection 74
7.5.2.2. Approach C3: Deep feature-based change detection 76
7.5.3. Computational considerations 79
8. Conclusions 80
8.1. Conclusions on the state-of-the-art review 80
8.2. Conclusions on the experiments performed 81
8.3. Recommendations 82
8.3.1. General system architecture 83
8.3.2. System development planning 84
Appendix 86

Picterra Sàrl 2017, all rights reserved

List of Abbreviations

Abbreviation	Description
AL	Active Learning
AOI	Area of interest
AUC	Area Under the ROC Curve
CGD	Crowdsourced Geospatial Data
CNN	Convolutional Neural Network
CVA	Change Vector Analysis
DEM	Digital Elevation Model
DN	Digital Number
DSM	Digital Surface Model
DTM	Digital Terrain Model
EO	Earth Observation
ESA	European Space Agency
FCIR	False Color Infrared
GPU	Graphic Processing Unit
GRD	Ground Range Detected (Sentinel-1 product level)
GWR	Federal building usage database (construction date, number of flats, floors, etc.)
L8	Satellite Landsat-8
LC	Land cover
LU	Land use
nDSM	Normalized Digital Surface Model
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared (spectral band)
OA	Overall Accuracy
RE	Red edge (spectral band)
ResNet-50	Residual neural Network, a type of CNN architecture, with 50 layers
ROC	Receiving Operator Curve
RS	Remote Sensing
SAR	Synthetic Aperture Radar
S1 / S2	Satellite Sentinel-1 / Sentinel-2
SSL	Semi-Supervised Learning
S3VM	Semi-Supervised Support Vector Machine
SVM	Support Vector Machine (machine learning model for classification)
SWIR	Shortwave infrared (spectral band)
VGG-16	“Visual Geometry Group” neural network model with 16 layers
VGI	Volunteered Geographic Information
VHR	Very High Resolution (referring to spatial resolution of satellite data)

Picterra Sàrl 2017, all rights reserved

1. Executive summary

1.1. Executive Summary

The land use/cover statistics produced by BFS over all of Switzerland is foreseen to be updated from year 2020 on (Arealstatistik n°5) in shorter time intervals of 6 years. Partially or entirely automated methods are envisioned to support and reduce the high workload of the interpretation task. This part of the study aims at reviewing the state-of-the-art for land cover/use classification, change detection and complementary data sources. First, a review of the Arealstatistik class catalogue has been done regarding the criteria and spatial extent to be considered in order to discriminate them. From this, three groups of classes could be identified with different potential for automation.

The state-of-the-art review presents the most relevant machine learning and computer vision methods, which can serve either for land cover/use classification or change detection. It can be concluded that among the different learning strategies, the supervised learning will be the most promising in the context of the BFS Arealstatistik where training data is available in very large quantities.

The different approaches have been assessed with five criteria: “reduced interpretation”, “accuracy”, “robustness”, “sustainability”, and “adaptation cost” through a score between 1 and 5. The outcomes of the assessment is summarized in the table below.

Methods		Evaluation criteria					score
ID	Description	Reduced	Accuracy	Robustness	Sustainability	Adaptation Cost	
A1	Land cover/use classification using aerial imagery and handcrafted features	2	2	3	3	4	14
A2	Land cover/use classification using aerial imagery and unsupervised features	2	3	3	1	3	12
A3	**Land cover/use classification using aerial imagery and supervised deep features**	**4**	**4**	**4**	**3**	**4**	**19**
A4	Dense land cover/use classification using aerial imagery and supervised deep features	3	3	3	3	2	14
B1	Land cover classification using satellite imagery and Random Forest	3	3	3	4	3	16
B2	Land cover classification using satellite imagery and Active Learning	2	2	4	4	3	15
C1	**Land cover/use change detection using aerial imagery and pixel-based features**	**3**	**4**	**3**	**3**	**4**	**17**
C2	Land cover/use change detection using aerial imagery and object-based features	3	4	3	3	3	16
C3	**Land cover/use change detection using aerial imagery and unsupervised deep features**	**4**	**4**	**3**	**3**	**4**	**18**

The most promising approach is A3 “Land cover/use classification using aerial imagery and supervised deep features” which provides a good balance between the five criteria. Then follows the change detection approach C3 “Land cover/use change detection using aerial imagery and unsupervised deep features”, which shows a high potential for change detection. Then as third ranked approach follows the more traditional change detection approach C1 based on pixel features. Generally, it can be noticed that tremendous improvements in the classification accuracies of land use/cover of aerial scenes have been achieved with the recent developments of deep learning architectures specific to image data. Regarding change detection, some early works have been published using deep learning features, which present a potential for reducing interpretation.

Finally, the complementary data that could support the Arealstatistik are images from new satellite constellations, which provide very high revisit rates and spatial resolution. In addition, the forthcoming high-altitude platform systems (HAPS) providing regional real-time monitoring capabilities are very interesting in this context, as well as crowdsourced geospatial data (CGD) such as OpenStreetMap. Researchers are

Picterra Sàrl 2017, all rights reserved

exploring the combined usage of crowdsourced data and aerial images to improve classification models or the comparison of geo-tagged images and aerial imagery for change detection purposes.

The test experiments performed in this study have quantitatively assessed and compared the three most promising approaches A3, C1, and C3. Approach A3 for classification and C3 for change detection both based on deep learning presented the best results. The tests exploited a set of aerial RGB data over 66'950 samples points for classification and 26'448 sample points for change detection, representing about 1.6% of the entire set of Arealstatistik points. The major land cover and land use classes, representing around 70% of the data, obtained the highest accuracies for automatic change detection and classification.

As final recommendation, this study highlighted that machine learning and specifically deep learning and computer vision tools are currently allowing high accuracy aerial image scene classification and can provide more automation in the interpretation tasks of the Arealstatistik, estimated around 50% of the current workload saved. All the current auxiliary data should be exploited within the classification models and additionally, surface topology (DSM) and satellite data should be considered for an improved discrimination of certain classes and changes. A workflow combining expert interpretation and automation is recommended to let the system improve over time and maximize the accuracy of the interpretation with expert interpretation on focused difficult classes and machine learning predictions on large mainstream classes. The Arealstatistik could also be extended to a denser grid with predictions every 10m instead of 100m taking around 30 days of computations.

Picterra Sàrl 2017, all rights reserved

1.2. Résumé

Les statistiques de l'utilisation / couverture des sols produites par BFS sur toute la Suisse seront effectuées tous les 6 ans à partir de l'année 2020 (Arealstatistik « Statistique de la superficie » n ° 5) et entraîneront une augmentation de la charge de travail d'interprétation. L'interprétation pourra bénéficier du support de méthodes entièrement ou semi-automatiques. Cette partie de l'étude vise à examiner l'état de l'art pour la classification de l'utilisation / couverture des sols, la détection des changements et les sources de données complémentaires. D'abord, un examen du catalogue des classes utilisées pour Arealstatistik a été fait en ce qui concerne les critères de classes et l'étendue spatiale à prendre en compte pour les discriminer. Trois groupes de classe ont pu être identifiés par rapport à leur type et potentiel d'automatisation.

La revue de l'état de l'art présente les méthodes d'apprentissage automatique et les méthodes de vision par ordinateur les plus pertinentes qui peuvent servir soit la classification du sol, soit la détection des changements. Il peut être conclu que parmi les différentes stratégies d'apprentissage, l'apprentissage supervisé sera le plus adapté au contexte de l'Arealstatistik, où les données d'entrainement sont disponibles en très grandes quantités.
Les différentes approches ont été évaluées avec cinq critères: «interprétation réduite», «précision», «robustesse», «durabilité» et «coût de l'adaptation» noté entre 1 et 5.

Les résultats de l'évaluation des différentes approches sont résumés ci-dessous :

Methodes		Critères d'évaluation					score
ID	Description	Réduite	Précision	Robustesse	Durabilité	Coût d' adaptation	
A1	Land cover/use classification using aerial imagery and handcrafted features	2	2	3	3	4	14
A2	Land cover/use classification using aerial imagery and unsupervised features	2	3	3	1	3	12
A3	**Land cover/use classification using aerial imagery and supervised deep features**	**4**	**4**	**4**	**3**	**4**	**19**
A4	Dense land cover/use classification using aerial imagery and supervised deep features	3	3	3	3	2	14
B1	Land cover classification using satellite imagery and Random Forest	3	3	3	4	3	16
B2	Land cover classification using satellite imagery and Active Learning	2	2	4	4	3	15
C1	**Land cover/use change detection using aerial imagery and pixel-based features**	**3**	**4**	**3**	**3**	**4**	**17**
C2	Land cover/use change detection using aerial imagery and object-based features	3	4	3	3	3	16
C3	**Land cover/use change detection using aerial imagery and unsupervised deep features**	**4**	**4**	**3**	**3**	**4**	**18**

L'approche la plus prometteuse est A3 « Land cover/use classification using aerial imagery and supervised deep features » qui fournit un bon équilibre entre les cinq critères. Ensuite vient l'approche de détection de changement C3 « Land cover/use change detection using aerial imagery and unsupervised deep features », qui montre un potentiel élevé de détection de changement. Puis, la troisième approche de détection de changement C1, plus traditionnelle et basée sur les caractéristiques des pixels, pourrait être plus approprié pour contrôler précisément le taux de détection de changement manqué. D'une manière générale, on constate que des améliorations considérables dans la précision de la classification des scènes aériennes de couverture / d'occupation du sol ont été réalisées avec le développement récent d'architectures d'apprentissage profond spécifiques aux images. En ce qui concerne la détection de changements, certains travaux récents ont été publiés en utilisant des fonctions d'apprentissage profond qui présentent un potentiel pour réduire l'interprétation.

Picterra Sàrl 2017, all rights reserved

Enfin, les données complémentaires qui pourraient soutenir les prochaines campagnes d'Arealstatistik, sont issues des nouvelles constellations de satellites qui fournissent une très haute résolution spatiale, temporelle (revisite quotidiennes) et des coûts réduits, ainsi que des systèmes de plate-forme à haute altitude (HAPS) qui fourniront un suivi régional en temps réel. De plus, les données géospatiales de crowdsourcing (CGD) fournissent une quantité massive de données telles que les médias sociaux avec des images géo-étiquetées, ou la cartographie faite par les utilisateurs comme dans OpenStreetMap. Des chercheurs explorent l'utilisation combinée de données de crowdsourcing et d'images aériennes pour améliorer les modèles de classification ou la comparaison de panoramas géo-taggés avec des images aériennes à des fins de détection de changement.

Les tests effectués dans cette étude ont permis d'évaluer et de comparer quantitativement les trois approches les plus prometteuses : A3, C1, et C3. Les approches A3 et C3, toutes deux basées sur de l'apprentissage profond, ont obtenu les meilleurs résultats. Les tests ont été basé sur des images couleurs (RVB) aériennes correspondant à 66'950 points d'interpretation pour la classification et 26'448 points pour la détection de changement, représentant 1.6% de la totalité des points interprétés lors d'une campagne Arealstatistik. Les classes majeures de couverture et d'utilisation du sol représentant à elles seules 70% de la totalité des données, ont obtenu les plus hauts scores de détection et de classification.

Au final, cette étude démontre que l'apprentissage automatique et plus particulièrement l'apprentissage profond (« deep learning ») et les outils de visions par ordinateurs permettent actuellement d'obtenir de très hautes performances en classification de scenes aériennes et peuvent fournir plus d'automatisme dans les tâches d'interprétation d'Arealstatistik, avec une estimation de réduction de la charge de travail de 50%. Il est recommandé d'utiliser toutes les données auxiliaires dans les modèles de classification et de détection de changement. De plus, la topologie avec un modèle numérique de surface (MNS) et les données satellites fréquente devraient être considérées pour améliorer la discrimination de certaines classes et changements. Un système combinant les interprètes experts et de l'automatisation est recommandée pour permettre au system de s'améliorer dans le temps et maximiser la qualité de l'interpretation, avec les experts interprétant les classes les plus difficiles et l'apprentissage automaitque prédisant les classes plus fréquentes et faciles. L'interprétation pourrait être étendues à une grille plus fine avec des prédictions tous les 10m au lieu de 100m et prenant environ 30 jours de calculs.

Picterra Sàrl 2017, all rights reserved

1.3.Zusammenfassung

Die vom BFS über die ganze Schweiz produzierte Landbedeckungs-/Landnutzungsstatistik soll ab dem Jahr 2020 alle 6 Jahre aktualisiert werden (ab Arealstatistik Nr. 5). Die Verwendung von teilweise oder vollständig automatisierten Methoden wird angestrebt um den hohen Zeitaufwand der Interpretationsaufgaben zu reduzieren. Dieser Teil der Studie zielt darauf ab, den Stand der Technik (state-of-the-art) für die automatische Klassifizierung von Landbedeckungs-/ Landnutzungsklassen, für das Detektieren von Veränderungen dieser Klassen und hinsichtlich zusätzlicher Datenquellen zusammenzustellen und auszuwerten. Zuerst wurde eine Betrachtung des Arealstatistik-Klassenkatalogs hinsichtlich der verwendeten Klassenkriterien und der relevanten räumlichen Ausdehnungen durchgeführt. Drei Gruppen von Klassen wurden identifiziert mit jeweils unterschiedlichem Automatisierungspotential.

Die Studie präsentiert die wichtigsten Methoden aus den Bereichen Maschinelles Lernen (machine learning) und computer vision, die geeignet sind um automatisch Landbedeckungen und Landnutzungen zu klassifizieren und Veränderungen zu erkennen. Es kann gefolgert werden, dass unter den verschiedenen Lernstrategien (learning strategies) das überwachte Lernen (supervised learning) wegen der Verfügbarkeit sehr großer Trainingsdatensätze der vielversprechendste Ansatz im Zusammenhang mit der Arealstatistik sein wird.
Die verschiedenen Ansätze wurden anhand von fünf Kriterien bewertet: "reduzierte Interpretation", "Genauigkeit", "Robustheit", "Nachhaltigkeit" und "Anpassungskosten" mit Werten zwischen 1 und 5. Die Ergebnisse der Bewertung sind in der folgenden Tabelle zusammengefasst.

Methods		Evaluation criteria					score
ID	Beschreibung	reduzierte	Genauigkeit	Robustheit	Nachhaltigkeit	Anpassungskosten	
A1	Land cover/use classification using aerial imagery and handcrafted features	2	2	3	3	4	14
A2	Land cover/use classification using aerial imagery and unsupervised features	2	3	3	1	3	12
A3	**Land cover/use classification using aerial imagery and supervised deep features**	**4**	**4**	**4**	**3**	**4**	**19**
A4	Dense land cover/use classification using aerial imagery and supervised deep features	3	3	3	3	2	14
B1	Land cover classification using satellite imagery and Random Forest	3	3	3	4	3	16
B2	Land cover classification using satellite imagery and Active Learning	2	2	4	4	3	15
C1	**Land cover/use change detection using aerial imagery and pixel-based features**	**3**	**4**	**3**	**3**	**4**	**17**
C2	Land cover/use change detection using aerial imagery and object-based features	3	4	3	3	3	16
C3	**Land cover/use change detection using aerial imagery and unsupervised deep features**	**4**	**4**	**3**	**3**	**4**	**18**

Der vielversprechendste Ansatz ist A3 “Land cover/use classification using aerial imagery and supervised deep features”, der eine ausgewogene Balance zwischen den fünf Kriterien erreicht. Es folgt Ansatz C3 "“Land cover/use change detection using aerial imagery and unsupervised deep features”, welcher ein hohes Potential für das Erkennen von Veränderungen zeigt. An dritter Stelle kommt der traditionellere pixel-basierte Ansatz C1. Generell kann festgestellt werden, dass durch die neueren Entwicklungen im Bereich des „deep learning“ außerordentliche Verbesserungen der Genauigkeiten von Landbedeckungs-/ Landnutzungsklassifizierungen von Luftbildern erzielt werden konnten. In Bezug auf das automatische Erkennen von Veränderungen wurden einige frühe Arbeiten veröffentlicht, die das Potential von „deep learning“ Ansätzen zur Verringerung des Interpretationsaufwandes zeigen.

Picterra Sàrl 2017, all rights reserved

Bezüglich zusätzlicher Datenquellen zur Unterstützung der Arealstatistik könnten Bilder neuer Satellitensysteme mit hoher räumlicher und zeitlicher Auflösung relevant werden. Daneben können in Zukunft auch Daten die von den noch experimentellen „high-altitude platform systems“ (HAPS) interessant werden sowie Daten, die durch crowdsourcing erfasst werden, z.B. OpenStreetMap. Die kombinierte Verwendung von derartigen Daten und von Luftbildern wird derzeit untersucht, um Klassifizierungsmodelle zu verbessern. Auch die unterstützte Erkennung von Veränderungen auf Luftbildern mittels geokodierter am Boden aufgenommener Fotos ist Objekt wissenschaftlicher Untersuchungen.

Die in dieser Studie unternommenen Tests ermöglichen die Evaluation und den qualitativen Vergleich der drei aussichtsreichsten Ansätze: A3, C1 und C3. Die Ansätze A3 und C3, welche beide auf tiefgehendem Lernen (deep learning) basieren, haben die besten Resultate erzielt. Die Tests basieren auf Luft-Farbbildern (RGB), welche 66'950 Punkten für die Klassifikation und 26448 Punkten für die Detektion von Änderungen entsprechen und somit insgesamt 1.6% der interpretierten Punkte im Rahmen einer Arealstatistik durchgeführten Kampagne repräsentieren.

Schlussendlich zeigt diese Studie auf, dass mittels automatischem, Computer-unterstütztem Lernen und insbesondere mittels tiefgehendem Lernen (deep learning) und maschinellem Sehen (computer vision) sehr gute Resultate in der Klassifizierung von Luftbildern erzielt werden können, was eine gewisse Automatisierung im Rahmen der Interpretationsaufgabe der Arealstatistik erlaubt und somit den Arbeitsaufwand um geschätzte 50 % reduziert. Es wird empfohlen, alle Hilfsdaten in den Klassifikations-Modellen und Detektion von Änderungen zu verwenden. Des Weiteren sollten die Topologie mittels einem digitalen Oberflächenmodell und regelmässig aufgenommenen Satellitenbilder in Betrachtung gezogen werden, um die Unterscheidung gewisser Klassen und Veränderungen zu verbessern. Ein System wird empfohlen, welches Experten-basierte Interpretation sowie Automatisierung verbindet zur Verbesserung des Systems im Verlaufe der Zeit und zur Maximierung der Interpretationsqualität. Dabei interpretieren Experten schwierigere Klassen und automatische Lernmethoden produzieren Vorhersagen für häufigere und einfachere Klassen. Die Interpretation könnte durch ein feinmaschigeres Netz erweitert werden und Vorhersagen alle 10 Meter statt alle 100 Meter könnten innert ca. 30 Tage Rechenzeit produziert werden.

Picterra Sàrl 2017, all rights reserved

2. General

This report presents the outcomes of the feasibility study, which in a first phase reviews the state-of-the-art on approaches for a more efficient classification of land cover and land use, and on the data supporting such approaches, and in a second phase performes experiments on real data used by BFS for land cover/use interpretation. BFS contracted Picterra to perform this study, which reviews the state-of-the-art, tests promising approaches and assesses the potential for automation of BFS Arealstatistik operations in the future.

The land use statistics produced by BFS over all of Switzerland are planned to be generated every 6 years from year 2020 on (Arealstatistik n°5), which represents a significant reduction of the time interval previously used for completing a survey.

The outcomes of this study shall allow the identification of the most promising elements for partial or complete automation within a future workflow. Automation approaches for these elements (specific land cover or land use classes or groups of classes, detection of changes, etc.) shall enable the reduction of the generally very time-consuming interpretation tasks.

In order to identify the elements from which the BFS Arealstatistik surveys could strongly benefit, first the BFS catalogue of land cover and land use classes is examined with the aim to understand class characteristics, complexity and define meaningful class groups. Then, the review of the currently available technologies for automatic image analysis is performed, with the state-of-the-art in remote sensing image change detection and image classification, and the exploration of potential complementary data sources. A comparative assessment of the different methods available is been done and the most promising approaches are tested using the aerial data used for interpretation. Conclusions are drawn on the state-of-the-art review and the tests performed, and recommendations are provided to support the development of a system architecture supporting the forthcoming BFS Arealstatistiks.

Picterra Sàrl 2017, all rights reserved

3. BFS land cover/use classes

To ensure that this state of the art review addresses correctly the requirements it is necessary to examine the catalogue of land cover/use classes used for the Arealstatistik by BFS, i.e. their characteristics and definition. Different groups of classes will be defined with respect to their complexity in terms of their composition (visual similarities/dissimilarities, multiples elements, etc.) and the data required for their identification.

Before presenting a summary of the classes, a schematic of the sample point (red cross), the interpretation area (white box) and the reference surface area (red box) is presented in Figure 1. The interpretation area represents the context used to classify the majority of the classes. Additionally, a reference surface area of 625m2, placed to the North-East of the sample point by convention, is used to estimate certain surface criteria within certain class definition. The interpretation surface being centered on the sample point, the interpretation is rotation invariant (the zone can be flipped or rotated without changing the interpretation). Most likely, the pixels to be considered in order to classify a sample point will be the ones lying in this interpretation area.

Figure 1 - Schematic of the different areas to be considered around a sample point: the interpretation surface 2500m^2 (in white) centered around the sample point and at the top-right, the reference surface of 625m^2 (in red).

The different land cover / land use classes established for the Arealstatistik surveys are summarized in the table below with their name, ID, description of elements they include or exclude, criteria for their classification. Additionally, three further characteristics have been introduced for this study:

- **Extent**: minimum region around the point of interest to be considered for interpretation
  - 0: <1m, the very close neighborhood of the point needs to be considered
  - 1: 25m, a zone of 25x25m needs to be considered
  - 2: 50m, a zone of 50x50m needs to be considered = equivalent to interpretation surface

Picterra Sàrl 2017, all rights reserved

  - 3: 100m, a zone of 100x100m needs to be considered, which represents a “wall-to-wall” coverage of the Arealstatistik grid with one point every 100m.
  - 4: >100m, a zone larger than 100x100m needs to be considered
- **Data types**: which data types are providing the characteristics required for classification
  - Standard data (currently used by interpreter):
    - Spectral: color information (R, G, B, NIR) at the point required
    - Spatial: spatial patterns around the point required, for example:
      - Texture: color information and its spatial pattern around the point is required
      - Shape: information about the shape of the object around the point (e.g. right angles for buildings, etc.)
    - Terrain: terrain characteristics required, such as altitude, slope or exposition derived from a DTM at the point of interest.
    - Cadastral information: for example, building usage derived from the GWR data
  - Additional data:
    - Surface context: information about the surface required, such as height (DSM - DTM) or surface smoothness around the point of interest.
    - Geographical location: for example, information on the administrative unit (e.g. canton), the climatic zone, etc. required.

- **Composition**: if the class is composed of a single element type (e.g. greenhouse, high visual similarity), of multiple simple elements (e.g. trees), or a mix of elements (e.g. sealed surfaces, low similarity between roads and artificial water storage for example).
  - “single”: single type of elements, uniform class.
  - “multiples”: multiple simple elements, class made of a set of uniform elements with varying organization.
  - “mixed”: mixed elements, high class variability.

Picterra Sàrl 2017, all rights reserved

Class group	Class name	ID	Description	Criterion	Extent	Data types	Composition
LC - urban	Sealed surfaces	11	Includes roads, pedestrian side path, waste water treatment pools, manure hips, artificial water storage, sealed surface below tree canopy even if not directly seen on image	At the point, No minimal surface	0: <1m	Standard data	mixed
	Buildings	12	Includes verandas, winter garden, ruins if point on a foundation wall, silos and tanks, flat vegetated roofs. Does not include caravans, construction site temporary buildings, small garden houses	At the point, inside the built surface	1: 25m	Standard data + Additional data (surface height)	mixed
	Greenhouses	13	Includes greenhouses at man height; Does not include verandas, winter gardens, protected agricultural fields, low height greenhouses found in gardens.	At the point, inside the built surface	1: 25m	Standard data + Additional data (surface height)	single
	Structures of vegetable growing systems	14	Includes gardening surfaces, trees over a gardening surface, horticulture, tree nursery, tomb in cemetery, flower fields, small garden houses	At the point, inside the garden surface	1: 25m	Standard data	mixed
	lawn	15	Includes intensively used grassland (part of golf course, grassy runway, etc.), usually flat surface, low slope, of a certain size Does not include agricultural grassland	At the point, No minimal surface	1: 25m	Standard data	single
	Trees on landscaped surface	16	Includes at least 3 trees on landscaped surface, cemetery, camping sites, sport surfaces Not to be confused with LC 47 (group of trees over pasture)	At the point, either inside a group (min 3) of trees, or isolated tree at no more than 25m to a group of tree (3 min)	2: 50m	Standard data + Additional data (surface height)	multiples: trees
	Small mixed structures	17	Includes small surfaces of lawn, bushes, isolated trees, flowers mixed altogether.	At the point, inside a mixed surface close to buildings.	1: 25m	Standard data	mixed
LC - vegetation	Herbaceous vegetation	21	Includes Natural grass and plants: arable land, pasture with high biodiversity, also close to housing surfaces and infrastructures (ex: roads) and inside agricultural land	- Agricultural and non-productive surfaces : at the point on grassy homogeneous surfaces >625m2 or heterogeneous grassy surfaces (at least >50% of sand, humous, talus) - Urban areas and infrastructures (cf LU), at the point, heterogeneous grassy surface with no minimum surface	1: 25m	Standard data	mixed
LC - bushes & agriculture	Bushes and shrubs	31	Includes bushes, shrubs, fern trees and bushy trees below 3m	At the point on an homogenous surface >625m2	1: 25m	Standard data + Additional data (surface height)	single

Picterra Sàrl 2017, all rights reserved

	Bushy surfaces overgrown by vegetation	32	Includes bushes and trees below 3m with medium surface covering rate 50 to 80%	At the point on the bushy surface	1: 25m	Standard data + Additional data (surface height)	mixed
	Fruit trees with low stems	33	Includes fruit trees plantation with more than 3 trees per acre as well as the surfaces inside the plantation and at its border	At the point on the fruit tree on between fruit trees or at the border of the plantation	3: 100m	Standard data + Additional data (surface height)	mixed
	Vineyards	34	Includes plant vines, inter-ranks (max 3m), tool sheds, vine nurseries	At the point on the vine plants or inter-ranks	1: 25m	Standard data	mixed
	Horticultural crop	35	Includes all plantations (small fruit trees, bush, plants..) and tree nurseries for commercial purposes as well as inter-ranks (max 3m)	At the point on the vine plants or inter-ranks	1: 25m	Standard data	mixed
LC - forest	Dense forest stands	41	Includes forest stand with tree above 3m, fruit trees with high stems, poplar and willow plantation: >covering rate >60%	At the point on the forest stand or near the border of the stand if well delineated. Min 25m width stands becoming after 50m max a continuous stand >50 m width . Trees >3m	2: 50m	Standard data + Additional data (surface height)	multiples: trees
	Forest corners	42	Same as 41, max 25 m long near a forest stand above 25m width	At the point of the stand, max 25 width becoming after 25 m max a continuous stand > 25 m width. Trees >3m	1: 25m	Standard data + Additional data (surface height)	multiples: trees
	Forest strips	43	Same as 41, 42. Isolated stand between 25-50 width and forest stripes >50 m long and 25-50 m width close forest stands (41)	At the point on the forest stand or near the border of the stand if well delineated. 25-50 m width For isolated stand: length >25m If closed to 41, length >50m	2: 50m	Standard data + Additional data (surface height)	multiples: trees
	Spars forest stand	44	Includes same as 41, 42, 43. Trees height >3m Min width 50 m Low coverage 20-60%	At the point within the stand	2: 50m	Standard data + Additional data (surface height)	multiples: trees
	Bushy forest stand	45	Includes Bushes: - Min width: 25m -Medium coverage >60% - Height >3m - Height exception for green alder and trailing pine < 3m	At the point on the forest stand or near the border of the stand if well delineated.	2: 50m	Standard data + Additional data (surface height)	multiples: trees
	Linear forest stand	46	Includes: - Forest stand min 25 m length and max 25 m width , coverage >60% - Height>3m (except for cut hedges)	At the point on the tree crown	1: 25m	Standard data + Additional data (surface height)	multiples: trees

Picterra Sàrl 2017, all rights reserved

			- Tree fruits with high stems on housing surfaces and infrastructures - poplar and alder plantations - Corner forest				
	Groups of trees	47	Includes tree stand with a min of 3 trees within the afforestation with low coverage, - Tree fruits with high stems, forest corners ends-up by sparse tree groups Exclude all class from 41-46	At the point on the tree crown. Coverage < 20% if width >50m Coverage < 60% if width >25-50m If width <25m no min coverage applies Height >3m	2: 50m	Standard data + Additional data (surface height)	multiples: trees
LC - rocks	Salient rocks	51	Includes excavated building sites, quarries, rocky river /lake banks, cliff.	At the point on a rocky surface min 625m2 or on an heterogeneous surface cover >80% of rocks	1: 25m	Standard data	single
	Loose rock/soil	52	Includes scree, sand excavated construction site, humus, permafrost, river bed, gully	At the point on loose rocks and screes surface min 625m2 or on an heterogeneous surface cover >80% of rocks	1: 25m	Standard data + Vector25	multiples: rocks
	Rocky surfaces	53	Includes 21 or 31 + 51 or 52	At the point on the rocky surface, cover with 50%-80% of 51 and /or 52 covered at max 50% by shrubs (31)	1: 25m	Standard data	multiples: rocks
LC - water, snow	Water body	61	Includes Lakes, rivers, torrents within natural or qusai natural environments but exclude watercourse width <6m under tree canopy	At the point on the surface	0: <1m	Standard data	single
	Glacier and firn, névé	62	Glacier, firn, névé, screes on glacier, waterbody<25m2 on glacier, avalanche debris cone	At the point on the glacier or firn min surface 0,25Ha or ice, névé >80% of the reference surface	2: 50m	Standard data	multiples: snow
	Humid Biotope	63	Includes wetland classified by OFEV covered with vegetation (47,21,31)	At the point on homogeneous surface 625m2, >80% of heterogeneous cover inside of the reference surface is humid, reed <50%	1: 25m	Standard data + OFEV classification	multiple: humid soil
	Reed bed	64	Includes reeds and waterbodies < 25m2 inside reed beds	At the point on homogeneous surface 625m2, >80% of heterogeneous cover inside of the reference surface is made of reeds, reed >50%	1: 25m	Standard data	single

Picterra Sàrl 2017, all rights reserved

LU - urban	Industrial surfaces > 1ha	101	Includes shopping centers, industrial harbor zones, fuel stations, sawmill, post centers, trees and bushes in the surface Do not include roads in the surface, congress centers, military centers	At the point, surface of min. 1ha	4: >100m	Standard data + Additional data (surface height)	multiples: sealed surface + buildings + roads
	Industrial surfaces < 1ha	102	As above	At the point, surface of less than 1ha	3: 100m	Standard data + Additional data (surface height)	multiples: sealed surface + buildings + roads
	Individual houses and houses with 2 independent housing	103	Includes houses with swimming pools, holiday houses	At the point, inside the building surface	1: 25m	Standard data + Additional data (surface height)	multiples: buildings
	Aligned houses and terrace housing surfaces	104	Same as above but with at least 3 individual houses or terrace housing of more than 2 houses	At the point on the surface	1: 25m	Standard data + Additional data (surface height) + (cadaster)	multiples: buildings + pastures + grass + shrubs + roads (sealed and unsealed)
	Residential Building surface	105	Surface of buildings, houses of more than 2 floors for dwelling purpose. Includes also the dwelling part of hospital, boarding school, hotel etc… as well as their direct surroundings	At the point on the surface	1: 25m	Standard data + Additional data (surface height) (cadaster)	multiples: buildings + pastures + grass + shrubs + roads (sealed and unsealed)
	Public building surfaces	106	Includes the buildings describe in the NOGA nomenclature	At the point on the surface	1: 25m	Standard data + NOGA	multiples: buildings + pastures + grass + shrubs + roads (sealed and unsealed)
	Agricultural building surfaces	107	Includes housing and farm building and their surroundings	At the point on the surface	1: 25m	Standard data	multiples: buildings + pastures + grass + shrubs + roads (sealed and unsealed)
	Indeterminate buildings surfaces	108	Includes surfaces and buildings that can not be attributed to classes 101-107	At the point on the surface	1: 25m	Standard data	multiples: buildings + pastures + grass + shrubs + roads (sealed and unsealed)

Picterra Sàrl 2017, all rights reserved

	Motorway surfaces	121	Includes surfaces delineated by wildlife protection fence, or roads with multiple traffic lanes separated by safety barriers	At the point inside the surface delineated by protection fence	1: 25m	Standard data	mixed
	Road surfaces	122	Includes smooth surfaces used for circulation of vehicles or pedestrians and surroundings surfaces	At the point on the surface	1: 25m	Standard data	mixed
	Parking surfaces	123	Includes all smooth surfaces often compacted used for vehicles parking as well as direct surroundings	At the point on the surface	1: 25m	Standard data	mixed
	Rail surfaces	124	Includes surfaces used for rail transportation and their surroundings surfaces	At the point on the rail surfaces	1: 25m	Standard data	mixed
	Aerodromes	125	Includes surfaces used for aerial transportation and their surroundings surfaces	At the point on the aerodrome	1: 25m	Standard data	mixed
	Energy supply facilities	141	Includes surfaces used for production and distribution of energy	At the point on the facility	1: 25m	Standard data	mixed
	Water treatment plant	142	Includes surfaces used to treat waste water and their direct surroundings	At the point on the water treatment plant	1: 25m	Standard data	mixed
	Other facilities of supply and elimination	143	Communication, water supply and waste collection infrastructures and surroundings	At the point on the facilities	1: 25m	Standard data	mixed
	Dumps	144	Includes surface used for final storage of household refuse, excavation material and debris and surroundings	At the point on the surface of the landfill site	1: 25m	Standard data	mixed
	Extraction sites	145	Includes surfaces used to extract raw material also on lakes and rivers	Centre of the point on the surface	1: 25m	Standard data	mixed
	Construction sites	146	Includes temporary and long term construction sites, modifying the original land use of the surface to housing and or to infrastructure	A the point on the surface in construction	1: 25m	Standard data	mixed
	Brownfields and closed down buildings	147	Includes abandoned surfaces inside housing and infrastructures surfaces which did not find new assignment but also ruins	At the point on the surface	1: 25m	Standard data	mixed
	Public parks	161	Includes public green spaces for recreational purposes, harbors and zoos and all infrastructures belonging to the parks	At the point on the surface	1: 25m	Standard data	mixed
	Sport facilities	162	Includes public sport infrastructures and facilities	At the point on the surface	1: 25m	Standard data	mixed
	Golf course	163	Includes golf courses, buildings and every surfaces inside the golf course perimeter	At the point on the surface	1: 25m	Standard data + list of golf courses in Switzerland	mixed
	Camping site	164	Includes Camping and caravanning sites and buildings and every surfaces inside the camping site	At the point on the surface	1: 25m	Standard data + list of camping site in Switzerland	mixed
	Family gardens	165	Includes cultivated land, gardens surfaces outside of the adjacent housing surfaces with no commercial goals and with leisure purpose + infrastructures and surfaces inside of the family garden surface	At the point on the surface	1: 25m	Standard data	mixed
	Cemeteries	166	Includes surface used to burry human being or animal as well as to leave funerary urn + buildings and surfaces inside of the cemetery surface	At the point on the surface	1: 25m	Standard data	mixed

Picterra Sàrl 2017, all rights reserved

LU - Agriculture	Arboriculture	201	Includes Agricultural surfaces mainly used for arboriculture. At least 3 trees at a maximum distance of 25 m + infrastructures and surfaces inside of the plantation	At the point or between the canopy trees or between canopy and edge plantation	1: 25m	Standard data	Multiples: trees
	Viticulture	202	Includes agricultural land used for viticulture + infrastructures and surfaces inside of the surface	At the point inside the vineyard	1: 25m	Standard data	mixed
	Horticulture	203	Include all plantations (small fruit trees, bush, plants..) and tree nurseries for commercial purposes	At the point on the production surface	1: 25m	Standard data	mixed
	Arable land in broad sense	221	Includes arable land, artificial pastures and “assolement”	At the point on the arable land (one block, same structure, low slope)	2:50m	Standard data + "Terres cultivables" de Res Finger	mixed
	Natural pasture in broad sense	222	Includes surfaces located in a permanent zone of habitat, covered with herbaceous vegetation reap at least once a year with no “assolement”	At the point on the natural pasture	2:50m	Standard data	mixed
	Local pastures in broad sense	223	Includes local pasture mainly used for grazing in a permanent zone of habitat	At the point on the natural pasture	2:50m	Standard data	mixed
	Reaped Mountain pastures	241	Includes mountain pastures located outside of a permanent zone of habitat	At the point on reaped mountain pastures	2:50m	Standard data + statistics 2004/09	mixed
	Grazed mountain pastures	242	Includes mountain pastures with seasonal use or exploited from inhabited farms with high seasonal variations of livestock number	At the point on the grazed mountain pasture	2:50m	Standard data + statistics 2004/09	mixed
	Grazed high-mountain pastures for sheeps	243	Includes isolated surfaces, difficult to access in high mountain used for sheep grazing	At the point on the surface . Agricultural use on more than 50% of the surface	2:50m	Standard data + statistics 2004/09	mixed
LU – Wooded areas	Forest Stands	301	Includes dense or sparse forest stands outside of housing and infrastructure surfaces. Forest stand can be exploited for sylvicultural purpose	At the point inside the forest stand without agricultural use neither housing or identifiable infrastructures	2:50m	Standard data	mixed
	Afforested surfaces	302	Incudes forest stands that following sylvicultural measure had become forest.	At the point on the afforested surface + no forest on the previous acquisition	2:50m	Standard data	mixed
	Forest cuts	303	Includes forest stand that has been partially cut (sparse cut or logging) following sylvicultural activities	- At the point on the cut, - the forest stand on the previous acquisition was dense or sparse -Coverage <60%	2:50m	Standard data	mixed
	Damaged forest surfaces	304	Includes forest stands, dense or sparse that has been damaged by storms or fires	- At the point on the cut, - the forest stand on the previous acquisition was dense or sparse -Coverage <60%	2:50m	Standard data	mixed

Picterra Sàrl 2017, all rights reserved

	Lakes	401	Includes lakes, reservoirs, ponds in zoos and gravel pits, water surfaces within ports, temporary lakes after flooding/mudflow/landslide, lakeshores wet biotope outside of settlements; Excluded: water surfaces within settlements, for aquaculture, waste water treatment facilities	At the point, area > 25m²	1: 25m	Standard data + Additional data (surface height)	single
	Running waters	402	Includes rivers, streams, rocks/sand/gravel in river or stream bed, vegetated or fortified banks, temporary running waters after mudflows; Excluded: steams < 6m under forest; within public parks	At the point	1: 25m	Standard data + Additional data (surface height)	single
	Technical barriers against flooding	403	Includes all related parts of such barriers; also if point is located between two structures within a stream bed which have a max. distance of 50 m.	At the point	1: 25m	Standard data + Additional data (surface height, TLM)	single
	No usage	421	Very heterogeneous: e.g. military drill grounds, areas damaged/covered by land slide, former alpine pastures covered with bushes, glaciers, bogs, etc.	At the point	2: 50m	Standard data + Additional data (surface height, TLM)	mixed
	Technical barriers against avalanches, rockfall	422	Includes all related parts of such barriers; also if point is located between two structures within < 50 m or < 25m for “avalanche humps”	At the point	1: 25m	Standard data + Additional data (surface height, TLM)	single
	Alpine sports infrastructure	423	Very heterogeneous: e.g. ski lifts, ski slopes, trails within skiing area, etc.; Excluding buildings within skiing area, water reservoirs used for artificial snow	At the point	2: 50m	Standard data + Additional data (surface height, TLM)	mixed
	Landscape interventions	424	Temporally disturbed areas, e.g. due to installation of pipes, ameliorations of cultivated land; Excluding: construction sites, mining	At the point	1: 25m	Standard data + Additional data (surface height)	mixed

Picterra Sàrl 2017, all rights reserved

The following information can be deduced from the Arealstatistik LU/LC class catalogue:

- **Minimal mapping unit (spatial resolution)**: generally, a sub-meter resolution (~50cm) is required to distinguish elements composing a class and at the sample point. This restricts the investigation of complementary data to very high spatial resolution sources of data.
- **Scene extent**: in average the scene extent to be considered to identify the class is 50x50m, with exceptionally hundreds of meters required for few classes (e.g. industrial area > 1ha) or less than a meter around the sample point (e.g. water body). This will set the focus on the review to methods dealing with a scene of pixels more than individual pixels or small objects.
- **Data types**: the class interpretation requires the standard set of data, mainly spectral information for land cover classes and auxiliary information layers (e.g. GWR) for land use classes. In addition to the standard set of data, the surface height (normalized DSM) around the sample point will be required for most of the classes. This data will allow the machine to deduce similar information than the interpretation to the stereo imagery used by the interpreters.
- **Composition**: classes can be split in three groups of composition: “single”, “mixed” and “multiples” categories. The group of “single” class composition should be the easiest to classify automatically and should represent a high potential for automation. The group “mixed” composition will be complex to handle by method based only on spectral information and will definitely require the auxiliary data in order to be classified automatically. Finally, the group of “multiples” class composition should benefit from methods classifying densely the scene (at pixel or superpixel level) into a certain set of restricted simple classes (trees, rocks, buildings, roads). From such classification could be tested certain pre-established rules (above or below a certain percentage, distances between trees, etc.).

4. Methodology

This section presents the methodology used for the comparative assessment of the methods, presented in section 1 The following five assessment criteria, presented in Figure 2, have been established in the proposal and validated at the beginning of the project during the kick-off meeting. No significant modifications of the criteria has been done since their validation.

Criteria	Description
Reduced interpretation	The capability to minimize user interpretation. For example, an active learning strategy, or a high number of different data allowing to reduce the number of labeled examples required for reaching a similar accuracy.
Accuracy	Capability of the type of model to adapt and discriminate the classes of interest (all classes, only some groups of classes or only few classes). The theoretical risk of error of the method, if known and bounded, is taken into account.
Robustness	Capability to generalize from a small number of labeled samples (e.g. what minimal amount of training data required), to be unaffected by class imbalance, intrinsic data variability (atmospheric effects, shadows, wrong labeling of annex data e.g. OpenStreetMap, etc.), to integrate time series, and thus reduce generally the need for calibration procedures.
Sustainability	Possibility to integrate easily new type of data, to handle large data sets, to re-train the model at low computational cost, etc.
Adaptation cost	General complexity of the algorithm, pre-existence of tools for its implementation at scale, and ease of adaptation to the specificities of the classification problem for its operationalization.

Figure 2 - Assessment criteria

The comparative assessment of the methods is based on the above-mentioned criteria, with a defined weight and rated on a scale from 1 to 5 representing the strength (5) and weaknesses (1) of each method. Here are presented the scale and weight for the different criteria:

Picterra Sàrl 2017, all rights reserved

Criteria: / Score:	Reduced interpretation	Accuracy	Robustness	Sustainability	Adaptation cost
1	Reduced interpretation preserving accuracy not guaranteed.	Desired accuracy level not reachable for any class group.	Strongly affected by class imbalance and variability	Only training from scratch possible	No tools existing
2	Interpretation can be reduced by only a few interpretation per class (< 10) or few classes can avoid interpretation	Desired accuracy level reached under certain conditions for a specific group of class.	Affected by class variability and imbalance	Training model with additional batch data possible (batch training)	Partial availability of tools, entire development required.
3	Interpretation can be reduced to a small set of interpretation per class (~100) or a majority of classes can avoid interpretation	Desired accuracy level reached for a specific group of class.	Can cope with class imbalance and variability, but requires a large amount of training data (> 10000 samples / class)	Training model with new single data possible (online training)	Tools existing as source code/scripts, development required.
4	Interpretation is reduced to its minimal with few interpretation per class (< 10) required	Desired accuracy level reached for several groups of classes.	Not affected by class imbalance and variability, requires some training data (~ 1000 samples / class)	Training model with new data and additional features (online training)	Tools existing as COTS software components and/or source code/scripts, integration required.
5	Interpretation can be entirely removed	Desired accuracy level reachable for all type of classes of interest.	Can integrate complex data such as time series, vector layers, etc., and requires only a small training dataset (~100 samples / class)	Training model with new data, additional features and new classes (online training)	Existing as a software

Note: With regard to discussions during the kick-off meeting all criteria are given the same weight.

Picterra Sàrl 2017, all rights reserved

5. State-of-the-art

Two major axes for automatization are present in the Arealstatistik workflow: the land cover and land use change detection and the classification, i.e. automatic interpretation of the classes. Based on the classification results and data of a preceding Arealstatistik survey, change detection can be performed in order to assess if there is a need for further interpretation/classification. If a change is detected, then automatic classification or visual interpretation can be performed in order to determine the new land cover and land use class.

A schematic of this workflow, being it automatic or not, is shown in Figure 3:

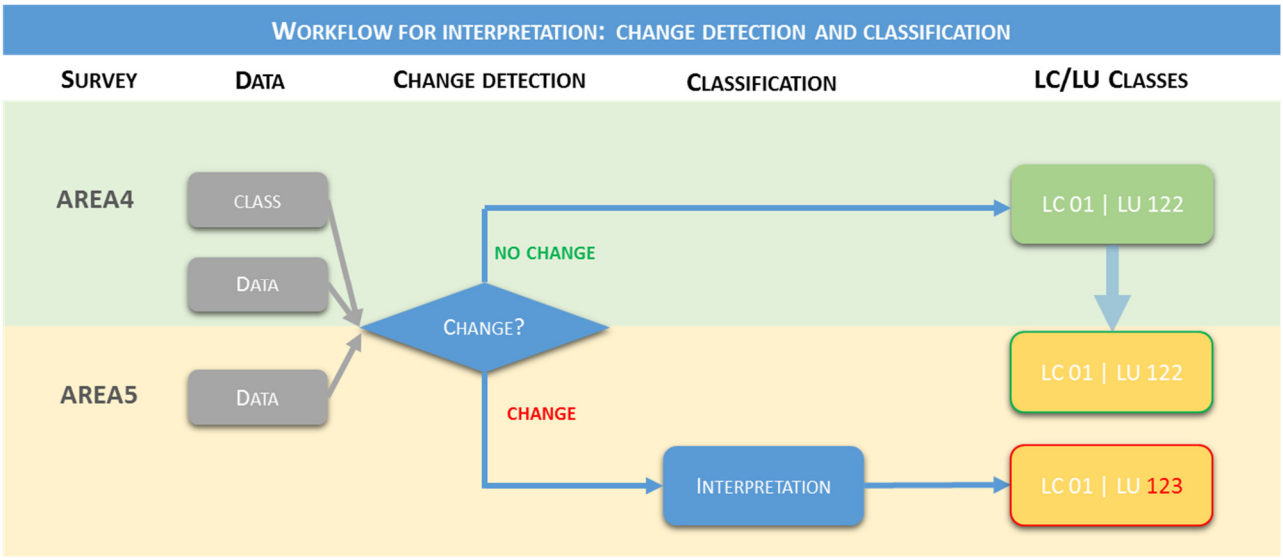


Figure 3 - Flowchart of the LC/LU classification with change detection

In automatic classification and change detection in aerial imagery, key methodological components are directly derived from the strong current of research and development within the machine learning & computer vision communities. The remote sensing community is increasingly quickly adopting and transferring such components to remote sensing challenges such as aerial land cover/use classification or change detection.

The methods can always be separated into three major groups of components: the type of features to be extracted, the learning strategy and the classification model, also called decision function, used. Therefore, each method for aerial land cover classification includes these three groups of components and the possibilities in terms of combination are extremely large. However, there are certain typical scenarios of combination of components schematized in Figure 4, which have been broadly used and have proven to work well together.

This state-of-the-art is organized as follows: Section 5.1 reviews these different groups of components, section 5.2 reviews and assesses the land cover/land use classification methods and finally section 5.3 reviews and assess the change detection methods.

Picterra Sàrl 2017, all rights reserved

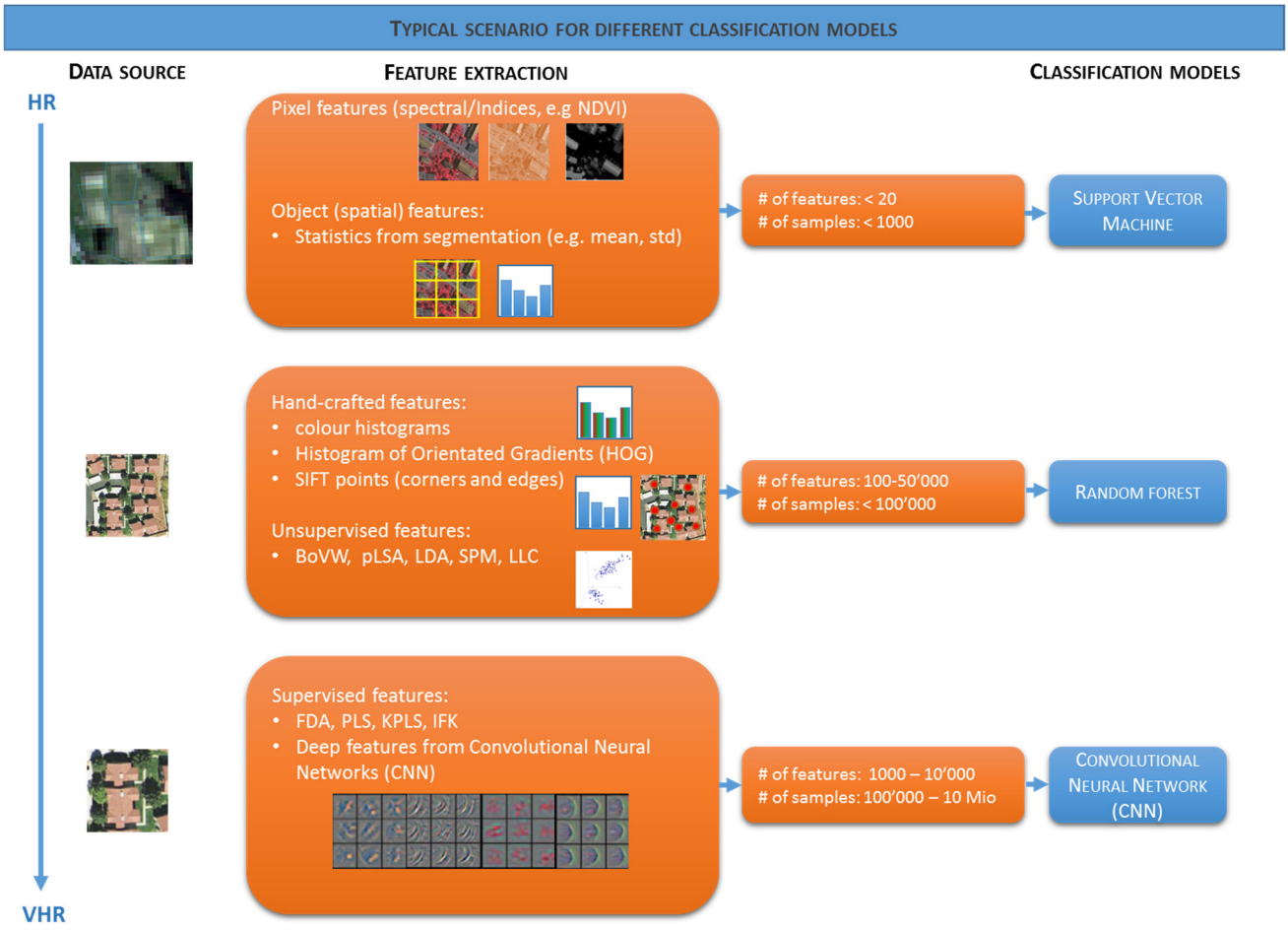


Figure 4 - Typical combination of data sources, feature extraction, selection and classification models.

5.1. Machine learning and computer vision review

5.1.1. Learning strategies

In general, the learning algorithms are grouped based on the training data available: "Is there labeled data available?", "Is there raw unlabeled data available additionally?", "To what extent?", etc. These different situations define the following categories of learning strategies: unsupervised, supervised or semi-supervised learning (Camps-Valls et al. 2009), and active learning (Tuia et al. 2011). The major learning strategies are presented in the following sub-sections.

5.1.1.1. Unsupervised learning

The aim of unsupervised learning is to find similarities among unlabeled data samples in order to group them. Most of the unsupervised learning algorithms assume that distinct clusters in the data correspond to different labels. This is often used as pre-processing step in order to reduce the amount of data when learning a model by either learning on each cluster or learning using only the cluster centers. Typical unsupervised classification algorithms are k-means or partitioning tree. A semantic class defined by a user (e.g. typically a land use class) is not meant to be equivalent to a cluster in the data. This only becomes true when the data features are carefully designed for the user needs and when the similarity criterion reflects the class similarities. In situations where this assumption is granted, meaning that the clusters or groups of clusters are reflecting the classes of interests, the unsupervised clustering algorithms will be able to provide the user with classes of interest. Therefore, unsupervised learning algorithm are to be considered in the frame of this study as a pre-

Picterra Sàrl 2017, all rights reserved

processing step allowing to either reduce the number of features, i.e. dimensionality reduction, or reduce the number of data samples by partitioning the data before performing supervised learning.

5.1.1.2. Supervised learning

Supervised learning algorithms are trying to find the boundaries of classes of interest, defined in the training data by labels attached to each data sample. The training stage aims at finding parameters of the model, which minimizes the error on the training set of samples, i.e. minimizing the empirical risk. However, in most supervised models a regularization of the model complexity is introduced to avoid overfitting the training data and therefore better generalize to unseen data samples. The trade-off between the complexity of the model and its prediction error directly relates to the bias and variance trade-off. Figure 5 below schematizes this trade-off.

In practice, a model with a high bias would always predict class LC 41 instead of real class LC 21, no matter how many times the classification if performed. On the other hand, a model with high variance, therefore overfitting the training data, would be too specific andwould predict randomly one class from the set { 41, 42, 43, 44, 45, and 46} instead of LC 21. This lack of generalization capability due to overfitting is regularized with the model complexity as mentioned previously.

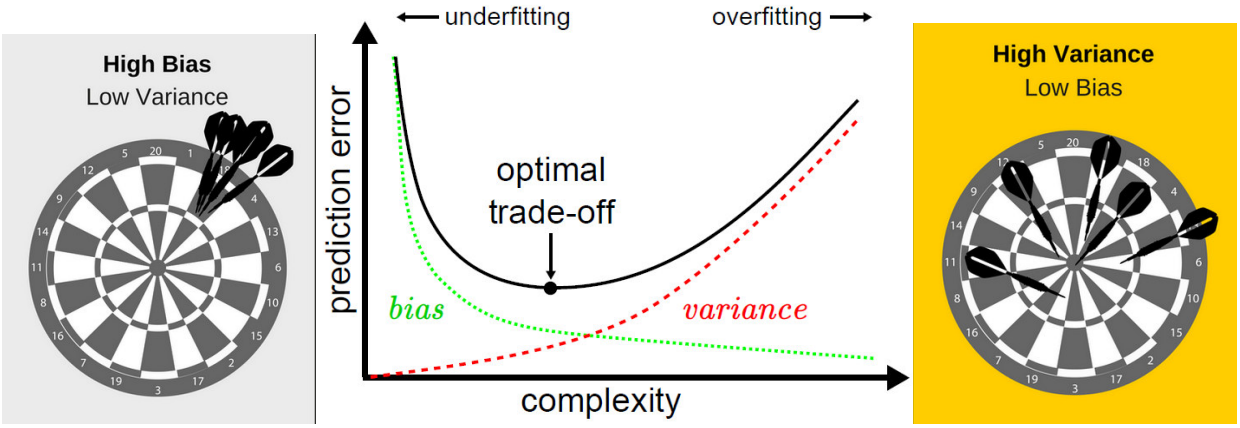


Figure 5 - The bias and variance in prediction with respect to the model complexity. For complex models, the bias is low while the variance is large. The optimal trade-off is found for a certain model complexity. Figure from https://elitedatascience.com/bias-variance-tradeoff

In practice, the problem of overfitting is predominant since training sets are usually limited and complex models are available. Therefore, regularization is performed at two different levels: inside the model, which during its learning phase penalizes complex parameters in order to provide the simplest model possible; and by performing cross-validation. Cross-validation splits artificially the training set in multiple folds, where one is kept apart during training to evaluate (test) the prediction error on unseen data. For example, if not enough data to allow some of it to be kept back for evaluation (testing). At the extreme is leave-one-out cross-validation (LOOCV) in which the test data consists of only one sample and the training is made of the remaining data.

The most common supervised classification models are *k*-NN (the *k* nearest neighbors in the training set are defining the predicted class), the naïve Bayes classifier (class probability density functions are known and maximum likelihood is defining the predicted class), decision trees (tree “leaves” are defining the predicted class), Support Vector Machines (decision boundary with maximum margin between the classes is found) or neural networks.

Picterra Sàrl 2017, all rights reserved

5.1.1.3. Semi-supervised learning:

Semi-supervised algorithms are designed by using the following paradigm: labeled data is typically limited because difficult and costly to obtain, unlabeled data is easily available in large quantities (e.g. all data outside the points on the 100x100m grid considered for the Arealstatistik). Therefore, exploiting both data types (labeled, unlabeled) can increase the classification accuracy by considering that:

- The class boundaries are forced to lie in regions where unlabeled data is less present (low density regions). In this situation, each data cluster is supposed to have a single class label.
- Similarities among all data samples are learned and represented on a graph. The labeled data samples can then propagate their class label information on the graph u until a global stable state is obtained over the graph and all data samples are labeled.

Typically, semi-supervised learning allows to gain in accuracy and to reduce the interpretation effort. However, it requires that the above assumptions (low density separation, data cluster equivalent to a single class) hold true and that massive amounts of data can be handled.

5.1.1.4. Data augmentation

Another way to overcome the limited size of a training set, i.e. of labelled data, is to perform “data augmentation”. This method applies a set of transformations to the training data samples: rotation, scaling, illumination changes, translations, etc. This approach artificially increases the size of the training set and allows the classification model to be more robust to such transformations after training. This is typically used with convolutional neural networks, which require a large set of training samples.

In the context of the interpretation made for the Arealstatistik, the sampled point is at center of the interpretation area, which makes the classification insensitive to rotation and translation. All the classes could benefit from rotation, flipping and illumination transformation. Yet, an increase in accuracy is not guaranteed however the robustness is strengthened.

5.1.1.5. Active learning

The Arealstatistik surveys demonstrated the labor intensity of labeling data. Labeled data is the base for the survey statistics, and also for any machine learning to perform automatic classification or change detection. In this context, Active Learning (AL) enriches both the information provided as input to the supervised classifier and improves the classifier's performance by iteratively developing the ground truth according to a process involving an interaction between the expert user and the automatic classification system. AL algorithms improve the accuracy of a model while minimizing the number of required labeled training data. The classification model is first trained on a small part of the labeled data, then the model returns the sample that has the most uncertain classification result and is most likely incorrectly classified to be interpreted and labeled by the expert user. The classification model is then re-trained from these new training data to increase the generalization performance of the model. This continues iteratively until the desired accuracy is obtained or a certain number of samples have been interpreted.

The general principle of AL is the ranking of the data sample according to a defined heuristic, which takes into account the following concepts:

- Uncertainty: Data samples to be labeled are the most uncertain ones, being at the boundary between two or more classes. Therefore, once labeled, these samples will have a strong impact on the class boundaries.
- Diversity: Data redundancy is reduced, avoiding the labeling of similar data samples to maximize the impact of newly labeled data samples on the classification model.

More precisely, three main families of heuristics exist: the committee-based heuristics, which are based on multiple classification models exploit the level of disagreement between the models to rank the samples, the

Picterra Sàrl 2017, all rights reserved

large margin-based heuristics which are based on the distance to the class boundaries, and the posterior probability-based heuristics which use the class posterior probabilities to find diverse uncertain samples.

A large number of heuristics have been proposed in the literature, including diverse components specific to the classification model being used, the type of data or the way the labels are collected. For example, a heuristic that takes into account the geographic location of the sample in relation to the previous ones and the presence of nearby roads to optimize the in-situ validation with a motor vehicle has been proposed (Demir et al. 2012).

It has been shown that in remote sensing applications (Tuia et al. 2011, Stumpf et al. 2014, Rougier et al. 2016), active learning and the combination of active learning and semi-supervised learning (Munoz-Mari et al. 2012) could reduce significantly the amount of labeled data required during training and therefore the expert interpretation time and also help for adapting a classifier to new data (Tuia et al. 2012). In the case of very high-resolution imagery, adaptations to segmented images have been proposed (Vargas et al. 2014, Stumpf et al. 2012). In the specific case of change detection, some active learning strategies have been applied (Alhichri et al. 2012), however it has attracted less attention from the machine learning community since it is a simpler case with only two classes (change, non-change).

5.1.2. Feature extraction

The feature extraction process has a key role in automatic classification and detection. The trend in machine learning moved from first focusing mainly on classification models (decision function) to focusing on feature extraction approaches. Only recently the combination of both features and decision function learned altogether has driven a strong current of research, with Convolutional Neural Networks (CNNs) being at the forefront. The CNNs learn to extract features and perform classification from training data.

Classification and change detection methods started at the pixel level, exploiting directly the pixel's spectral values as data features. However, considering single pixels has quickly been insufficient to describe correctly the class or change of interest, due to the increasing spatial resolution of satellite sensors now meeting the spatial resolution of aerial sensors. With such meter or sub-meter spatial resolution, a pixel is lacking high-level semantic information of the classes, which is needed to perform automatic land cover / land use classification. This spatial contextual information can be extracted in an unsupervised way using hand-crafted features, where engineering knowledge about the problem is used to extract the most discriminative information. However, when a large amount of training data is available the features can be extracted in a supervised manner. This aims at providing the classification model with a set of discriminative features for the classification problem at hand.

A review of the different type of features (Cheng et al. 2017, Xia et al. 2017) is performed hereafter. Features can be separated in three major groups: hand-crafted features which are extracted in a pre-defined way, Unsupervised learning of features which is directly driven by the data distribution and supervised learning of features which is based on a training set of labeled data. Further grouping of features can be established based on the level of semantic information captured by the features: low, mid, and high levels.

5.1.2.1. Hand-crafted and unsupervised feature learning

In aerial scene classification, the information is carried by a set of characteristics including spectral, colour, texture, shape, keypoints (corner, edges), composition of objects, and most often of a combination of characteristics. Hand-crafted features are trying to extract these different types of characteristics in order to best describe a scene.

- **Pixel-based features**: spectral information or other colour transformations (e.g. HSV). Such features do not exploit any contextual information and are only used in remote sensing applications with low spatial resolution.
- **Object-based features**: given a segmentation of the image, shape features can be extracted at each segment (perimeter, area, width/length ratio, eccentricity, etc.). Also spectral statistics of the

Picterra Sàrl 2017, all rights reserved

segments (mean, variance, min/max, etc.) and morphological descriptors can be exploited (Aptoula 2014). The drawback of object-based features is the need for reliable and accurate segmentation, which is challenging to reliably ensure, and the low level of semantic information captured. Such features were mainly used for land cover classification applications.

- **Scene-based features**: an entire image scene, which can represent quickly tens or hundreds of thousands of pixels, needs to be summarized in fewer but relevant information. Some approaches are directly quantizing the information, others are first detecting certain salient elements (edges, corners) beforehand.
    - **Hand-crafted**: The most widely used descriptor of an image is its color histogram (CH). It has the advantage of being rotation and translation invariant, however it is insensitive to the spatial distribution of the pixels. Therefore, it has been mainly used as a complementary feature for aerial scene classification (dos Santos et al. 2010, Yang et al. 2010). The extraction of spatial information has been widely exploited in remote sensing image analysis, with texture descriptors such as grey level co-occurrence matrix or Local Binary Patterns (Newsam et al. 2004, dos Santos et al. 2010, Aptoula 2014), edge-based descriptors such as GIST or Histogram of Oriented Gradients "HOG", which both are made of histograms involving filters at different scales and orientations (Cheng et al. 2014), keypoint descriptors such as Scale Invariant Feature Transform "SIFT", which first detects keypoints in the image such as corners and strong edges and then describe them using histograms of edge orientations in a similar way than GIST and HOG descriptors. The SIFT descriptor has several extensions or adaptations such as SURF (faster), DAISY (not invariant to scale and rotation), BRIEF (binary comparisons invariant to contrast and brightness), ORB (same as BRIEF but invariant to rotation), BRISK (as ORB but invariant to scale), or FREAK (based on human vision principles).
    - **Unsupervised**: The previous hand-crafted features become obsolete and suboptimal when subtle class differences are present and large amount of data is available. For example, clustering methods (see section 5.1.1.1) and dimensionality reduction techniques such as PCA, or autoencoder neural networks (Zhang et al. 2015) can provide adapted "words" from which the scenes can be best described. The words are typically formed by vector-quantizing local features with a clustering method. For example, the Bag-of-Visual-Words "BoVW" method uses k-means for the quantization of SIFT, GIST or HOG local features to create words and produces histogram of word occurrences (Yang et al. 2010, Xu et al. 2010, Zhao et al. 2016, Zhao et al. 2016b), while sparse coding methods produce a linear combination of words created using dictionary learning approaches (Sheng et al. 2012, Zheng et al. 2013, Cheriyadat 2014). A number of methods are trying to overcome the limitations of hand-crafted features by combining several local features under constraints of the general data structure: Probabilistic Latent Semantic Analysis (pLSA), Latent Dirichlet allocation (LDA), Spatial Pyramid Matching (SPM), or Locality-constrained Linear Coding (LLC) (Cheng et al. 2017).

5.1.2.2. Supervised feature learning

For large training datasets, the features are best extracted in a supervised manner to provide the classifier directly with discriminative features.

- **Supervised dimensionality reduction methods**: the label information of the training data can be exploited to select the best discriminant features (LDA, FDA), or to find new linear combination of features (PLS, CCA). However, these methods are limited since unable to produce non-linear combination very often required in remote sensing. To overcome such limitation, the kernel trick (also used in SVM, see section 5.1.3.1) allows to non-linearize these methods (KFDA, KPLS, KCCA). Further extensions have been proposed such as the Improved Fisher kernel (IFK), or the semi-supervised PLS (SSKPLS) (Izquierdo-Verdiguier et al. 2014).
- **Deep feature learning methods (CNN-based)**: a major breakthrough has been made in the computer vision community with the CNNs, see section 5.1.3.3, which in 2012 provided a strong improvement in the ImageNet image recognition challenge. Since then, pre-trained CNNs have been used in very

Picterra Sàrl 2017, all rights reserved

different applications with very good results when sufficient training data is present. However, an interesting observation has been that even a pre-trained CNN on the ImageNet challenge, being images of everyday scenes, could perform extremely well on a very different task such as aerial scene classification (Razavian et al. 2014, Penatti et al. 2015). The only adaptation being to remove the last layers, tuned for the original classification, and using the outputs of the last convolutional layer as features. Some examples of CNN layer filters are presented in Figure 6.

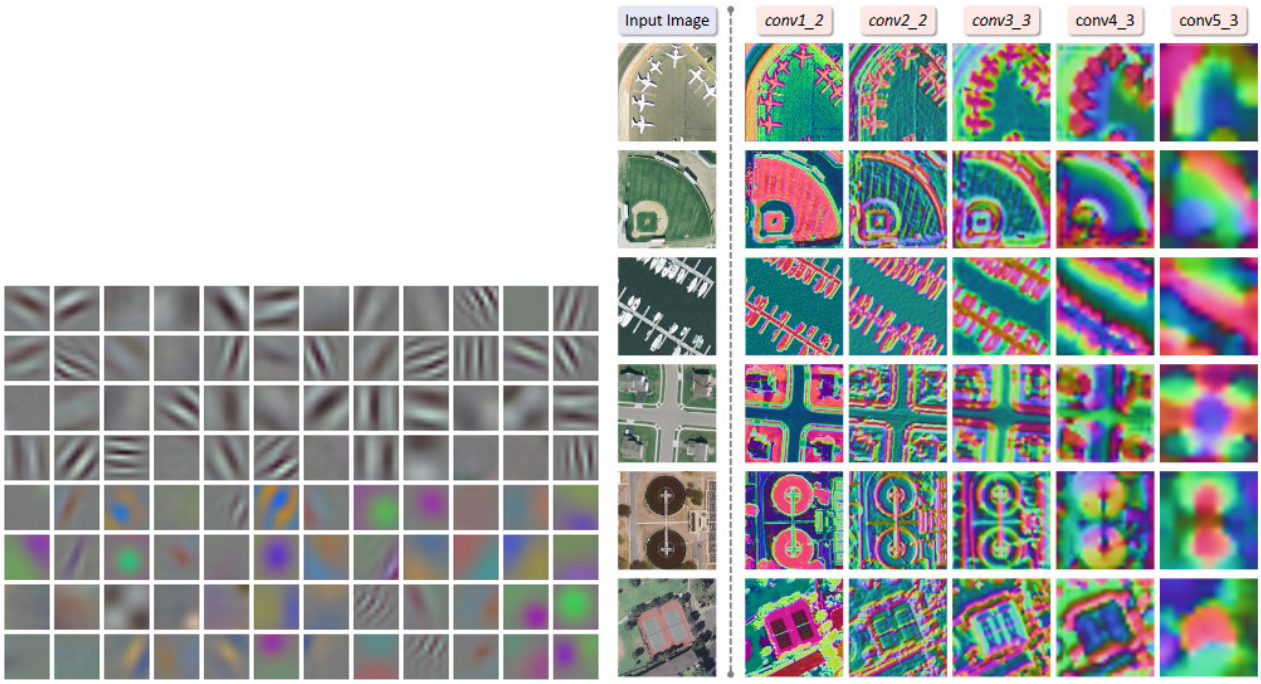


Figure 6 - Left: Examples of CNNs type of filters performed by the first convolutional layers (Castellucio et al. 2015). Right: Examples of feature maps derived from five different convolutional layers of VGG-VD-16 model (Hu et al. 2016)

5.1.3. Classification models

5.1.3.1. Support Vector Machines

SVMs are nonparametric statistical models able to deal with supervised classification and regression problems. Therefore, they have no assumption about the underlying data distribution. In the original SVM formulation, the model is presented with a set of data samples, and the SVM training algorithm aims at determining a hyperplane that discriminates training data samples from different classes, in a manner which optimally minimizes the empirical risk of error, i.e. maximizing the margin between the data samples at the boundary between the classes (Vapnik 2000). By finding the unique optimal solution, the SVM is having the lowest bias and variance for the given training data (Valentini et al. 2004).

SVMs allow non-linear separation of classes, more suited to remote sensing data, thanks to the usage of kernel functions. The crucial part for any kernel technique, including SVM, is the correct definition of a kernel function and its parameters that accurately reflects the similarity between the samples. The kernels commonly used with the SVM and other kernel-based classifiers are 1) linear kernel, 2) polynomial kernel, 3) radial basis function, or 4) sigmoid kernel. Radial basis function "RBF" kernels have been widely accepted as the most generic and able to meet the requirements of most classification problems. The kernel parameters are estimated by cross validation, as explained in section 5.1.1.2.

Over the past decade, numerous studies have been published in the scientific literature on the application of SVM classifiers in the analysis of RS imaging (Camps-Valls et al., 2009, Mountrakis et al., 2011, Maulik et al. 2017). Since the acquisition of labeled samples is costly and time-consuming, researchers have developed many SVM-based techniques to address this problem. In this context it should be mentioned that semi-supervised classifiers have been developed to solve problems that are characterized by a very low ratio

Picterra Sàrl 2017, all rights reserved

between the number of training samples available and the number of features (high dimensionality of the data), known as the Hughes phenomenon, by strengthening the learning process with the use of unlabeled data samples.

The S3VM is one of the semi-supervised alternative of SVM, which takes advantage of labeled and unlabeled samples in the learning phase. It attempts to push separations between classes in areas of low data density. It has shown very good performances in the classification of high-resolution satellite images (Camps-Valls et al., 2011).

5.1.3.2. Decision trees and Random Forest

Decision tree based learning algorithms are considered one of the best and mostly used supervised learning methods (see Section 5.1.1.2 for details on supervised learning algorithms) as their predictive models offer high accuracy, stability and easy interpretation. Moreover, they cope with non-linear relationships quite well.

A decision tree is a set of rules organized in a hierarchical manner. Each rule applies to a single feature. Once a tree is trained, certain insights on the features can be deduced by looking for example at which feature is involved first at the top rule. Decision tree based classification algorithms are extremely popular in computer vision and robotic applications because being very fast, simple, and allowing for a non-linear classification that is often required. In addition, approaches have been proposed for obtaining very precise results and generalizing better than a simple decision tree, among which the so-called "boosting" approaches (Bailly et al. 2007) and the sets of decision trees (forests) are widely used. These algorithms have been adopted in the remote sensing community and have proved to be very effective in many land classification applications, including recently the French land use/cover classification (Inglada et al., 2017). General disadvantage of decision trees is that they are prone to overfitting (high variance). Ideal classifier must handle and minimize both error due to variance and due to bias (please refer to Section 5.1.1.2 for details).

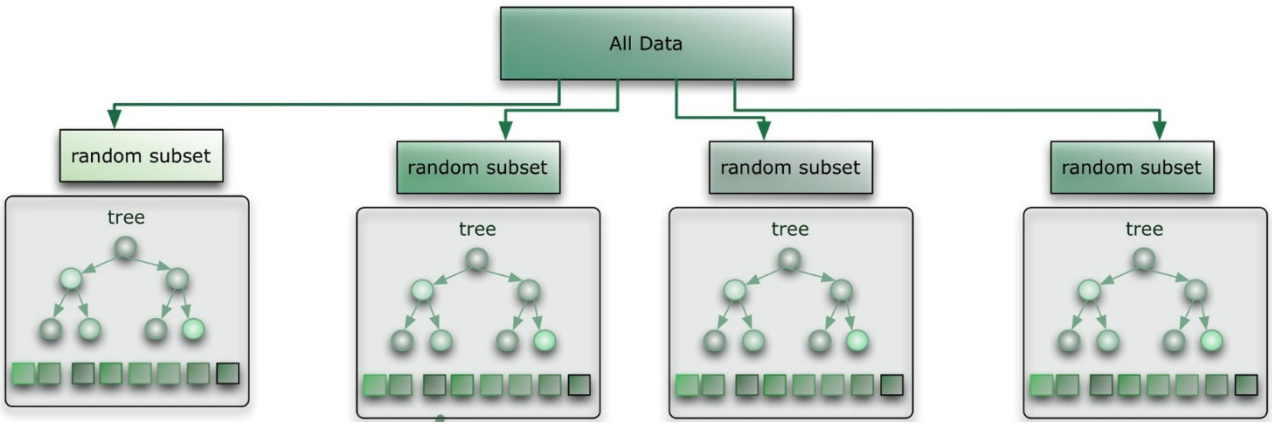


Figure 7 - Example of a random forest made of 4 decision trees. More detailed information on the decision tree methods can be found in this tutorial: https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/

The Random Forest, as a collection of decision trees whose results are aggregated into one final result using for example majority voting, can mitigate this issue very well and finds the optimal bias-variance trade-off (see Figure 5). It reduces the variance of prediction while retaining low bias, thanks to their inner mechanism, which are bootstrapping the dataset and splitting the feature space among decision trees (Svetnik et al. 2003). Bootstrapping is dividing the data into overlapping subsets, which are given to the trees for training. Therefore, each tree is specific to a part of the feature space and to a part of the training set. Random Forest can be trained in an online fashion, meaning training the forest not once but each time a new batch of data is available and without requiring all previously available data. One example of such implementation is the “Mondrian Forests” (Lakshminarayanan et al., 2014), which follows a tree-growing strategy.

Picterra Sàrl 2017, all rights reserved

5.1.3.3. Convolutional Neural Networks (CNN)

Deep learning has gained renewed interest in recent years when the development of very performant graphics processors (GPUs) has greatly improved and allowed the massive parallel processing of data. Artificial neural networks are models allowing a so-called "deep" learning because of their architecture including many layers of neurons. This architecture efficiently exploits the parallel processing of the training data to obtain the connection weights of the neurons producing the desired classification. The use of these methods has been democratized very recently by being applied in many fields including remote sensing. The enthusiasm for neural networks on problems of image processing or geo-spatial elements is strongly related to convolutional neural networks. The latter use image patches of different sizes and propagate them in the first layers of the network, making it possible to extract contextual characteristics adapted to the classification problem to be solved (Romero et al., 2016). CNNs convolutional layers act as feature extraction, see deep feature learning in section 5.1.2.2, and the last layers of the network, typically fully connected, act as a classifier and transform these features into a classification label. The ImageNet challenge of image categorical classification has revealed the discriminative power of CNNs, reaching unprecedented accuracies. Some of the available CNNs pre-trained on everyday scenes are: *AlexNet*, *CaffeNet, VGG-VD-16, GoogLeNet or Resnet,* wich differ in their architecture (number of layers, total number of parameters, etc.). For more details on CNNs and the ResNet architecture, please see the Appendix. The following table and Figure 8 summarize information about the CNN models pre-trained on ImageNet dataset.

CNN trained model	Input image size	Pre-processing	Number of layers	Total number of parameters	Performance benchmark in ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
AlexNet	224x224	Subtracting the mean RGB value, computed on training dataset, from each pixel	8	60'000'000	Winner 2012
GoogleLeNet			22	5'000'000	Winner 2014
VGG-16			16	138'000'000	2nd place 2014
ResNet			152	2'000'000	Winer 2015

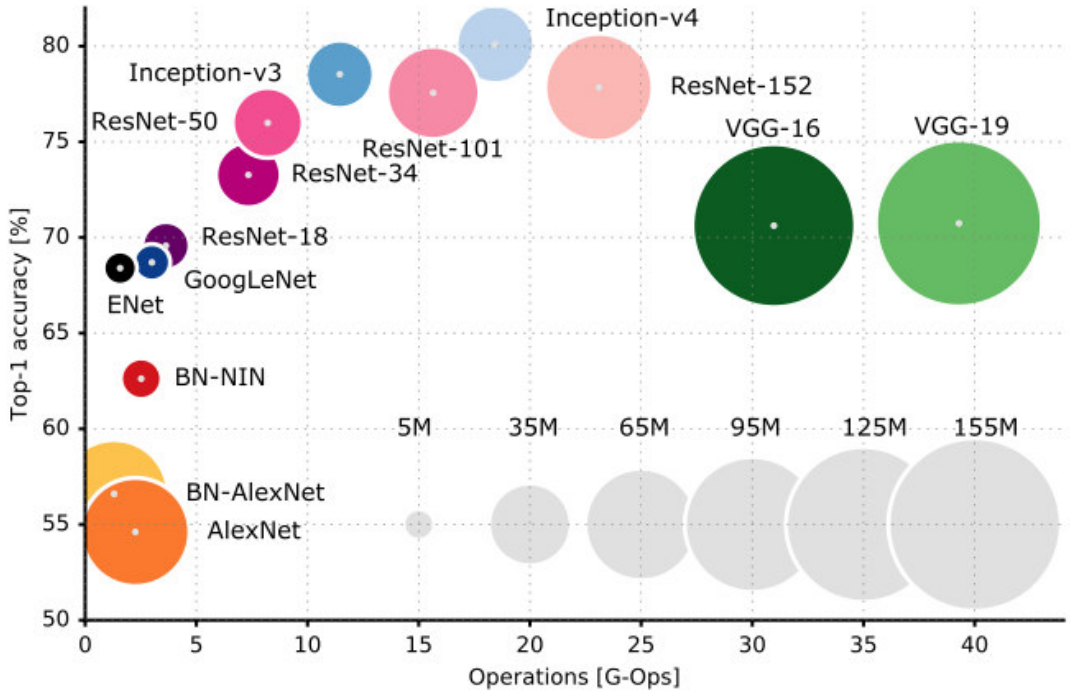


Figure 8 - Accuracy vs. efficiency in terms of operations required of selected CNN pre-trained models. The size of the disk representing the model is related to the total number of parameters.

In remote sensing applications, CNNs have shown extremely good performances with very high-resolution aerial imagery, for classification at the pixel level (dense labeling) (Sherrah 2016, Volpi et al. 2017), and at the

Picterra Sàrl 2017, all rights reserved

scene level (Penatti et al. 2015, Weng et al. 2017, Xia et al. 2017, Cheng et al. 2017). Semi-supervised learning with CNNs can be done by giving pseudo-label from maximum likelihood classification on unlabeled data and fine-tuning the network with this data (Dong-Hyun 2013). This brings the decision boundaries of the CNN in low density regions.

5.1.3.4. General comparison of classification models

The three major families of classification models: SVM, Decision trees and Random Forest, and CNNs, are compared in terms of general classification criteria which shows their differences and the situations in which they are best adapted. The Random Forest will have a great flexibility regarding the type of data it can handle, while the SVM will require well selected and normalized features. The CNN will have a strong modelling power for images, by integrating the feature extraction and classification in its learning process, but it will be limited in the integration of other data sources. The following table summarizes these different characteristics.

Pros/cons	Support Vector Machines (kernel methods)	Decision trees (boosted, random forest methods)	Convolutional Neural Network
Reduced interpretation			
Possibility to perform active learning	✓	✓	- (AL with pre-trained models only. Not really adapted to be trained with few samples.)
Accuracy			
Non-linear modelling	✓	- (random forest only)	✓
Possibility to model very different class types	- (dependent on the feature extracted)	- (dependent on the feature extracted)	✓
Robustness			
Robust to noise	✓	- (random forest only)	✓
No need for feature standardization (different range, continuous variables)	✗	✓	- incl. in certain architectures (e.g. batch normalization)
Robust to small training set	✓	-	✗
Robust to overfitting	✓	-	- incl. in most architectures (e.g. drop out layers)
Sustainability			
Handle large number of training samples	✗	-	✓
Handles multiple feature types (continuous, discrete, categorical, etc.)	✗	✓	✗
Handle large nb of features	✗	✓	✓
Adaptation cost			
feature extraction incl.	✗	✗	✓
Feature selection incl.	✗	✓	✓

Picterra Sàrl 2017, all rights reserved

5.2. Land cover / land use classification review

This chapter reviews land cover / land use classification approaches exploiting the strength of machine learning and computer vision with very high resolution remote sensing data. This completes the previous review of machine learning and computer vision approaches, where all the features and classification models were presented, by presenting the datasets used for benchmarking methods and analysing the type of classes that can be discriminated.

The land cover and land use classification based on remote sensing data has evolved from pixel-based methods relying mainly on the power of the classification model, to object-based and scene-based methods incorporating context by using hand-crafted features or convolutional neural networks. The contextual methods are bringing richer information which is required for land use classification and very beneficial to land cover classes also. As discussed in section 5.1.2, the usage of carefully chosen hand-crafted features or learned features from training data will depend on the classes to be discriminated and the amount and quality of the training data at hand.

Benchmark datasets

The review of the datasets used for benchmarking methods in aerial scene classification (land use / cover) shed an interesting light on the type of data exploited (e.g. spatial resolution, extent considered, number of images, etc.) and the type of classes which can be discriminated (e.g. number of classes, intra-class variability, etc.). The spatial resolution is most often similar to the data exploited for the Arealstatistik (25-50cm). Some datasets include even a range of spatial resolutions (RSSCN7 and NWPU-RESISC45) in order to increase the classification challenge. The extent of the image scenes considered are in a range from 50m to 600m, with an exception for a dataset, which includes low resolution satellite imagery providing large extent (7.6km). Such extents are adapted to the type of class to be discriminated and is in the range of the Arealstatistik grid of 100x100m.

The major datasets used in publications for benchmarking methods are described here below:

Datasets	Total # images	# Images / class	# classes	Spatial resolution (cm)	Image size (pixels)	Image extent (m)	Image source	Reference
UC-Merced	2100	100	21	30	256x256	77m	USGS National Map Urban Area Imagery collection	(Yang et al. 2010)
WHU-RS19	1005	50	19	50	600x600	300m	Google Earth imagery	(Sheng et al. 2012)
SIRI-WHU	2400	200	12	200	200x200	400m	Google Earth imagery	(Zhao et al. 2016)
RSSCN7	2800	400	7	20 to 150	400x400	80m to 600m	Google Earth imagery	(Zou et al. 2015)
RSC11	1232	100	11	20	512x512	102m	Google Earth imagery	(Zhao et al. 2016b)
AID	10000	330	30	50	600x600	300m	Google Earth imagery	(Xia et al. 2017)
NWPU-RESISC45	31500	700	45	20 to 3000	256x256	50m to 7.6km	Google Earth, Landsat	(Cheng et al. 2017)

All datasets are considering R,G,B color imagery only.

Datasets are including land use classes (e.g. commercial area, dense residential area, park, etc.), land cover classes (e.g. chaparral, snowberg) or even specifications type of classes (e.g. church, school). This demonstrates that the research community is facing similar challenges than the Arealstatistik land use / cover classification. The list of all the classes considered in these datasets with their most similar (if existing) class from BFS Arealstatistik catalogue is presented in the following table.

Picterra Sàrl 2017, all rights reserved

BFS LC/LU Class, Groups or Domains	UCMerced	WHU-RS19	SIRI-WHU	RSSCN7	RSC11	AID	NWPU - RESISC 45
LU 107 Agricultural building		farmland		farmland		farmland	rectangular farmland
							circular farmland
LU 2 Agriculture	agricultural		agriculture				
LC 21 grassland		meadow	meadow	grassland	grassland	meadow	meadow
	chaparral						chaparral
LU 420 unproductive land			idle land			bare land	
LU 161 Public park		park	park			park	
						square	
							terrace
LU 242/243 high mountain pasture		mountain				mountain	mountain
LU 300 Forest	forest	forest	forest	forest	dense forest	forest	forest
					sparse forest		
LU 402 Rivers LC 61 water surface	river	river	river	river		river	river
LU 401 Lakes		pond	pond			pond	
LC 61 water surface			water	lake			lake
							sea
							wetland
LC 62 Glacier, lasting snow “névé”							snowberg
							ice
							ship
	harbor	port	harbor		harbor	port	harbor
	beach	beach				beach	beach
							island
							palace
LU 100 Buildings	buildings				high buildings		
LU 105 Residential buildings	dense residential	residential area	residential	residential region	residential area	dense residential	dense residential
LU 104 house alignment	medium residential					medium residential	medium residential
LU 103 individual houses	sparse residential				low buildings	sparse residential	sparse residential
LU 101 & 102 industrial area		industrial area	industrial	industrial region		industrial	industrial area
		commercial area	commercial			commercial	commercial area
						center (shopping mall)	

Picterra Sàrl 2017, all rights reserved

LU 141 Power distribution area							thermal power station
LU 164 camping	mobile home park						mobile home park
LU 106 Public building A15 Church						church	church
LU 106 Public building A11 School						school	
LU 108 undetermined buildings						resort	
LU 124 railways		railway station				railway station	railway station
					railway		railway
LU 121 highways	freeway				roads		freeway
LU 122 roads							ground track field
		bridge				bridge	bridge
	overpass	viaduct	overpass		overpass	viaduct	overpass
	intersection						intersection
							roundabout
LU 123 parking lot	parking lot	parking lot		parking lot		parking	parking lot
LU 125 airports		airport				airport	airport
	runway						runway
	airplane						airplane
LU 162 Sport field						stadium	stadium
						playground	
	baseball diamond					baseball field	baseball diamond
							basketball court
LU 162 Sport field B11 Football field		football field					
LU 162 Sport field B12 Tennis, squash court	Tennis court						tennis court
LU 163 golf course	Golf course						golf course
LU 101 & 102 Industrial area A81 Storage tanks	Storage tanks				storage tanks	storage tanks	storage tank
		desert				desert	desert
							cloud

Sample images for the different categories of the datasets are presented in appendix.

Picterra Sàrl 2017, all rights reserved

In addition to these datasets for scene classification, the following two datasets are for benchmarking pixel-wise (dense labeling) classification into fewer classes. This type of classification may be of interest for the "multiples" group of Arealstatistik classes, made of a composition of multiple similar elements (e.g. trees) or a clear percentage of certain cover (e.g. snow cover in % for névé). Moreover, they integrate the near-infrared channel and the digital surface model in addition to the standard color channels, which is closer to the Arealstatistisk data setup.

Datasets	Total # images	Total # pixels	# classes	Spatial resolution (cm)	Image size (pixels)	Image extent (m)	Image source	Reference
Vaihingen	33 (16 training)	168'287'871	6	9	2494x2064	~200m	ISPRS WG III/4. ISPRS 2D Semantic Labeling Contest	(Gerke 2015)
Potsdam	38 (17 training)	1'368'000'000	6	5	6000x6000	300m		Potsdam dataset link in section 5.5.

All datasets are considering NIR, R, G, and a normalized DSM (or DHM). Potsdam dataset contains additionally the blue spectral band.

Different approaches relying on CNNs have been proposed to produce a dense pixel-wise labeling, with an architecture avoiding down-sampling (Sherrah 2016) or with a deconvolutional stage learning how to up-sample the class scores back to the original resolution (Volpi et al. 2017). The integration of several data sources: spectral (NIR-R-G) and surface (normalised DSM, nDSM) data has been investigated using the FuseNet deep neural network architecture (Eitel et al. 2015). Moreover, the multi-class detection of geospatial objects (cars, tanks, sport fields, etc.) has been investigated with a collection of part detectors (Cheng et al. 2014b). In Figure 9 are presented examples of dense labeling on these two datasets.

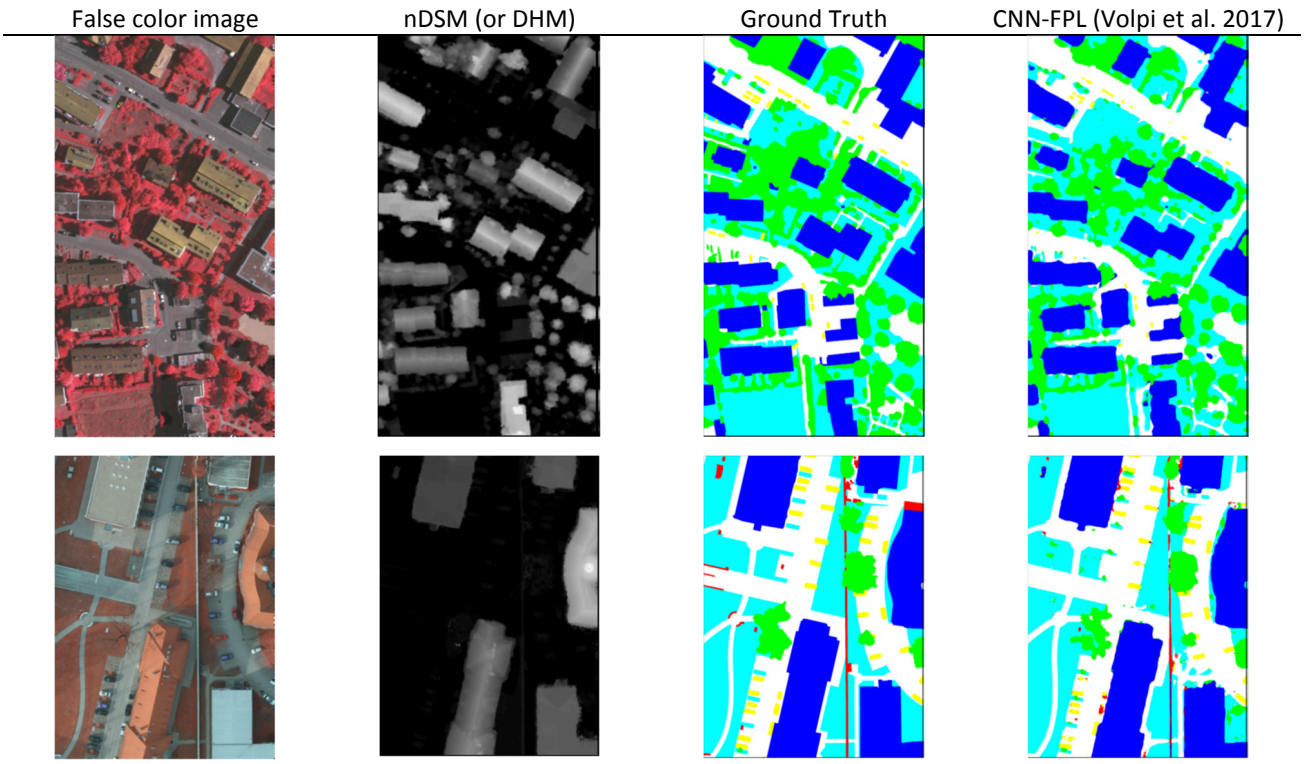


Figure 9 - Example for the Vaihingen (top row) and Potsdam (bottom row) datasets.
Legend: White = impervious surface, Blue = buildings, Cyan = low vegetation, Green = trees, Yellow = cars, Red = clutter.

Picterra Sàrl 2017, all rights reserved

Approaches description and evaluation:

In this sub section, the following approaches are described and evaluated:

Name	ID
Land cover/use classification using aerial imagery and handcrafted features	A1
Land cover/use classification using aerial imagery and unsupervised features	A2
Land cover/use classification using aerial imagery and supervised deep features	A3
Dense land cover/use classification using aerial imagery and supervised deep features	A4
Land cover classification using satellite imagery and Random Forest	B1
Land cover classification using satellite imagery and Active Learning	B2

Here follows the description and evaluation of the different approaches.

Method	ID: A1	
Description	**Land cover/use classification using aerial imagery and handcrafted features** Automatic land cover/use classification from aerial VHR data by extracting handcrafted features from aerial scene and performing linear SVM classification	
Application ranges	Classification of land cover/use/specifications (up to 45 classes for NWPU-RESISC45 dataset)	
Features	Handcrafted features: SIFT, LBP, Colour Histogram (CH), GIST	
Classification model	Linear SVM (Xia et al. 2017) as classification model	
Data and accuracies	UC Merced dataset (Training 80%): 46.90% +- 1.76 OA, GIST (Xia et al. 2017) WHU-RS19 (Training 60%): 51.87% +- 3.40 OA, CH (Xia et al. 2017) RSSCN7 (Training 50%): 60.54% +- 1.01 OA, CH (Xia et al. 2017) AID dataset (Training 50%): 37.28% +- 0.46, CH (Xia et al. 2017) NWPU-RESISC45 (Training 20%): 27.52% +- 0.14 OA, CH (Cheng et al. 2017) The above OA are representing the mean and standard deviation after repeating 10 times the experiment, for the best type of features.	
Authors/References	(Xia et al. 2017) (Cheng et al. 2017)	
Development status (« Technology Readiness Level »)	**High readiness level**: Most features (SIFT, LBP, CH) available in open source libraries (e.g. openCV). http://opencv.org/	
License	BSD *license (OpenCV, python scikit-image)*	
		Score
Reduced interpretation	Only certain classes could potentially benefit of a reduced interpretation from these handcrafted features	1
Accuracy	Accuracy levels are below 60% overall classes. For the class group “single” accuracy could be probably higher reaching an acceptable level (>85%).	2
Robustness	Features are robust to a certain number of effects (illuminations, scale, etc.), however it will need a large amount of data to handle class variability.	3
Sustainability	Model can be retrained efficiently with new data (Linear SVM can handle online learning approaches).	3
Adaptation cost	The features can be extracted from the aerial data and DSM and added to the auxiliary data to a standard classifier. All elements are existing in standard toolboxes.	4

Figure 10 - Method A1: Land cover/use classification using aerial imagery and handcrafted features

Picterra Sàrl 2017, all rights reserved

Method	ID: A2
Description	**Land cover/use classification using aerial imagery and unsupervised features** Automatic land cover/use classification from aerial VHR data by exploiting handcrafted & unsupervised mid-level features to perform aerial scene classification
Application ranges	Classification of land cover/use/specifications (up to 45 classes for NWPU-RESISC45 dataset)
Features	Handcrafted features: BoVW, BoVW+SPM, LLC (Cheng et al. 2017), IFK, LDA, pLSA, SPM
Classification model	Linear SVM (Xia et al. 2017) to classify
Data and accuracies	UC Merced dataset (Training 80%): 83.79 +- 2.32 OA, IFK (CH) (Xia et al. 2017) WHU-RS19 (Training 60%): 86.95 +- 1.31 OA, IFK (SIFT) (Xia et al. 2017) RSSCN7 (Training 50%): 86.77 +- 1.11 OA, IFK (CH) (Xia et al. 2017) AID dataset (Training 50%): 77.33 +- 0.37 OA, IFK (SIFT) (Xia et al. 2017) NWPU-RESISC45 (Training 20%): 44.97% +- 0.28 OA, BoVW (Cheng et al. 2017) The above OA are representing the mean and standard deviation after repeating 10 times the experiment, for the best type of features.
Authors/References	(Xia et al. 2017) (Cheng et al. 2017)
Development status (« Technology Readiness Level »)	**Medium readiness level**: Applied research Detailed publications, some published codes: https://github.com/briansrls/SIFTBOW (BovW) https://github.com/kushalvyas/Bag-of-Visual-Words-Python (BovW) https://github.com/yassersouri/Maahigir (IFK)
License	BSD *license (OpenCV used in the codes above)*
	Score
Reduced interpretation	Only certain classes can benefit of a reduced interpretation from these unsupervised features, since high enough accuracies were only obtained on RSSCN7 and WHU-RS19 classes. — 2
Accuracy	Accuracy levels are close to 85% for the three simpler datasets, at 77% for AID dataset with more complex classes and 45% for the most complex one. The class group “single” and “mixed” could reach an acceptable level (>85%). — 3
Robustness	Features are robust to a certain number of effects (illuminations, scale, etc.), however it will need a large amount of data to handle class variability. — 3
Sustainability	Compared to purely handcrafted features (A1), once new data is present all features have to be recomputed. — 1
Adaptation cost	The features can be extracted from the aerial data and DSM and added to the auxiliary data to a standard classifier. Not all elements are existing in standard toolboxes. — 3

Figure 11 - Method A2: Land cover/use classification using aerial imagery and unsupervised features

Picterra Sàrl 2017, all rights reserved

Method	ID:	A3
Description	**Land cover/use classification using aerial imagery and supervised deep features** Automatic land cover/use classification from aerial VHR data by exploiting deep features obtained from CNNs pre-trained on massive computer vision datasets and transferred to remote sensing aerial imagery. This transfer is performed by either learning a classifier based on these features (SVM, Random Forest, etc) or by fine-tuning the pre-trained CNNs last fully connected layers (Cheng et al. 2017). The fusion of several pre-trained models together has shown the best results on UC Mercred dataset (Penatti et al. 2015).	
Application ranges	Classification of land cover/use/specifications (up to 45 classes for NWPU-RESISC45 dataset)	
Features	Extracted features from pre-trained CNNs: AlexNet, GoogleLeNet, VGG-16, OverFeat, CaffeNet	
Classification model	Linear SVM (Cheng et al. 2017, Xia et al. 2017) or extreme learning machine (Penatti et al. 2015) when CNNs used as feature extraction only, otherwise CNN last fully connected layers act as classifier. Extension of CNN to semi-supervised classification with pseudo-labels on unlabeled data for fine-tuning (Dong-Hyun 2013).	
Data and accuracies	UC Merced dataset: 99% OA, OverFeat + CaffeNet (Penatti et al. 2015) WHU-RS19 (Training 60%): 96.24 +- 0.56 OA, CaffeNet or VGG-16 (Xia et al. 2017) RSSCN7 (Training 50%): 88.25 +- 0.62 OA, CaffeNet or VGG-16 (Xia et al. 2017) AID dataset: 89.6% +- 0.3, CaffeNet or VGG-16 (Xia et al. 2017) NWPU-RESISC45 dataset: 79.8% +- 0.13 OA, VGGNet-16 or AlexNet (Cheng et al. 2017) The above OA are representing the mean and standard deviation after repeating 10 times the experiment, for the best type of features.	
Authors/References	(Xia et al. 2017) (Cheng et al. 2017), (Penatti et al. 2015)	
Development status (« Technology Readiness Level »)	**High readiness level**: Many pre-trained models available in open source libraries (python Keras, Theano, Tensorflow, Caffe, etc.) and cloud computing libraries, e.g. Microsoft Azure Cognitive Toolkit (AlexNet, GoogleLeNet, ResNet, VGG-16, CaffeNet).	
License	-	
		Score
Reduced interpretation	All classes can benefit of a reduced interpretation from these supervised deep features, since high accuracies were obtained on all classes.	4
Accuracy	Accuracy levels are close to 100% for UCMerced dataset, close or above 90% for the other datasets, except NWPU dataset which has an additional complexity with a mixture of resolutions. All the class groups (“single”, "multiples" and “mixed”) could reach an acceptable level (>85%). Class group "multiples" may need a specific processing to reach an acceptable accuracy level.	4
Robustness	Features are robust to a large number of effects (illuminations, scale, orientations, organisations, etc.), however it will need a significant amount of data to adapt to the classes and learn appropriate semantic level of features.	4
Sustainability	Model can be retrained with new data (data and classification error backpropagated through the network), but classifier needs to be trained from scratch afterwards since features produced by the network may have changed after retraining. If using CNN last layers as classifier, retraining is possible.	3
Adaptation cost	The features can be extracted from the aerial data and DSM and added to the auxiliary data to a standard classifier. All elements are existing in standard toolboxes.	4

Picterra Sàrl 2017, all rights reserved

Figure 12 - Method A3: Land cover/use classification using aerial imagery and supervised deep features

Method	ID:	A4
Description	**Dense land cover/use classification using aerial imagery and supervised deep features** Automatic land cover/use classification from aerial VHR data by exploiting CNN to directly retrieve at each pixel of a scene (dense labeling) the class information.	
Application ranges	Classification of land cover and land use (6 classes)	
Features	Learned features inside the CNN architecture: either CNN trained from scratch (Volpi et al. 2017), with pre-trained VGG model (Sherrah 2016) or trained CNN with OpenStreetMap labels and fine-tuned on real class labels (Maggiori et al. 2017)	
Classification model	CNN	
Data and accuracies	Aerial VHR multispectral and normalized digital surface model (height) data: Vaihingen dataset: 89.1% OA (Sherrah 2016) Potsdam dataset: 90.3% OA (Sherrah 2016)	
Authors/References	Researchers: M. Volpi, Uni. Zurich, Switzerland	
Development status (« Technology Readiness Level »)	**Medium readiness level**: Applied research detailed publications, research code downloadable https://sites.google.com/site/michelevolpiresearch/codes/dense-labeling	
License	Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) Creative Commons licensing.	
		Score
Reduced interpretation	Certain class groups can benefit of a reduced interpretation from this dense labelling, since high accuracies were obtained on specific classes.	3
Accuracy	Accuracy levels are close to 90% for the two datasets. The class group ("multiples") could strongly benefit fromthis and avoid confusion between very similar classes, by allowing precise measures on the reference area.	3
Robustness	Features are robust to a large number of effects (illuminations, scale, orientations, organisations, etc.), however it will need a large amount of data to adapt to the classes.	3
Sustainability	Model can be retrained with new data (data and classification error backpropagated through the network)	3
Adaptation cost	The auxiliary data should be integrated to the CNN architecture. Elements are not existing in standard toolboxes. Detailed publications can allow development in established deep learning framework (e.g. Theano, Tensorflow, CNTK, ..)	2

Figure 13 - Method A4: Dense land cover/use classification using aerial imagery and supervised deep features

Picterra Sàrl 2017, all rights reserved

Method	ID: B1	
Description	**Land cover classification using satellite imagery and Random Forest** Automatic land cover classification from Landsat-8 satellite data mapping the entire French metropolitan territory. Training data from Corine Land Cover map, French National Geo. Institute, Land Parcel Information System for agriculture. This reference data contains errors due to recent changes not updated, however these errors are overcome by the classification algorithm.	
Application ranges	Classification of land cover (17 classes) at 10m resolution	
Features	198 features: 6 spectral bands plus 3 radiometric Indices (NDVI, NDWI, brightness) for 22 different dates.	
Classification model	Random Forest with 100 trees, a maximum depth for each tree of 25 levels and a minimum number of 5 samples per node. Local model trained on each cell of a regional climatic zone grid (N cells).	
Data and accuracies	Satellite (Sentinel-2 and Landsat-8) Accuracy: 0.86 Kappa coefficients (max value = 1) with Landsat-8 (similar accuracy for Sentinel-2 but not mentioned explicitly) Accuracy is the mean over 10 experiments with each time different training and validation sets.	
Authors	Researchers: J. Inglada, CESBIO, France	
Development status (« Technology Readiness Level »)	**Medium readiness level**: Applied research detailed publications, classification output and tools http://www.cesbio.ups-tlse.fr/multitemp/?p=10104#more-10104 http://tully.ups-tlse.fr/jordi/iota2 Ioata2 processing chain for parallel processing Orfeo Toolbox 5.4 & OpenCV 2.11 (Random Forest)	
License	-	
		Score
Reduced interpretation	Land cover high level class groups (17 classes) can benefit of a reduced interpretation from this classification approach, since high accuracies were obtained using auxiliary not updated reference data.	3
Accuracy	Accuracy levels are around 0.86 Kappa for the 17 land cover classes. Land use classes are not addressed with this method and low resolution data will introduce errors for class with small extent around the point.	3
Robustness	Robustness and repeatability of model: High, local models can adapt to region specificities and preserve a high accuracy over time.	3
Sustainability	Model can be retrained with new data (random forest parameters stable, no need for performing a new parameter search). Moreover, trees can easily deal with heterogeneous data type and Random forest can accept easily additional trees in a majority voting setup for example.	4
Adaptation cost	Implementation complexity: Classification model by itself available (random forest), setup of regional models & data preparation time consuming.	3

Figure 14 - Method B1: Random Forest for national land cover/use mapping

Picterra Sàrl 2017, all rights reserved

Method	ID: B2	
Description	**Land cover classification using satellite imagery and Active Learning** Automatic land cover/use classification from very high-resolution data using active learning to minimize the user interpretation.	
Application ranges	Classification of land cover/use (9 classes) at 2.4m resolution Classes: Residential, commercial buildings, trees, vegetation, harvested fields, bare soil, roads, parking lots and water.	
Features		
Classification model	Support Vector Machine (RBF kernel) and different active learning strategies. The "Margin Sampling" and "Multiclass Level Uncertainty" are providing the best overall accuracies for the multispectral VHR dataset.	
Data and accuracies	Satellite (Quickbird) multispectral VHR (4 bands: NIR-R-G-B, 2.4m resolution) Accuracy: 75% with 50 samples, 90% with 350 samples ("Margin Sampling") and 95% with 1100 samples. Full training set of 7000 samples provides 97% accuracy. Random sampling of samples provides 5% less accuracy than active learning with the same number of samples. See (Tuia et al.) for confidence intervals on the accuracies plotted on charts.	
Authors	Researchers: D. Tuia, M. Volpi, L. Copa, M. Kanevski, J. Munoz-Mari. (Tuia et al. 2011)	
Development status (« Technology Readiness Level »)	**Medium readiness level**: Applied research detailed publications & Matlab toolbox available http://isp.uv.es/soft_classification.html ALTB "Active Learning Toolbox"	
License	-	
		Score
Reduced interpretation	Land cover high level class groups (9 classes) can benefit of a strong reduced interpretation from this classification approach, allowing to reach highest accuracies already with only a fraction of the reference data. Actually, this may be most effective within the change detection in order to set the change detection threshold efficiently with less training data representing changes than the classification benefiting from a large training set.	2
Accuracy	Accuracy levels are 90% with 350 samples ("Margin Sampling") and 95% with 1100 (15% of full training set) samples for the 9 land cover/use classes. 97% is reached with the full training set.	2
Robustness	Active learning strategies adapt to many different type of classifiers, by using committee of classifiers, probabilities of output or margin of SVM classifiers. Could provide on the specific set of classes high accuracies with around 1100 samples (~120 pixels / class).	4
Sustainability	Models can deal with heterogeneous data type with the committee of classifiers setup for example.	4
Adaptation cost	Some elements existing as toolbox (classification models, active learning strategies), adaptation to aerial data and different type of features/classification models (e.g. pre-trained CNNs) to be developped.	3

Picterra Sàrl 2017, all rights reserved

5.3.Change detection review

In this chapter, change detection approaches are reviewed which exploit the strength of machine learning and image processing in remote sensing (aerial/satellite) applications.

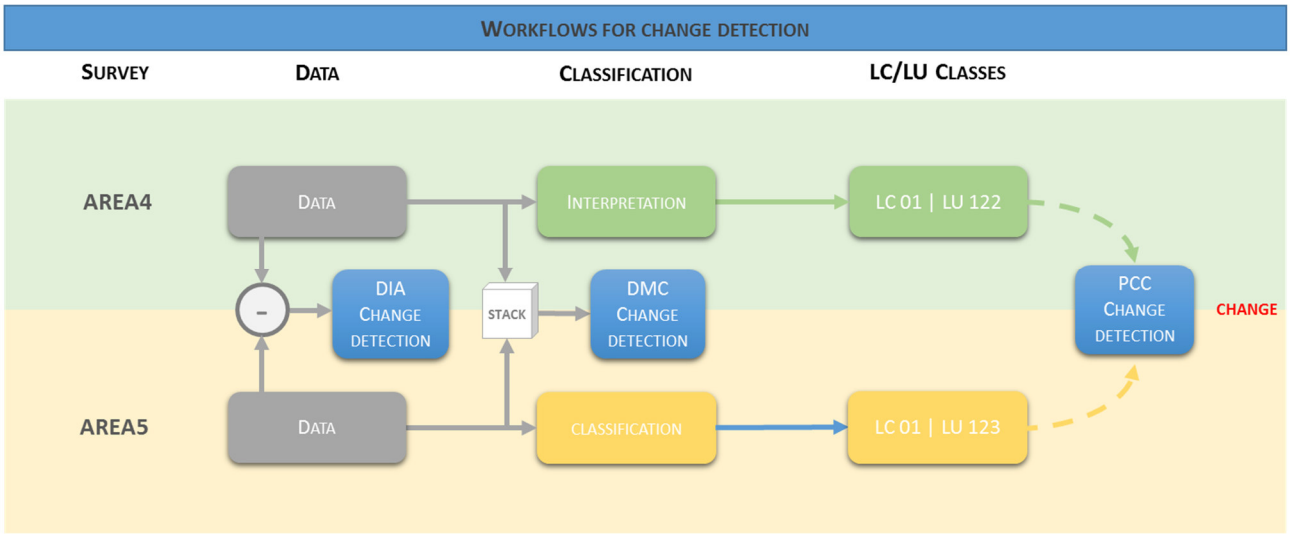


Figure 15 - Flowchart of the different approaches to perform change detection: DIA, DMC and PCC

General change detection approaches can be separated into three main categories:

- **DIA**: Difference Image Analysis: This approach subtracts two images descriptors (pixel by pixel, or deep features by deep features) at two different time instants to form a new image difference. As well the use of indices instead of spectral bands can provide better enhancement of the changes. For example the Vegetation index differencing is the difference between the Normalized Difference Vegetation Index (NDV I = (NIR - Red)/(NIR + Red)) at two time instants. This is the preferred solution in unsupervised change detection since all unchanged pixels collapse towards zero (Bruzzone and Serpico, 1997).
- **DMC**: Direct Multi-date Classification: This approach combines the multi-date images into a single one by stacking the spectral information of each pixel. This has the advantage of using all the information available for classifying accurately the different occurring changes. The advantage is that information on land-cover types of the unchanged regions is preserved. Detection of changes and identification of change types can be performed. However, this requires more complex classification models and sufficient training data covering all possible transitions.
- **PCC**: Post-classification comparison: This approach directly compares the classes obtained from a classification of each single date image separately to deduce the land-cover or land-use transitions that have taken place (e.g. vegetation to road). This is what BFS performs after having completed an Arealstatistik survey and comparing it to a previous one, therefore PCC does not bring added-value in comparison to a direct land-cover/land-use classification, they are actually similar. Moreover, it is prone to cumulation of classification errors.

In the future Arealstatistik approach which is explored in this study, the preceding survey results are known and change detection should be a first step to avoid interpreting non-changed points. In case of a change it is already very helpful to know that a change occurred without knowing the change type or class transition. Therefore, the DIA approach is most appropriate and DMC may be considered.

Picterra Sàrl 2017, all rights reserved

The change detection methods should be robust to the following elements:

- Illumination changes and shadows
- Non-permanent changes (cars, animals, temporary installations for repair, etc.)
- Data heterogeneity over time: format/specifications evolving through time
- Seasonal/intra-class changes (season difference between surveys, change of agricultural field type, etc.)

Illumination changes and shadows

Methods for compensating shadows have been proposed such as decomposing colors in intensity invariant components (Tsai 2006), pairing regions of shadows and non-shadows (Guo et al. 2013), and thresholding the hue to intensity ratio image (Storey et al. 2017). The general variations of illumination are usually taken into account at the feature extraction level with hand-crafted or learned features, or at the classification model level when using pixel or object-based features. An appropriate threshold handling illumination changes can be selected when using the Change Vector Analysis (CVA) approach (Bovolo et al. 2012), or the iteratively-reweighted multivariate alteration detection (IR-MAD), can be used since it is better in handling illumination differences (Nielsen 2007). The estimation of the surface reflectance, an intrinsic property of the scene, allows to be invariant to both sensor specificities and illumination changes (Nimisha et al. 2015). Even differences in view point can be handled by warping and co-registering the images using SURF keypoint descriptors and correlation between Local Binary Similarity Patterns (LBSP) allow a certain robustness to illumination changes (Javadi et al. 2016).

Non-permanent changes

The non-permanent changes due to a temporary occlusion of the scene (vehicle, left overs, temporary installations, etc.) are difficult to identify and to discard from the change detection. They will introduce a false detection which is not problematic but only time consuming since requiring a validation by the interpreter.

Data heterogeneity

The data heterogeneity can be well appreciated with analogic black & white photographs in the first two Arealstatistik surveys, then analogic colors in the third and all digital data in the current one. Methods or features can be robust to handle such types of differences. For example, the comparison of black & white and color imagery has been proposed based on GLCM texture descriptors (Erener et al. 2009). Differences in resolution can be handle by hand-crafted features robust to scale (e.g. SIFT, SURF), using objects derived from a common segmentation, or by fusion of the different data (Ferraris et al. 2016).

Seasonal/intra-class changes

The changes due to the difference in seasons (flowers, leaves colors, etc.) and the changes which occur but do not impact the land cover/use class of a scene will have to be learned by the model. For example, a tree can be moved and replanted 10m away, which will cause a change in the scene but might not impact the class type. The learning of the relevant changes can be done via a supervised, semi-supervised or active learning strategy. Unsupervised approaches have interesting methods for transforming the data in a way that highlights changed areas. Then a threshold on the magnitude of change can be applied, like with Change Vector Analysis (CVA) (Bovolo et al. 2012, Im et al. 2008) or Canonical Correlation Analysis (CCA) (Nielsen 2007). Alternatively, the difference between bi-temporal images can be represented in a feature space using kernel functions (Volpi et al. 2013). In these feature spaces, unsupervised methods such as k-means clustering or supervised methods such as SVM can be performed (de Morsier et al. 2013, de Morsier et al. 2015).

The features to be extracted for change detection range from the pixels to complex texture descriptors, passing by object-based features (derived from a segmentation). In the Arealstatistik context, the classes are related to a certain contextual area (in average 50x50m) around the sample point. Therefore, the change information has to consider the entire set of the individual pixels changes. There will be a trade-off between obtaining directly change information at the pixel-level which may be too sensitive and inducing a high number

Picterra Sàrl 2017, all rights reserved

of false detection, and a too coarse change information directly at the scene level with a risk of missing class changes.

Here are listed and described the different features to be considered for change detection:

- **Pixel-based features**: most of the change detection methods are built on top of the difference made between each image pixel (DIA) (Bovolo et al. 2012, Volpi et al. 2013, de Morsier et al. 2013).
- **Object-based features**: the object-based change detection methods require to perform a segmentation which fits to both images in time. Either two independent segmentations are performed and then overlapping segments are compared or a unique segmentation is performed, and the features derived from the segment used for detecting changes (Im et al. 2008, Masroor et al. 2013).
- **Hand-crafted features**: higher level features, as presented in section 5.1.2.1 as "hand-crafted" features, can be integrating in change detection methods. The HOG descriptors and gray level differences (Singh et al. 2014) spatially regularized via Markov-Random Fields have been used for change detection. Moreover, intensity-base, edge-based, and color-based features such as local binary pattern (LBP) or local edge pattern (LEP) were used for detection of changes corresponding to damaged roofs (Thomas et al. 2014).
- **Deep learning features**: with the challenging distinction of significant changes and irrelevant changes, more complex features such as supervised deep features extracted from CNNs can be the appropriate type of descriptors for detecting changes (El Amin et al. 2016, Chu et al. 2016). A Difference Image (DI) is computed from the extracted features instead of the pixels.

Benchmark datasets

In the change detection community, no standard dataset is shared across the researchers to benchmark different methods of change detection with aerial imagery. Here are two datasets which are closest to the Arealstatistik context of land cover / land use changes in very high-resolution imagery, however they have not been used in other publications than their authors:

- AICDDataset: 1000 pairs of 800x600 images and resolution 0.5m/pixel. Images are simulated using a realistic 3D engine with varying viewpoints, tilted view angles (20° off-nadir). http://www.computervisiononline.com/dataset/1105138664 (Bourdis et al. 2011)
- AirChange Benchmark: 13 image pairs of size 952x640 and resolution 1.5m/pixel. http://web.eee.sztaki.hu/remotesensing/benchmark.html (Benedek et al. 2009)

Picterra Sàrl 2017, all rights reserved

Approaches description and evaluation:

In this sub section, change detection approaches are described and evaluated:

Method	ID: C1	
Description	**Land cover/use change detection using aerial imagery and pixel-based features** Automatic change detection from aerial VHR data by creating a difference image and learning a classification model of change vs. non-change pixels.	
Application ranges	Change detection of land cover	
Features	Pixel features (spectral bands, band ratios and indices such as NDVI)	
Classification model	SVM, Maximum Likelihood	
Data and accuracies	Authors are evaluating changes at each pixel. Reported accuracies are on different datasets at lower resolution than for Arealstatistik setup.	
Authors/References	(Bovolo et al. 2012, Volpi et al. 2013, de Morsier et al. 2013)	
Development status (« Technology Readiness Level »)	**High readiness level:** - Ease of implementation of pixel-based features - Standard classifiers existing in open source libraries, toolboxes and softwares (PCI Geomatics Geomatica, ITT ENVI EX Change detection) http://www.pcigeomatics.com/pdf/geomatica/tutorials/Change_Detection.pdf http://www.harrisgeospatial.com/portals/0/pdfs/enviex/ENVIEX_ChangeDetection_Workflow_Slick_A4.pdf SOCET GXP ACD Simple/Relative Difference & Spectral Change detection	
License	Commercial license for PCI Geomatics, SOCET GXP, ENVI Harris	
		Score
Reduced interpretation	The change detection at the pixel level might lead to a too high false change detection rate, however it will have a very high potential for avoiding to miss changes and still reducing the number of interpretation.	3
Accuracy	It is not straight forward to assess the accuracy for evaluating change vs. non-change at the scene level, however the change type information per pixel could allow to deduce information at the scene level.	4
Robustness	Features are not specifically robust, however the classifier integrating all pixels of the scene could handle a certain class variability with a certain amount of training data.	3
Sustainability	Model can be retrained efficiently with new data (Linear SVM can handle online learning approaches). Cannot directly handle new data sources or type of features.	3
Adaptation cost	The features can be extracted from the aerial data and DSM pixels and added to the auxiliary data to perform a image feature difference map and provide it to a standard classifier. All elements are existing in standard toolboxes.	4

Picterra Sàrl 2017, all rights reserved

Method	ID: C2	
Description	**Land cover/use change detection using aerial imagery and object-based features** Automatic change detection from aerial VHR data by segmenting into object segments and learning a classification model of change vs. non-change segments.	
Application ranges	Change detection of land cover	
Features	Object features (segment average & std of spectral values, band ratios and indices such as NDVI, segment), if different segmentation: object overlap, object shape	
Classification model	SVM, Maximum Likelihood	
Data and accuracies	Authors are evaluating changes at each object. Reported accuracies are on different datasets at lower resolution than for the Arealstatistik setup.	
Authors/References	(Im et al. 2008, Masroor et al. 2013)	
Development status (« Technology Readiness Level »)	**Medium readiness level**: - Difficult choice of segmentation method and parameters adapted to the different change types. Comparison of objects or object-based features not standardized. - Standard classifiers existing in open source libraries, toolboxes and and softwares (PCI Geomatics Geomatica, ITT ENVI EX Change detection) http://www.pcigeomatics.com/pdf/geomatica/tutorials/Change_Detection.pdf http://www.harrisgeospatial.com/portals/0/pdfs/enviex/ENVIEX_ChangeDetection_Workflow_Slick_A4.pdf	
License	Commercial license for PCI Geomatics	
		Score
Reduced interpretation	The change detection at the object level might lead to a too high false change detection rate, however it will have a very high potential for reducing the number of interpretation and avoiding to miss changes.	3
Accuracy	It is not straight forward to assess the accuracy for evaluating change vs. non-change at the scene level, however the change type information per object could allow to deduce information at the scene level.	4
Robustness	Object-features are robust to small variations in illuminations and reflectance, but the definition of an adequate segmentation is difficult. The classifier integrating all object information of the scene could handle a certain class variability with a certain amount of training data.	3
Sustainability	Model can be retrained efficiently with new data (Linear SVM can handle online learning approaches). Cannot directly handle new data sources or type of features.	3
Adaptation cost	The features can be extracted from the aerial data and DSM segmented into objects and added to the auxiliary data to perform feature comparison and provide it to a standard classifier. All elements are existing in standard toolboxes but more processing elements are required due to the segmentation.	3

Picterra Sàrl 2017, all rights reserved

Method	ID: C3
Description	**Land cover/use change detection using aerial imagery and unsupervised[1] deep features** Automatic land cover/use change detection from aerial VHR data by exploiting CNNs to extract deep features from multi(bi)-temporal image data. Consequently, a Difference Image (DI) is created by processing and thresholding the obtained features (El Amin et al. 2016, Chu et al. 2016). Using CNNs to generate DI enables to avoid the radiometric correction procedures for change detection, enhances the difference in change areas remarkably, and suppress the noise effectively compared with traditional method of DI creation methods.
Application ranges	Change detection of land cover/use/specifications (up to 45 classes for NWPU-RESISC45 dataset)
Features	Extracted features from pre-trained CNNs: AlexNet, GoogleLeNet, VGG-16, OverFeat, CaffeNet; or Extracted features form newly created and trained CNNs.
Classification model	Convolutional Neural Network (CNN)
Data and accuracies	Authors using CNNs to detect changes on aerial images claim to improve the precision of change detection results by 30%. The Kappa coefficient reaches 90% in average as reported by (El Amin et al. 2016, Chu et al. 2016)
Authors/References	(El Amin et al. 2016) (Chu et al. 2016)
Development status (« Technology Readiness Level »)	**Low readiness level**: - few experimental tests properly performed in terms of change detection on aerial images and publications not well documented - but many pre-trained models available in open source libraries and cloud computing libraries, e.g. Microsoft Azure Cognitive Toolkit (AlexNet, GoogleLeNet, ResNet, VGG-16).
License	-

		Score
Reduced interpretation	The change detection at the scene level might lead to a difficult to control missed change detection rate, however the potential for reducing the number of interpretation is very high.	4
Accuracy	There are no reliable experiments to assess the accuracy of change vs. non-change of classes at the scene level. Only from the high classification results can be extrapolated a potential high change detection rate.	4
Robustness	Features are robust to a large number of effects (illuminations, scale, orientations, organisations, etc.), however it will need a significant amount of data to adapt to the class specific changes and learn appropriate semantic level of features.	4
Sustainability	Model can be retrained with new data (data and classification error backpropagated through the network). If the model is fine-tuned the classifier needs to be trained from scratch afterwards since features produced by the network may have changed after retraining. If using CNN last layers as classifier, retraining is possible.	3
Adaptation cost	The features can be extracted from the aerial data and DSM using pre-trained CNNs and added to the auxiliary data to perform feature difference map and provide it to a standard classifier. All elements are existing in standard toolboxes.	4

[1] Unsupervised with pre-trainedpre-trained CNN. If the pre-trainedpre-trained CNN is then fine-tuned, the feature extraction becomes supervised.

Picterra Sàrl 2017, all rights reserved

5.4.Complementary data sources

In this section is presented data sources which could complement the standard dataset used in Arealstatistik (see section 0) for the classification of land cover/use classes or the detection of changes.

5.4.1. Optical Earth Observation satellite data

Free data are available such as the recently launched satellite constellation Sentinel-2 which systematically acquires data over the global land and coastal areas between 56° South to 84° North, including all larger islands. Its images are provided as so-called Level-1C products, which are corrected to top-of-atmosphere reflectance in a cartographic geometry and can be transformed to bottom-of-atmosphere reflectance by removing atmospheric absorption effects. The sensor is designed and very valuable for all application involving the monitoring of state and changes of terrestrial features, e.g. in the domain of agriculture, urban changes or forestry. Major characteristics are the large number of spectral bands (13) bands covering the visible, near-infrared and shortwave infrared part of the electromagnetic spectrum, a high spatial resolution (bands at 10m, 20m and 60m), a very wide swath of 290 km (width on the earth surface scanned at each instant) and the frequent revisit rate (5 days at the equator, therefore 2 to 3 days above Switzerland).

The satellite Landsat-8 is the latest member of a satellite family which has a track record of 45 years of continuous operation. Especially the long archive makes this system unique for certain applications requiring historic data records. The field of application is similar to Sentinel-2, even though the spatial resolution (30m), swath (185 km) and especially revisit time (16 days) are less favorable.

In terms of sustainability, there exists a strong continuity across the satellite systems in order to enable long term monitoring of our Earth Planet. The Landsat series of satellites are running un-discontinuously since more than 40 years. The European Sentinels have the same intention in terms of monitoring. In addition, there are many new satellite constellations which are to be launched in the next years and the trend seems to be the beginning of a new era of small satellite constellations. This will ensure an access to very high-resolution imagery at high revisit time in the next decade.

In the context of the Arealstatistik, satellite data can have different roles:

- Replacing aerial data: only for VHR satellite data which could provide sufficient resolution (<1m)
- Complementing aerial data: satellite data can provide time series which allows to better characterize certain surfaces (e.g. agriculture) or distinguish temporary vs. permanent changes.

Picterra Sàrl 2017, all rights reserved

-

Table 1 summarizes the relevant characteristics of relevant optical satellite systems.

Table 1 - Selection of optical EO satellite system (non-exhaustive list).

Satellite	Launch	Spatial resolution	# of spectral bands	Spectral domain*	Swath	Revisit time	Price
Landsat-8 (OLI)	2013	15 m 30 m 30 m	1 5 3	Pan B, G, R, NIR SWIR	185 km	16 days	free
Terra (ASTER)	1999	15 m 30 m	4 6	B, G, R, NIR SWIR	60 km	16 days	
Sentinel-2A/B (constellation: 2)	2015/ 2017	10 m 20 m 60 m	4 6 3	B, G, R, NIR NIR, SWIR B, NIR, SWIR	290 km	5 days	
Rapideye (constellation: 5)	2003	6.5 m	5	B, G, R, RE, NIR	77 km	1 day	0.95 to 1.3 €/km^2
SPOT-6/SPOT-7 (constellation: 2)	2012/ 2014	1.5 m 6 m	1 4	Pan B, G, R, NIR	60 km	1 day	
Planet (constellation: 149)	2013/ 2017	3.0/3.7 m	4	B, G, R, NIR	24.6 km	1 day**	~0.10 €/km^2
SkySat* ** (constellation: 7)	2013/ 2016	0.9 m 2.0 m	1 4 (Video option)	Pan B, G, R, NIR	8 km	< 1 day	
OptiSAR# (constellation: 8)	2021	1.5x2.5 m	4 (Optical) (not confirmed) 1 (SAR)	B, G, R, NIR (not confirmed) X-Band	5 km	1 day	? not yet available
Pleiades 1A/B (constellation: 2)	2011	0.5 m 2 m	1 4	Pan B, G, R, NIR	20km	1 day	17 €/km^2

* Pan - panchromatic, B - blue, G - green, R - red, RE - red edge, NIR - near infrared, SWIR - shortwave infrared
** Current revisit rate is 2-7 days, from end of 2017 on daily, once all satellites are operational.
*** Commercially operational from end of 2017 on; further launches scheduled for 2017.
# 8 SAR-satellites with X- and L-Band sensors, in tandem mode with 8 optical VHR satellites.

The price for images from the Planet constellation every 14 days for a duration of five years is estimated, based on prices quotes made over different areas and periods, at: ~0.10 €/km^2 (26 x 5 x 41300 km^2), a total of around 540'000 € to monitor Switzerland over this period.

Further satellite constellations which will provide daily revisits:

- Hera Systems: 48 satellites, ultra-high spatial resolution (< 1m)
- Satellogic: > 100 satellites, very-high spatial resolution (1m), also 30m hyperspectral
- Black Sky Global: 60 satellites, very-high spatial resolution (1m)

And additional constellations will most likely be launched in the next three to five years.

Picterra Sàrl 2017, all rights reserved

5.4.2. High Altitude Platforms data

At the same time than the rise of satellite constellations, the high-altitude platforms are under development and may directly compete with the latter. The idea is to place at high altitude (e.g. ~20km which is higher than civil flights but lower than EO satellites) platforms which would cover a certain region continuously. Such platform can provide high capacity telecommunications but also Earth Observation services in a continuous manner and at very high resolution. Moreover, the costs associated will be lower than with satellites, since the expensive rocket launch would be avoided. None of these platforms have yet committed to a fixed design of Earth Observation sensor but are definitely considering such payloads regarding surveillance and general monitoring applications. Different types of platforms are currently under development:

- Thales Aliena Space "Stratobus", an unmanned zeppelin, see Figure 16.
- Stratxx, an unmanned zeppelin.
- Airbus "Zephyr", a solar powered unmanned aerial vehicle, see Figure 16.
- OpenStratosphere, a solar powered unmanned aerial vehicle.
- Google "Loon" ballon project, however less likely to carry a EO payload.

Thales Aliena Space | Airbus

Figure 16 - Example of some High-Altitude Platforms under development

5.4.3. Crowdsourced Geospatial data

Nowadays, the usage of geospatial data is fully standardized and democratized with tools such as Google maps accessible from any smartphone, providing the current geolocation of the device via the embedded GPS. This democratization of the access to geospatial data had also its counter-part with more and more geospatial content generated by the users: geotagged pictures uploaded on social networks, online provision of the current location for adapted services (e.g. uber, shop recommendations, etc.) or mapping of known 2D environment on a platform such as OpenStreetMap.

Geospatial layers: The major source of geospatial information collected via crowdsourcing, user-generated content, also called Volunteered geographic information (VGI), are the OpenStreetMap (OSM) layers. The use of layers from OSM to perform a land use classification has been studied (Estima et al. 2013), however the minimum mapping unit is at a lower level than Arealstatistik (~50cm). This massive source of reference data can be exploited to initialize and pre-train CNNs. This large but possibly inaccurate reference data allows obtaining network layers well adapted to the classification task and only requiring a fine-tuning using a small amount of accurately labeled data (Maggiori et al. 2017).

Street-level images and Social Media: the social media (e.g. blogs, wiki, Instagram, Flickr, Facebook) are currently gathering massive amounts of geo-tagged images. Among the largest, Flickr represents 122 millions of users and a total of 10 billion images shared. A large collection of georeferenced panoramas and images was present on the Panoramio platform (94 millions), now part of Google Maps, on 360cities platform, or on Mapillary (100 millions). The latter is offering an immersive street-level view of the world but in a crowdsourced manner compared to Google Street View. Computer vision techniques are exploited to detect objects and understand text and signs. Specific to Switzerland, the VideoStreetView initiative is offering a combination of GIS map and video at the street-level. With respect to land use/cover classification, different

Picterra Sàrl 2017, all rights reserved

attempts of exploited such data alone or together with traditional aerial images have been performed (Lefèvre et al. 2017): Flickr pictures were used to predict the presence or absence of snow, individual trees could be detected from street-level and aerial images, roads observed in both street view and aerial images could be detected and change detection by coupling ground-level panoramas and aerial imagery (Ghouaiel et al. 2016), see Figure 17.

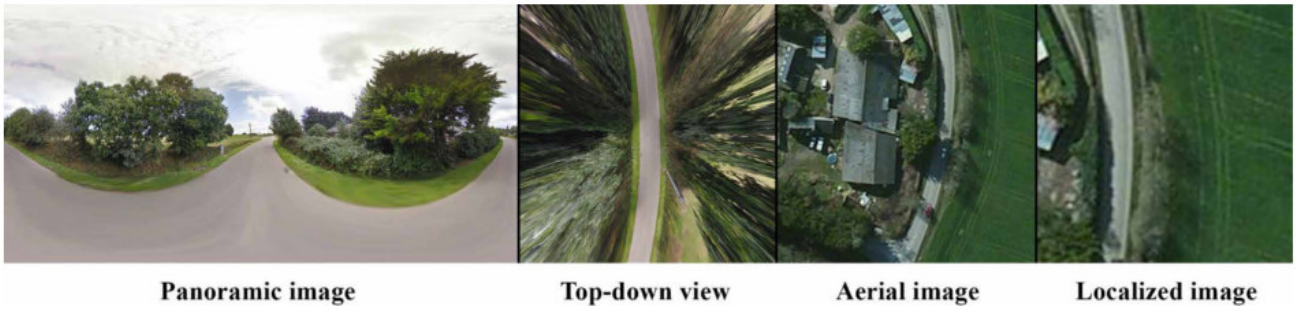


Figure 17 - Example of ground-level panorama, top-down view reconstructed, corresponding aerial image and its corresponding part localized in that image (from Lefèvre et al. 2017).

5.5.References

Datasets

UC-MERCED: Yang, Y., & 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*, 270. http://doi.org/10.1145/1869790.1869829

http://vision.ucmerced.edu/datasets/landuse.html

WHU-RS19: Sheng, G., Yang, W., Xu, T., & Sun, H. (2012). High-resolution satellite scene classification using a sparse coding based multiple feature combination. *International Journal of Remote Sensing*, *33*(8), 2395–2412. http://doi.org/10.1080/01431161.2011.608740

SIRI-WHU: Zhao, B., Zhong, Y., Xia, G. S., & Zhang, L. (2016). Dirichlet-Derived Multiple Topic Scene Classification Model for High Spatial Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(4), 2108–2123. http://doi.org/10.1109/TGRS.2015.2496185.

RSSN7: Zou, Q., Ni, L., Zhang, T., & Wang, Q. (2015). Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geoscience and Remote Sensing Letters*, *12*(11), 2321–2325. http://doi.org/10.1109/LGRS.2015.2475299

RSC11: Zhao, L., Tang, P., & Huo, L. (2016b). Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. *Journal of Applied Remote Sensing*, *10*(3), 35004. http://doi.org/10.1117/1.JRS.10.035004

AID: Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., … Lu, X. (2017). AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3965–3981. http://doi.org/10.1109/TGRS.2017.2685945

http://www.lmars.whu.edu.cn/xia/AID-project.html

NWPU-RESISC45: Cheng, G., Han, J., & Lu, X. (2017). Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proceedings of the IEEE*, 1–19. http://doi.org/10.1109/JPROC.2017.2675998

VAIHINGEN: Gerke, M. (2015). Use of the Stair Vision Library within the ISPRS 2D Semantic Labeling Benchmark (Vaihingen). http://doi.org/10.13140/2.1.5015.9683

POTSDAM: http://www2.isprs.org/commissions/comm3/wg4/2d-sem-labelpotsdam.html

Picterra Sàrl 2017, all rights reserved

Aptoula, E. (2014). Remote sensing image retrieval with global morphological texture descriptors. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(5), 3023–3034. http://doi.org/10.1109/TGRS.2013.2268736

Bailly, J. S., Arnaud, M., & Puech, C. (2007). Boosting: a classification method for remote sensing. *International Journal of Remote Sensing*, *28*(May), 1687–1710. http://doi.org/10.1080/01431160500469985

Camps-Valls, G., Tuia, D., Gómez-Chova, L., & Jiménez, S. (2011). *Remote sensing image processing*. *Synthesis Lectures on*. http://doi.org/S00392ED1V01Y201107IVM012

Camps-Valls, G., & Bruzzone, L. (2009). *Kernel Methods for Remote Sensing Data Analysis*. (G. Camps-Valls & L. Bruzzone, Eds.)*Kernel Methods for Remote Sensing Data Analysis*. Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/9780470748992

Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv Preprint arXiv:1508.00092*, 1–11. Retrieved from http://arxiv.org/abs/1508.00092

Cheng, G., Han, J., Zhou, P., & Guo, L. (2014). Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing*, *98*, 119–132. http://doi.org/10.1016/j.isprsjprs.2014.10.002

Cheriyadat, A. M. (2014). Unsupervised Feature Learning for Aerial Scene Classification. *Geoscience and Remote Sensing, IEEE Transactions on*, *52*(1), 439–451. http://doi.org/10.1109/TGRS.2013.2241444

Demir, B., Minello, L., & Bruzzone, L. (2012). A cost-sensitive active learning technique for the definition of effective training sets for supervised classifiers. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 1781–1784. http://doi.org/10.1109/IGARSS.2012.6351169

Eitel, A., Springenberg, J. T., Spinello, L., Riedmiller, M., & Burgard, W. (2015). Multimodal deep learning for robust RGB-D object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vol. 2015–Decem, pp. 681–687). IEEE. http://doi.org/10.1109/IROS.2015.7353446

Gong Cheng, Junwei Han, Peicheng Zhou, & Lei Guo. (2014). Scalable multi-class geospatial object detection in high-spatial-resolution remote sensing images. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 2479–2482). IEEE. http://doi.org/10.1109/IGARSS.2014.6946975

Hu, F., Xia, G., & Zhang, L. (2016). Deep Sparse Representations for Land-Use Scene Classification in Remote Sensing Images, 192–197.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., & Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, *9*(1), 95. http://doi.org/10.3390/rs9010095

Izquierdo-Verdiguier, E., Gomez-Chova, L., Bruzzone, L., & Camps-Valls, G. (2014). Semisupervised Kernel Feature Extraction for Remote Sensing Image Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(9), 5567–5578. http://doi.org/10.1109/TGRS.2013.2290372

Lakshminarayanan, B., Roy, D. M., & Teh, Y. W. (2014). Mondrian forests: Efficient online random forests. In *Advances in neural information processing systems* (pp. 3140-3148). https://arxiv.org/abs/1406.2673

Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop: Challenges in Representation Learning*, 1–6. Retrieved from https://www.kaggle.com/blobs/download/forum-message-attachment-files/746/pseudo_label_final.pdf

Picterra Sàrl 2017, all rights reserved

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2016). Fully convolutional neural networks for remote sensing image classification. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 5071–5074). IEEE. http://doi.org/10.1109/IGARSS.2016.7730322

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 645–657. http://doi.org/10.1109/TGRS.2016.2612821

Maulik, U., & Chakraborty, D. (2017). Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques. *IEEE Geoscience and Remote Sensing Magazine*, *5*(1), 33–52. http://doi.org/10.1109/MGRS.2016.2641240

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. http://doi.org/10.1016/j.isprsjprs.2010.11.001

Muñoz-Marí, J., Tuia, D., & Camps-Valls, G. (2012). Semisupervised classification of remote sensing images with active queries. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(10), 3751–3763. http://doi.org/10.1109/TGRS.2012.2185504

Newsam, S., Wang, L., Bhagavathy, S., & Manjunath, B. S. (2004). Using texture to analyze and manage large collections of remote sensed image and video data. *Applied Optics*, *43*(2), 210–217. http://doi.org/10.1364/AO.43.000210

Penatti, O. A. B., Nogueira, K., & Dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2015–Octob*, 44–51. http://doi.org/10.1109/CVPRW.2015.7301382

Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 512–519. http://doi.org/10.1109/CVPRW.2014.131

Romero, A., Gatta, C., Camps-valls, G., & Member, S. (2016). Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(3), 1349–1362. http://doi.org/10.1109/TGRS.2015.2478379.

Rougier, S., Puissant, A., Stumpf, A., & Lachiche, N. (2016). Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images. *International Journal of Applied Earth Observation and Geoinformation*, *51*, 60–73. http://doi.org/10.1016/j.jag.2016.04.005

Santos, J. A., Penatti, A. B., & Torres, R. S. (2010). Evaluating the Potential of Texture and Color Descriptors for Remote Sensing Image Retrieval and Classification. *Proceedings of the International Conference on Computer Vision Theory and Applications*, 203–208. http://doi.org/10.5220/0002843402030208

Sherrah, J. (2016). Fully Convolutional Networks for Dense Semantic Labeling of High-Resolution Aerial Imagery. *arXiv*, 1–22. Retrieved from http://arxiv.org/abs/1606.02585

Stumpf, A., Lachiche, N., Kerle, N., Malet, J. P., & Puissant, A. (2012). Adaptive spatial sampling with active random forest for object-oriented landslide mapping. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 87–90). http://doi.org/10.1109/IGARSS.2012.6351630

Stumpf, A., Lachiche, N., Malet, J. P., Kerle, N., & Puissant, A. (2014). Active Learning in the Spatial Domain for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(5), 2492–2507. http://doi.org/10.1109/TGRS.2013.2262052

Picterra Sàrl 2017, all rights reserved

Sventnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a tool for classification and regression in compound classification and {QSAR} modeling. *Journal of Chemical Information and Computer Science*, *43*(6), 1947–1958. http://doi.org/10.1016/j.rse.2008.02.011

Tuia, D., & Munoz-Mari, J. (2012). Putting the user into the active learning loop: Towards realistic but efficient photointerpretation. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 75–78). IEEE. http://doi.org/10.1109/IGARSS.2012.6351633

Tuia, D., Volpi, M., Copa, L., Kanevski, M., & Munoz-Mari, J. (2011). A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Signal Processing*, *5*(3), 606–617. http://doi.org/10.1109/JSTSP.2011.2139193

Valentini, G., & Dietterich, T. G. (2004). Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods. *Journal of Machine Learning Research*, *5*, 725–775. http://doi.org/10.1007/3-540-45428-4_22

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. New York, NY: Springer New York. http://doi.org/10.1007/978-1-4757-3264-1

Vargas, J. E., Saito, P. T. M., Falcao, A. X., Rezende, P. J. De, & Santos, J. a. Dos. (2014). Superpixel-Based Interactive Classification of Very High Resolution Images. *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, 173–179. http://doi.org/10.1109/SIBGRAPI.2014.49

Volpi, M., & Tuia, D. (2017). Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 881–893. http://doi.org/10.1109/TGRS.2016.2616585

Weng, Q., Mao, Z., Lin, J., & Guo, W. (2017). Land-Use Classification via Extreme Learning Classifier Based on Deep Convolutional Features. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 704–708. http://doi.org/10.1109/LGRS.2017.2672643

Xu, S., Fang, T., Li, D., & Wang, S. (2010). Object classification of aerial images with bag-of-visual words. *IEEE Geoscience and Remote Sensing Letters*, *7*(2), 366–370. http://doi.org/10.1109/LGRS.2009.2035644

Yang, Y., & 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*, 270. http://doi.org/10.1145/1869790.1869829

Zhang, F., Du, B., & Zhang, L. (2015). Saliency-Guided Unsupervised Feature Learning for Scene Classification. *Geoscience and Remote Sensing, IEEE Transactions on*, *53*(4), 2175–2184. http://doi.org/10.1109/TGRS.2014.2357078

Zheng, X., Sun, X., Fu, K., & Wang, H. (2013). Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint. *IEEE Geoscience and Remote Sensing Letters*, *10*(4), 652–656. http://doi.org/10.1109/LGRS.2012.2216499

Change detection methods

Alhichri, H., Bazi, Y., Alajlan, N., & Ahamad, S. M. (2012). Interactive change detection techniques in multitemporal multispectral remote sensing images. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 6173–6176). http://doi.org/10.1109/IGARSS.2012.6352666

Benedek, C., & Szirányi, T. (2009). Change detection in optical aerial images by a multilayer conditional mixed Markov model. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(10), 3416–3430. http://doi.org/10.1109/TGRS.2009.2022633

Picterra Sàrl 2017, all rights reserved

Bourdis, N., Marraud, D., & Sahbi, H. (2011). Constrained optical flow for aerial image change detection. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4176–4179. http://doi.org/10.1109/IGARSS.2011.6050150

Bovolo, F., Marchesi, S., & Bruzzone, L. (2012). A framework for automatic and unsupervised detection of multiple changes in Multitemporal Images. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(6), 2196–2212. http://doi.org/10.1109/TGRS.2011.2171493

Bruzzone, L., & Serpico, S. B. (1997). Detection of changes in remotely-sensed images by the selective use of multi-spectral information. *International Journal of Remote Sensing*, *18*(18), 3883–3888. http://doi.org/10.1080/014311697216702

Chu, Y., Cao, G., & Hayat, H. (2016). Change Detection of Remote Sensing Image Based on Deep Neural Networks. In *Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)* (Vol. 133, pp. 262–267). Paris, France: Atlantis Press. http://doi.org/10.2991/aiie-16.2016.61

De Morsier, F., Tuia, D., Borgeaud, M., Gass, V., & Thiran, J. P. (2015). Cluster validity measure and merging system for hierarchical clustering considering outliers. *Pattern Recognition*, *48*(4), 1474–1485. http://doi.org/10.1016/j.patcog.2014.10.003

De Morsier, F., Tuia, D., Borgeaud, M., Gass, V., & Thiran, J. P. (2013). Semi-supervised novelty detection using SVM entire solution path. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 1939–1950. http://doi.org/10.1109/TGRS.2012.2236683

Erener, A., & Technologies, G. I. (2009). A methodology for land use change detection of high resolution pan images based on texture analysis. *Italian Journal of Remote Sensing*, *41*(2), 47–59. http://doi.org/10.5721/ItJRS20094124

Ferraris, V., Dobigeon, N., Wei, Q., & Chabert, M. (2016). Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: a Fusion-Based Approach, 1–23. Retrieved from http://arxiv.org/abs/1609.06074

Guo, R., Dai, Q., & Hoiem, D. (2013). Paired Regions for Shadow Detection and Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2956–2967. http://doi.org/10.1109/TPAMI.2012.214

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 91–106. http://doi.org/10.1016/j.isprsjprs.2013.03.006

Im, J., Jensen, J. R. J., & Tullis, J. A. J. (2008). Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, *29*(September), 423, 399. http://doi.org/10.1080/01431160601075582

Mohammed El Amin, A., Liu, Q., & Wang, Y. (2016). Convolutional neural network features based change detection in satellite images. *First International Workshop on Pattern Recognition*, *10011*(figure 1), 100110W. http://doi.org/10.1117/12.2243798

Nielsen, A. A. (2007). The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transactions on Image Processing*, *16*(2), 463–478. http://doi.org/10.1109/TIP.2006.888195

Nimisha, T. M., Rajagopalan, A. N., & Aravind, R. (2015). Seamless change detection and mosaicing for aerial imagery. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2015–Octob*, 1–8. http://doi.org/10.1109/CVPRW.2015.7301393

Picterra Sàrl 2017, all rights reserved

Singh, P., Kato, Z., & Zerubia, J. (2014). A multilayer markovian model for change detection in aerial image pairs with large time differences. *Proceedings - International Conference on Pattern Recognition*, *1*, 924–929. http://doi.org/10.1109/ICPR.2014.169

Storey, E. A., Stow, D. A., Coulter, L. L., & Chen, C. (2017). Detecting shadows in multi-temporal aerial imagery to support near-real-time change detection. *GIScience & Remote Sensing*, *0*(0), 1–18. http://doi.org/10.1080/15481603.2017.1279729

Thomas, J., Kareem, A., & Bowyer, K. W. (2014). Automated poststorm damage classification of low-rise building roofing systems using high-resolution aerial imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(7), 3851–3861. http://doi.org/10.1109/TGRS.2013.2277092

Tsai, V. J. D. (2006). A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1661–1671. http://doi.org/10.1109/TGRS.2006.869980

Volpi, M., De Morsier, F., Camps-Valls, G., Kanevski, M., & Tuia, D. (2013). Multi-sensor change detection based on nonlinear canonical correlations. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 1944–1947. http://doi.org/10.1109/IGARSS.2013.6723187

Complementary data

Chen, C., Du, Z., Zhu, D., Zhang, C., & Yang, J. (2016). Land use classification in construction areas based on volunteered geographic information. *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, (17), 1–4. http://doi.org/10.1109/Agro-Geoinformatics.2016.7577633

Estima, J., & Painho, M. (2013). Exploratory analysis of OpenStreetMap for land use classification. *In: Geocrowd '13*, 39–46. http://doi.org/10.1145/2534732.2534734

Ghouaiel, N., & Lefèvre, S. (2016). Coupling ground-level panoramas and aerial imagery for change detection. *Geo-Spatial Information Science*, *19*(3), 222–232. http://doi.org/10.1080/10095020.2016.1244998

Lefevre, S., Tuia, D., Wegner, J. D., Produit, T., & Nassar, A. S. (2017). Toward Seamless Multiview Scene Analysis From Satellite to Street Level. *Proceedings of the IEEE*. http://doi.org/10.1109/JPROC.2017.2684300

Picterra Sàrl 2017, all rights reserved

6. Comparative assessment

This section presents the comparative assessment of the methods reviewed and summarizes the outcomes of the state-of-the-art review regarding the potential for the automatic classification and change detection of Arealstatistik land use/cover classes.

6.1. Comparative assessment of the methods reviewed

The reference to each method is done by a unique ID, which is also found in the section 5.1.3.4. The comparison is performed based on the six following criteria: Reduced interpretation, Accuracy, Robustness, Adaptation, Cost, Sustainability, which are generally described in Section 4 and specifically for each method in Section 5.1.3.4. In Table 2 below, the three highest ranked methods are in Bold.

Table 2 - Comparative assessment of methods

Methods		Evaluation criteria					score
ID	Description	Reduced	Accuracy	Robustness	Sustainability	Adaptation Cost	
A1	Land cover/use classification using aerial imagery and handcrafted features	2	2	3	3	4	14
A2	Land cover/use classification using aerial imagery and unsupervised features	2	3	3	1	3	12
A3	**Land cover/use classification using aerial imagery and supervised deep features**	**4**	**4**	**4**	**3**	**4**	**19**
A4	Dense land cover/use classification using aerial imagery and supervised deep features	3	3	3	3	2	14
B1	Land cover classification using satellite imagery and Random Forest	3	3	3	4	3	16
B2	Land cover classification using satellite imagery and Active Learning	2	2	4	4	3	15
C1	**Land cover/use change detection using aerial imagery and pixel-based features**	**3**	**4**	**3**	**3**	**4**	**17**
C2	Land cover/use change detection using aerial imagery and object-based features	3	4	3	3	3	16
C3	**Land cover/use change detection using aerial imagery and unsupervised deep features**	**4**	**4**	**3**	**3**	**4**	**18**

The method A3 “Land cover/use classification using aerial imagery and supervised deep features” is outstanding the other classification methods and provides a good balance between the five criteria. The second most promising approach is the change detection approach C3 “Land cover/use change detection using aerial imagery and unsupervised deep features”. It has a high potential for change detection, however the missed change rate is difficult to estimate from the literature review and will require tests. Then follows the more traditional change detection approach based on pixel features, which can allow to control the missed detection rate and still reduce the interpretation.

Picterra Sàrl 2017, all rights reserved

6.2.Comparative assessment in a Kiviat diagram

The compared methods are visualized in a Kiviat diagram, showing well the strengths and weaknesses of the different methods.

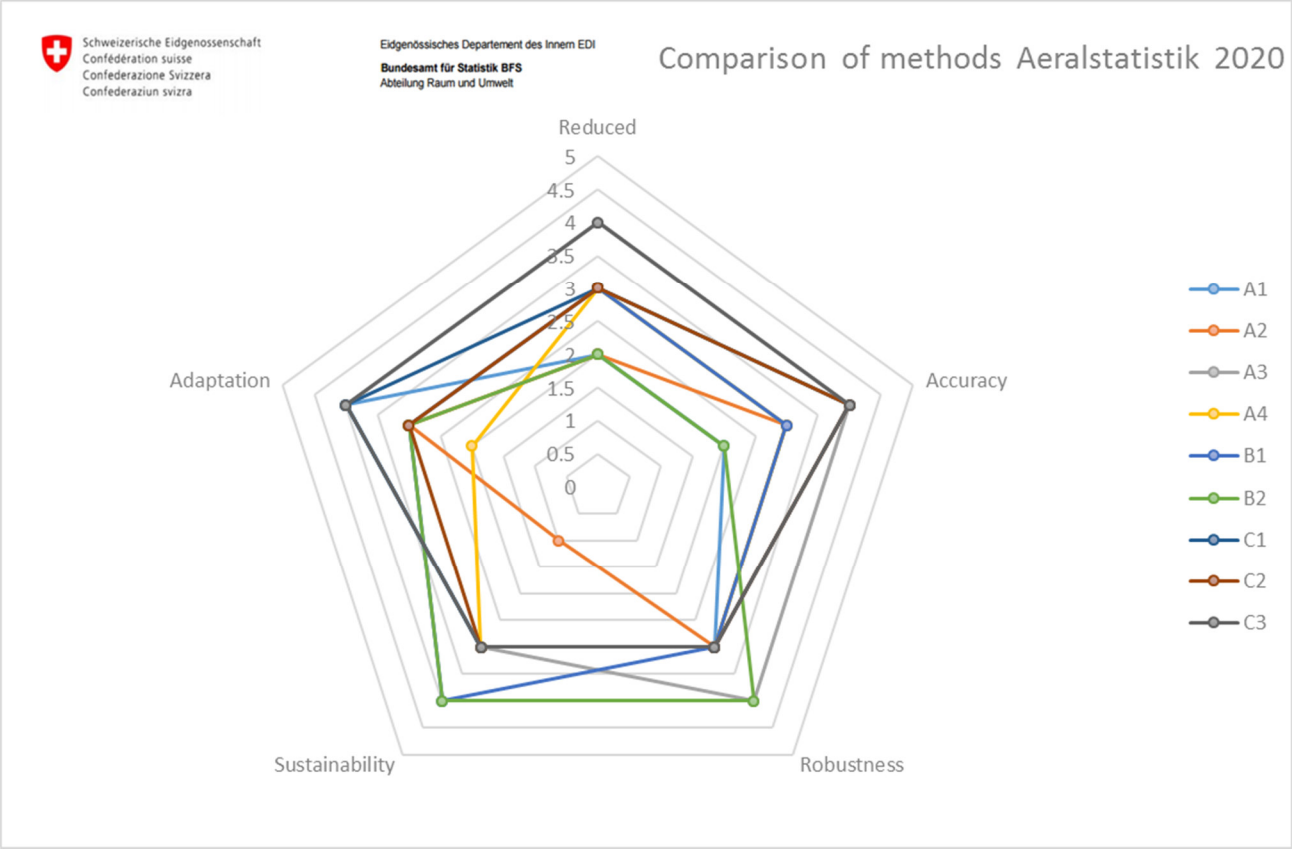
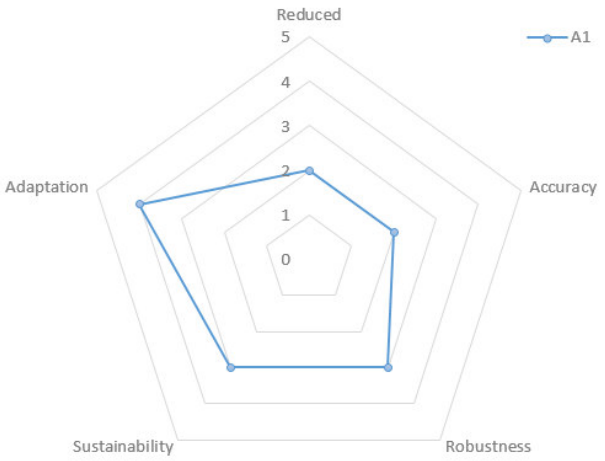
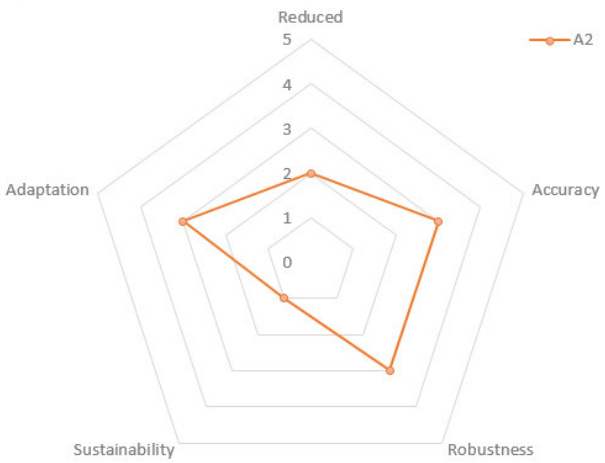


Figure 18 - Kiviat diagram as comparative assessment of methods

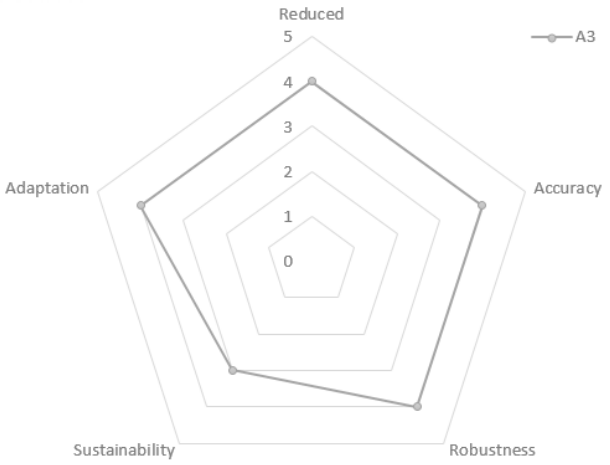
The Kiviat diagram gives an indication of which criterion is favored by certain methods. Very different methods could be investigated regarding a potential combination in order to obtain a high score in all criteria. For example, the approach B2 that is based on Active Learning seems complementary in terms of criteria scores to the first ranked A3 approach.

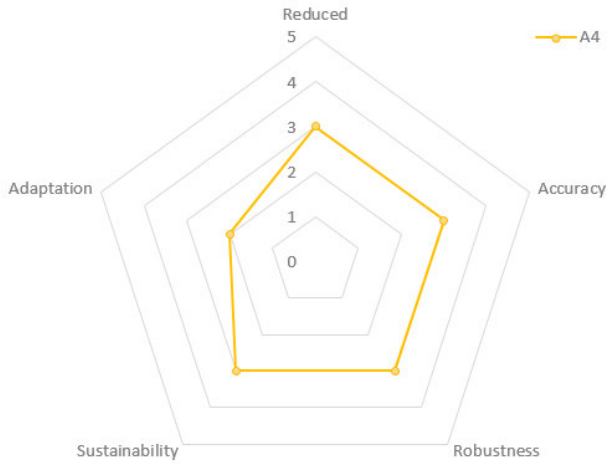
Here below the individual Kiviat diagrams:

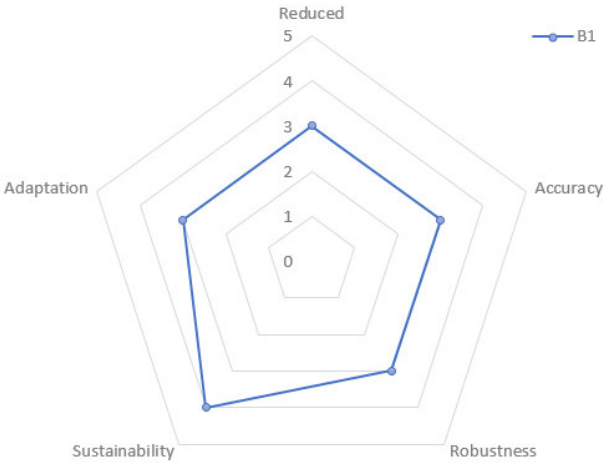


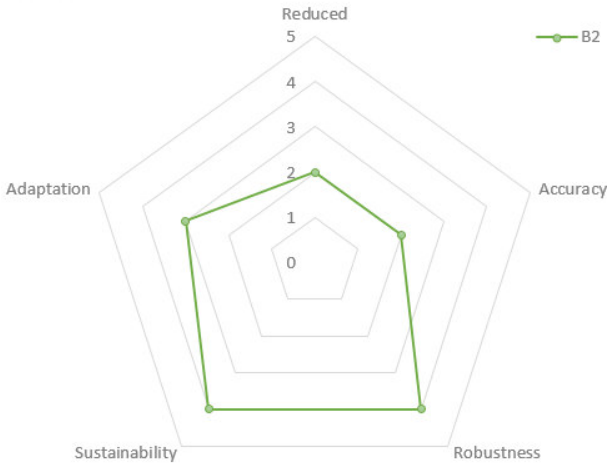


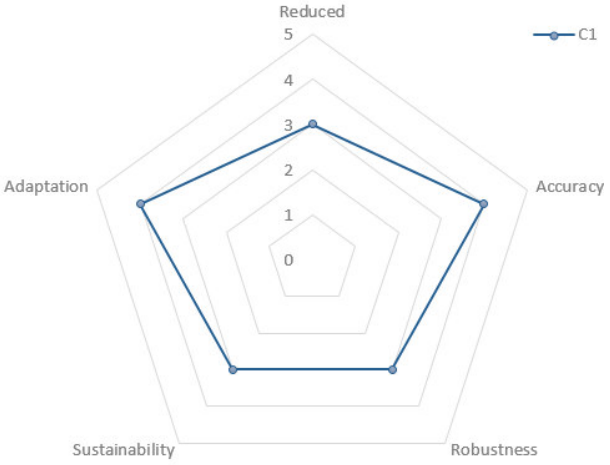
Picterra Sàrl 2017, all rights reserved

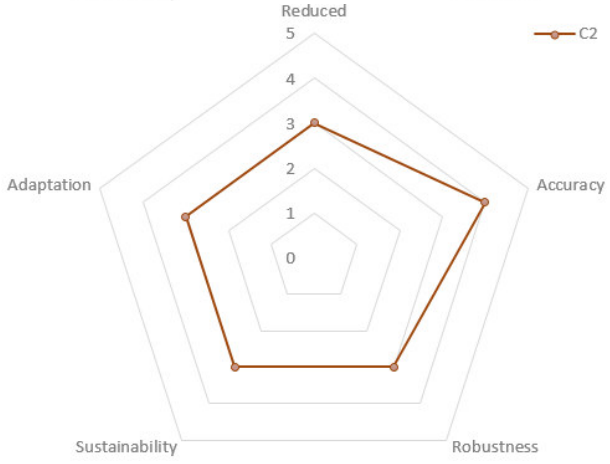


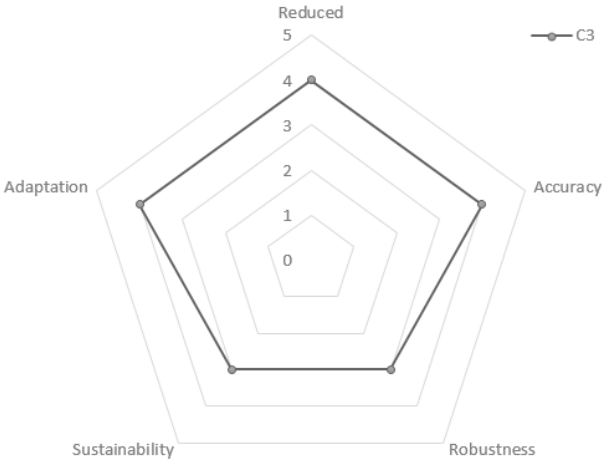












Picterra Sàrl 2017, all rights reserved

7. Test experiments

This section presents the tests performed in order to assess the potential of certain reviewed methods using the aerial data from Swisstopo and BFS Arealstatistik classes.

7.1. Introduction

The BFS Arealstatistik 2020 feasibility study is investigating, after the state-of-the-art review, certain tests with BFS Arealstatistik data in order to further assess the potential of approaches for automatic land use/cover classification and change detection. The following tests are performed:

- Classification: The quantitative assessment of approach A3 “Land cover/use classification using aerial imagery and supervised deep features” is performed:
  - Approach A3: Pre-trained CNNs from ImageNet (e.g. ResNet, VGG-16, etc.) to extract features
- Change detection: The quantitative assessment of approach C1 and C3 are performed:
  - Approach C1: Pixel-based change detection
  - Approach C3: Deep feature-based change detection

The following measures are reported:

- Optimal algorithm parameters and parameters sensitivity.
- Computational costs.
- Accuracy measures from stratified k-fold cross-validation, are reported per land cover and land use classes.
- Accuracy measures for top-3 classification results: assess the presence of the correct class within the TOP-3 predictions.
- ROC curves for best performing classes and Area under the ROC curves for change detection.

Let us define the following quantities:

- True Positives (TP): the number of correct classification or detection of changes.
- True Negatives (TN): the number of correct detection of non-changes.
- False Positives (FP): the number of samples wrongly classified as the class of interest or the wrongly detected changes.
- False Negatives (FN): the number of samples from the class of interest wrongly classified as another class or the missed changes.

The accuracy measures reported are defined as:

Accuracy measure	Classification	Change detection	Description
Overall accuracy	$\frac{TP}{Total}$	$\frac{TP+TN}{Total}$	General accuracy of the classifier. Shows the capability of the classifier, without specifically accounting for class imbalance.
Precision	$\frac{TP}{TP+FP}$		Shows the capability to detect all changes or a class of interest.
Recall	$\frac{TP}{TP+FN}$		Shows the capacity to avoid missing changes. This is particularly of interest for change detection.
F1-score	$2\,\frac{precision * recall}{precision + recall}$		The harmonic mean of precision and recall. This trade-off is specifically relevant for classification.
Area Under the ROC Curve (AUC)	For a set of thresholds, a set of pairs of TP and FP are obtained. For each FP, the associated TP is plotted as a ROC curve point. The higher the TP, the better. The AUC is the area below this curve.		The ROC curve describes the capacity of a classifier to deal with false positives and negatives for different thresholds of its output scores. The area under this ROC curve is a measure of the classifier accuracy performances.

Picterra Sàrl 2017, all rights reserved

These measures all lie in the range [0,1] with a F1-score of 1 being the optimal classifier making no false neither missed classification. AUC ranging between 1 and 0.5 is better than random, while below 0.5 is worse than a random classifier.

This will allow to understand which class groups are best adapted to automatic classification and change detection and support the design of a system for the semi-automatic Arealstatistik2020 interpretation.

7.2. Test data description

7.2.1. Aerial data

The currently used data by BFS are stereo pairs of aerial images produced by swisstopo and with 3 bands: either R,G,B or NIR,R,G. Additionally, the orthomosaics SwissImage at 25cm of resolution (50cm over alpine areas) are used and since 2016 (also as R,G,B or NIR,R,G), and finally the SwissImageRemoteSensing image tiles with directly 4 spectral bands in a bundle (NIR,R,G,B) are available are used currently.

The following table and figure describes and locate the tiles of aerial data that have been provided for the tests. Note that the tile "1047 – Basel" has been discarded due to a shape file with wrong data sample points. Additionally, the tiles over "1032 - Diessenhofen" have been discarded for change detection in order to keep a similar setup in terms of data type between AREA3 and AREA4 (both SWISSIMAGE).

Table 3 - Description of the test dataset

Legend:	SWISSIMAGE LV03 RGB data used for classification only	SWISSIMAGE LV03 RGB data used for classification and change detection	SWISSIMAGE RS LV95 RGB data used for classification only

NOTE: Data marked as gray were not used for the test experiments

Map code	Map name	AREA	Data	Tiles #	Filenames	Resolution	Class.	CD
1106	Moutier	AREA4	SWISSIMAGE LV03	4	"DOP25_LV03_1106-*_2014_1_14.tif"	25cm	X	-
			SWISSIMAGE FCIR LV95	4	"1106-*.tif"		-	-
1171	Bülach	AREA4	SWISSIMAGE RS LV95	6	"20160623_****_1308_0_*.tif"	25cm	X	-
1231	Urseren	AREA4	SWISSIMAGE RS LV95	4	"20160907_****_1308_0_*.tif"	25cm	X	-
1285	Les Diablerets	AREA4	SWISSIMAGE LV03	4	"DOP25_LV03_1285-2*_2013_1_13.tif"	25cm	X	X
			SWISSIMAGE FCIR LV95	4	"1285-2*.tif"		-	-
		AREA3	SWISSIMAGE LV03	4	"DOP50_LV03_1285-21_2004_2.tif"	50cm	-	X
1032	Diessenhofen	AREA4	SWISSIMAGE RS LV95	24	"20160623_08**_1308_0_*.tif"	25cm	X	-
		AREA3	SWISSIMAGE LV95	8	"DOP25_LV95_1032-11_2008_1_08.tif"		-	-
1088	Hauenstein	AREA4	SWISSIMAGE LV03	5	"DOP25_LV03_1088-22_2015_1_15.tif"	25cm	X	X
			SWISSIMAGE FCIR LV95	5	"1088-2*.tif"		-	-
		AREA3	SWISSIMAGE LV03	5	"DOP25_LV03_1088-**_2006_1_06.tif"	50cm	-	X
1166	Bern	AREA4	SWISSIMAGE LV03	5	"DOP25_LV03_1166-**_2013_1_13.tif"	25cm	X	X
			SWISSIMAGE FCIR LV95	5	"1166-**.tif"		-	-
		AREA3	SWISSIMAGE LV03	5	"DOP50_LV03_1166-**_2004_2.tif"	50cm	-	X
1204	Romont	AREA4	SWISSIMAGE LV03	4	"DOP25_LV03_1204-11_2013_1_13.tif"	25cm	X	X
			SWISSIMAGE FCIR LV95	4	"1204-**.tif"		-	-
		AREA3	SWISSIMAGE LV03	4	"DOP50_LV03_1204-11_2004_2.tif"	50cm	-	X
1206	Guggisberg	AREA4	SWISSIMAGE LV03	4	"DOP25_LV03_1206-41_2013_1_13.tif"	25cm	x	X
			SWISSIMAGE FCIR LV95	4	"1206-**.tif"		-	-
		AREA3	SWISSIMAGE LV03	4	"DOP50_LV03_1206-41_2004_2.tif"	50cm	-	x

Picterra Sàrl 2017, all rights reserved

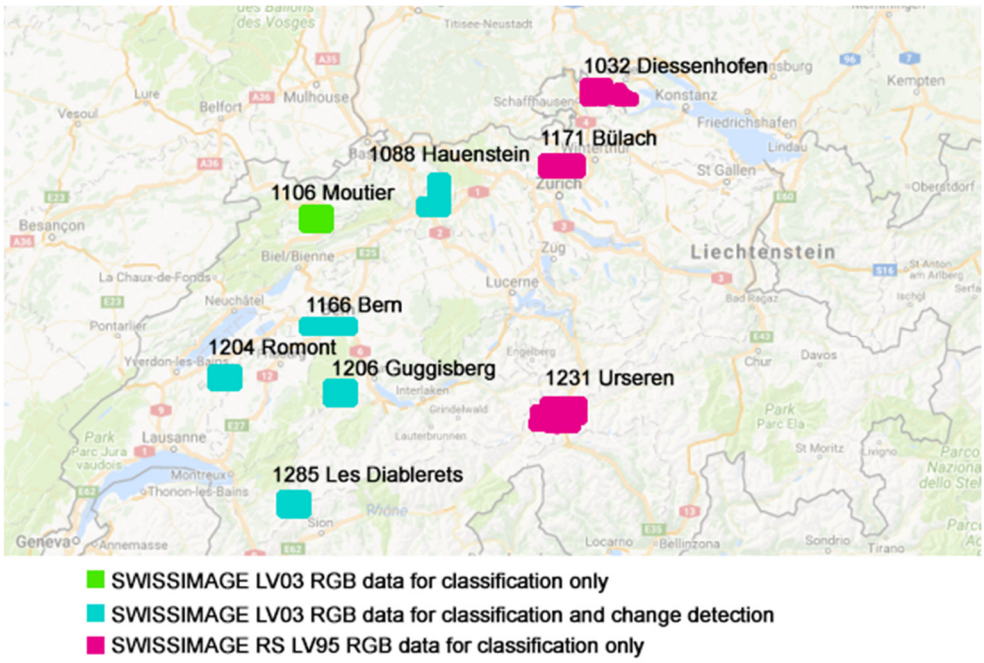


Figure 19 - Test data tiles, type and reference frame (LV03, LV95)

7.2.2. Auxiliary information

Many auxiliary sources of information are exploited by the Arealstatistik expert to derive the land cover / land use classes:

- Cadastral information from the different cantons
- National maps (1:25'000)
- Topographic landscape model (TLM) from Swisstopo: SwissTLM3D
- Terrain model for altitude, slope, exposition information from SwissAlti, Swisstopo
- GWR data on building usage

Remark: *Notice that the tests performed in the scope of this study are only based on aerial imagery. No additional information such as terrain slope, orientation, building usage (e.g. GWR) or cadastral information were used. Improvements in classification accuracies can be expected when feeding this auxiliary information in the classification models, typically in order to discriminate subtle land use classes in urban environment.*

7.2.3. Ground truth information

The aerial data is photo-interpreted by an expert and reviewed by a second one over the sample points on the grid of interpretation (point every 100m). In case of disagreements between two expert interpretations, an in-situ visit is performed to get a final interpretation. In addition, a small percentage of sample points is "blindly" interpreted by the second expert and allow to assess the interpretation variance.

Picterra Sàrl 2017, all rights reserved

For the test of this study, sample points over the data tiles were made available. Here a summary of the number of points for the different tiles:

Map code	Map name	Sample points	Tiles #	Filenames	Class.	CD
1106	Moutier	5046	4	“1106.shp”	X	
1171	Bülach	7928	6	“1171.shp”	X	
1231	Urseren	17443	4	“1231.shp”	X	
1285	Les Diablerets	5046	4	“1285.shp”	X	X
1032	Diessenhofen	10085	24	“1032.shp”	X	
1088	Hauenstein	6264	5	“1088.shp”	X	X
1166	Bern	5046	5	“1166.shp”	X	X
1204	Romont	5046	4	“1204.shp”	X	X
1206	Guggisberg	5046	4	“1206.shp”	x	X
Total		**66950**	**60**		**66950**	**26448**

Considering that the entire interpretation for one Arealstatistik survey is around 4’128’500 sample points, this study test comprises 66’950 sample points which represents 1.6% of the entire set of points covering Switzerland.

The proportion of the different classes present in the test dataset is represented below in Figure 20 and Figure 21, for classification (AREA4). The classes distribution between the test dataset and the entire Switzerland are very similar for both land cover and land use classes. Some noticeable differences are with land cover classes 51 and 52 which are over represented in the test dataset. For land use classes, the class 421 is over represented and class 401 is under represented in the test dataset.

Picterra Sàrl 2017, all rights reserved

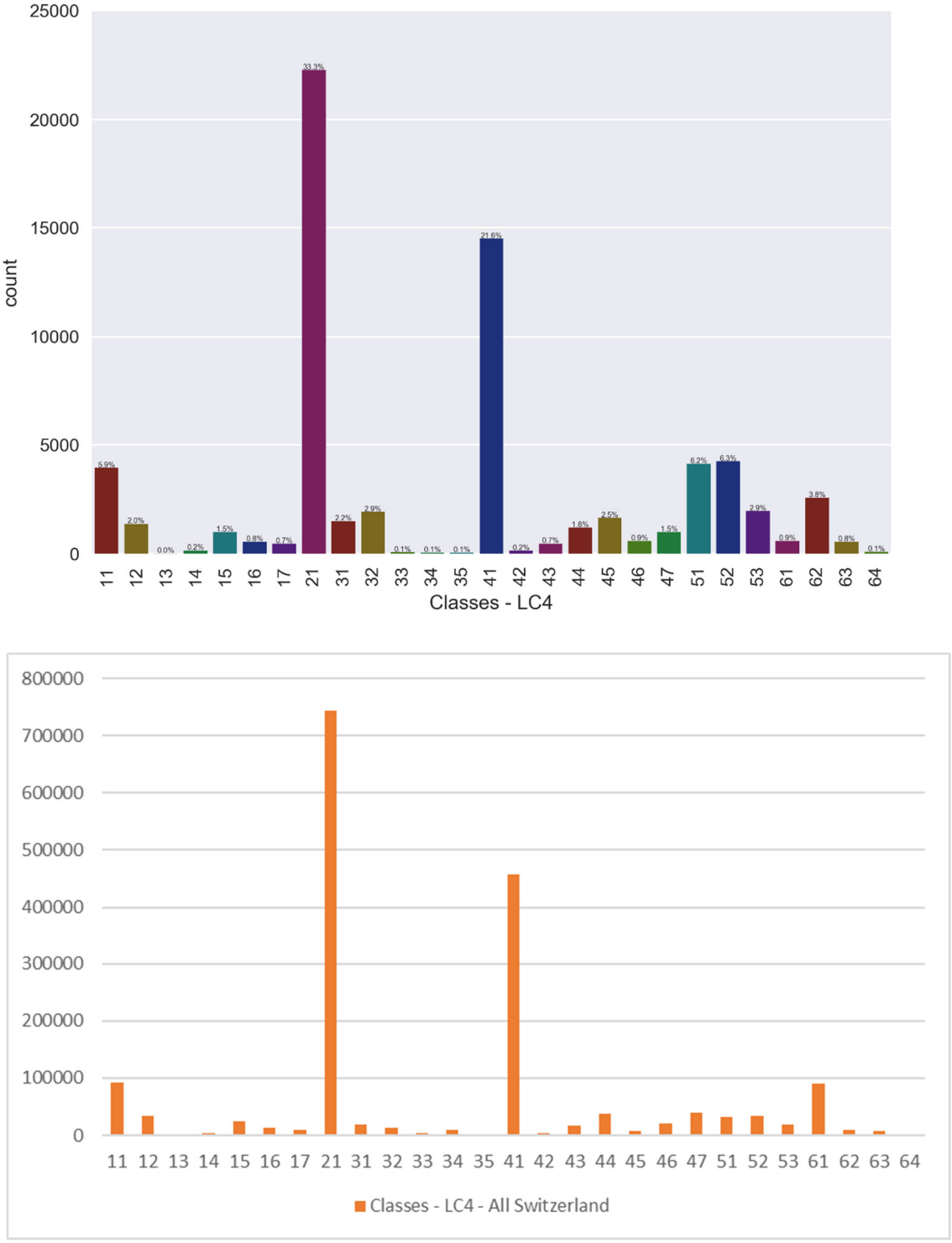


Figure 20 - Top: Distribution of the different land cover classes within the test dataset for classification (AREA4). Bottom: Distribution of the different land cover classes for the entire Switzerland.

Picterra Sàrl 2017, all rights reserved

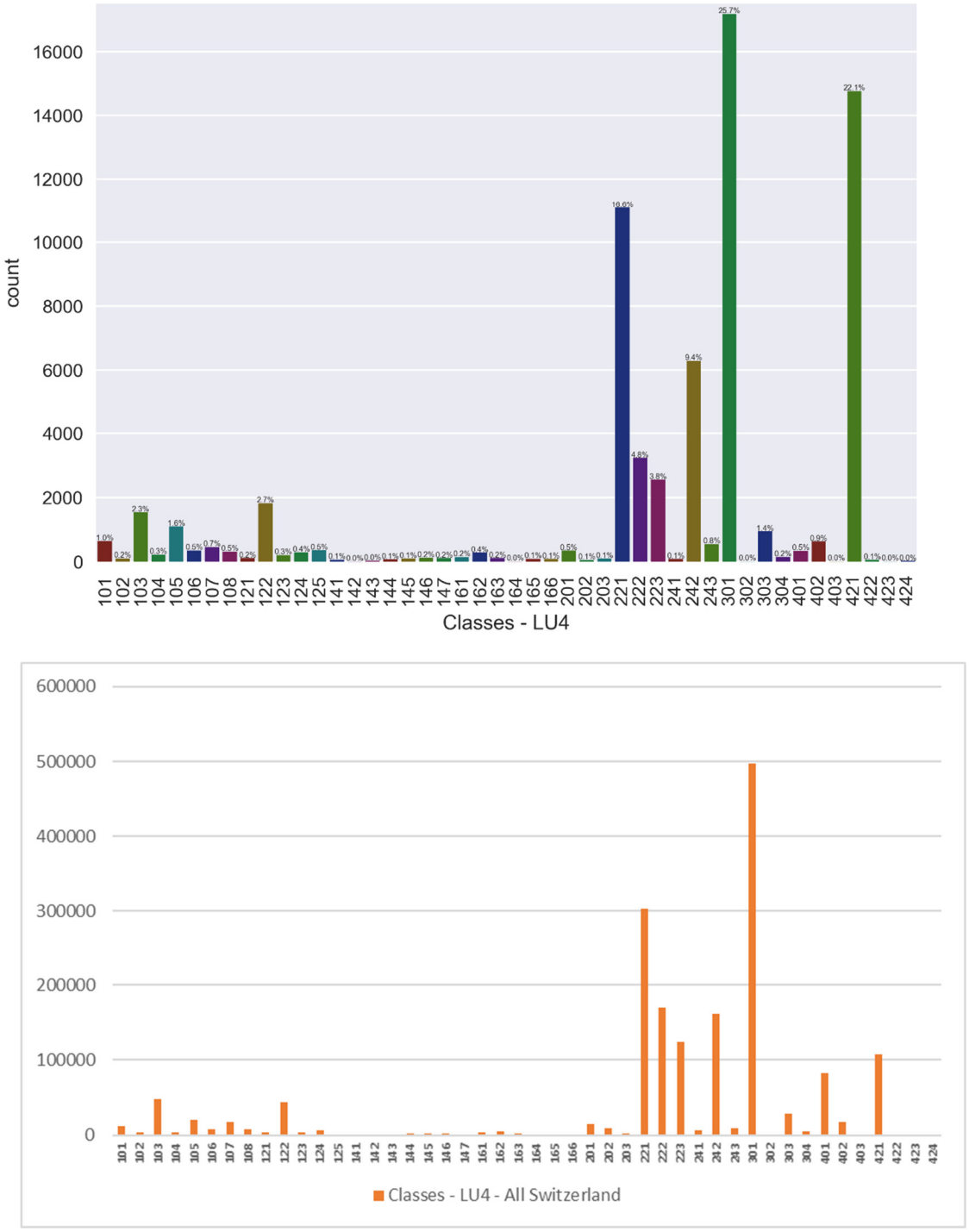


Figure 21 - Top: Distribution of the different land use classes within the test dataset for classification (AREA4). Bottom: Distribution of the different land use classes for the entire Switzerland.

The proportion of the different type of changes present in the test dataset is represented below, from AREA3 to AREA4:

Picterra Sàrl 2017, all rights reserved

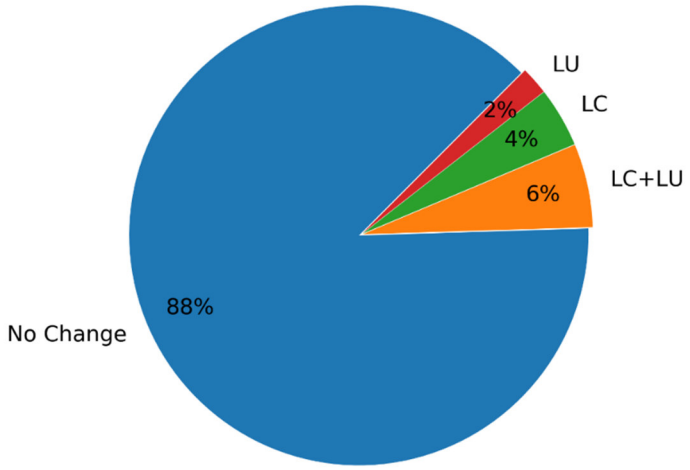


Figure 22 - Changes distribution among land cover and land use

Map code	Map name	Sample points	No change	Land Use change	Land Cover change	LU+LC change
1285	Les Diablerets	5046	4517	40	278	211
1088	Hauenstein	6264	5359	218	166	521
1166	Bern	5046	4203	176	134	533
1204	Romont	5046	4528	212	61	245
1206	Guggisberg	5046	4424	55	249	318
Total		26448	23031	701	888	1828
Percentage %		100%	87.1%	2.6%	3.4%	6.9%

7.3. Data pre-processing

7.3.1. Aerial data

In all the different classification and change detection tests, image patches around the sample points needed to be extracted. An extent of 50m around the sample point has been considered in order to extract the patches (see Figure 1, the white box for an example of patch):

The sample points and their extended area of 50x50m, which were not completely covered by the aerial data tiles, have been discarded. This typically happened at the border of a tile, between tiles or in the border of SwissImageRemoteSensing products containing black regions without data.

The image patches extracted where kept in their original geographic projection system (either CH1903 LV03 or LV95). The resolution of the patches was kept to the original, however in the case of change detection the patches are down sampled to the lowest resolution (AREA3) if a difference of resolution exists between AREA3 and AREA4. This is the case on certain tiles but not all of them.

7.4. Experiments and method description

7.4.1. General considerations

The experiments performed in this study have the purpose to validate certain methodological choices and provide a view on the classes more capable of being supported by automation.

It remains very important to underline here the scope and limitations of such experiments in comparison to the real Arealstatistik interpretation setup:

Picterra Sàrl 2017, all rights reserved

- Aerial data, in RGB colors, has been used for the tests. No additional auxiliary data was exploited at this stage.
- Near-infrared band in aerial data was not exploited. This might improve certain discrimination of vegetated classes.
- The pre-trained CNNs model where not fine-tuned towards BFS data in the frame of this study due to time constraints. The usage of fine-tuned models and data augmentation typically improves the class accuracies by 10% (see Cheng et al. 2017).
- From the state-of-the-art, the usage of DSM with CNNs models is known to improve classification and would definitely support change detection approaches.

7.4.2. Classification

The land cover and land use classifications are performed separately. A first step of feature extraction is performed (deep features extracted from pre-trained CNNs) and classifiers are trained to predict either land cover or land use classes.

7.4.2.1. Approach A3: Deep feature-based classification

The classification is based on pre-trained CNNs from ImageNet to extract features first and then a random forest classifier for the classification step. The following processing steps are performed:

- Each patch (AREA4) features are fed to the pre-trained CNN model, here ReSNet-50. More details on the ResNet-50 CNN model architecture can be found in the Appendix.
- A total of 2048 features per patch are extracted from the last convolutional layer of the CNN.
- Parameters for the Random Forest Classifier are searched in a stratified k-fold cross-validation scheme (k = 10 folds).
- Optimal parameters are used to train a Random Forest Classifier and predict on a test fold (out-of-training) samples.

The stratified k-fold cross-validation allows to estimate the overall accuracy and thus to search for optimal parameters on training data by considering folds of data equally distributed in terms of classes. Once optimal parameters are obtained on training and validation folds, the optimal Random Forest Classifier is used to predict on the test fold. The Figure 23 schematize the k-fold cross-validation strategy.

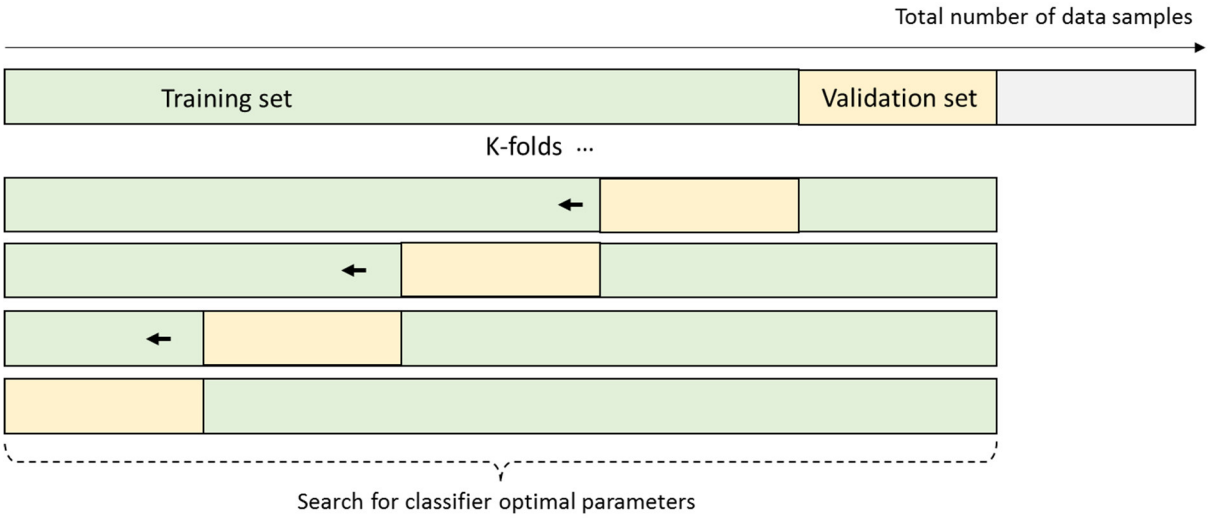


Figure 23 - Stratified K-fold cross-validation: training set and prediction on validation set allows to search for the optimal parameters best generalizing the classification on unseen data.

Furthermore, each fold becomes part of the training, validation, or test sets over the iterations. At the end, this allows to obtain a prediction on all the data samples without being biased. Each test set is predicted using

Picterra Sàrl 2017, all rights reserved

the rest of the data (training + validation folds), therefore each trained Random Forest Classifier predicts on unseen test data. Figure 24 illustrates this procedure of prediction on entire dataset.

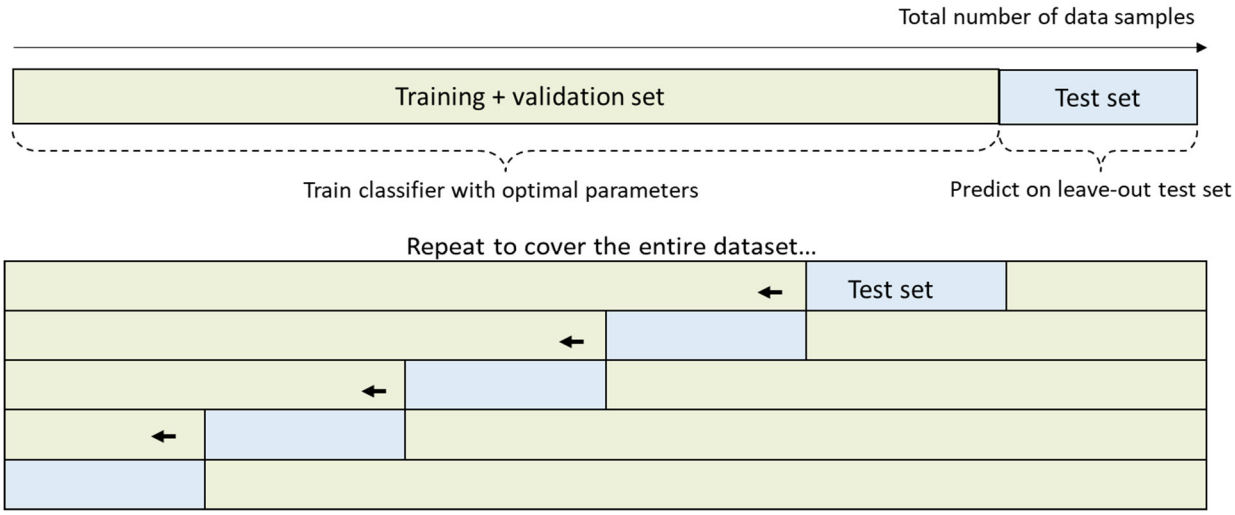


Figure 24 - Prediction over the entire datasets by successive stratified K-fold cross-validation followed by prediction on the left-out test set

The following parameters and ranges were optimized during training phase:

- Number of decision trees: [50, 200, **500**, 1000]
- Maximum depth of tree: [100, 1000, **2000**, 3000]
- Minimum samples per leave: [**1**, 10, 50]

Optimal parameters selected for the classification tests are highlighted in **bold**. Please note that permutations of all parameters were investigated, which leads to 3x4x3 = 36 different unique combinations of parameters.

7.4.3. Change detection

The change detection experiments aim at testing the approaches C1 and C3 and assessing the false/missed detection rates. As for the classification, first features were extracted, and then Support Vector Machines classifier were trained. The SVM was performing better than random forest due to the lower number of samples and prediction of only two classes (change vs. non-change). The optimization of parameters and training/validation/testing scheme is similar for change detection than classification. Additionally, two variants were performed by training a set of change detection classifiers, one per AREA3 class data samples, land cover or land use. We trained, validated and tested a classifier in the same manner as described above (stratified k-fold cross validation) with the subset of samples from the same AREA3 class. This allows to get classifiers which are more specific to certain change types and ultimately performing better than a general unique change detection classifier.

7.4.3.1. Approach C1: pixel-based feature change detection

The change detection approach based on pixel features is based on the following processing steps:

For each pair of patches AREA3&AREA4:

- Resampling of the image patch to the lowest resolution (25cm to 50cm for some AREA4 patches)
- Co-registration of the corresponding patches from AREA3 and AREA4
- Detection of shadows based on the LAB color intensity image transformation
- Mask of the pixels covered by shadows on AREA3 and AREA4 patches
- Computation of image difference: AREA4 - AREA3
- Change Vector Analysis based on image difference: for each pixel the magnitude and angle of change are retrieved:

Picterra Sàrl 2017, all rights reserved

- CVA magnitude = $\sqrt{\sum_{i=0}^{3}\left(x_{AREA4}^{i}-x_{AREA3}^{i}\right)^{2}}$, where i is the dimension index equal to 3 (RGB)
- CVA angle = angle in degree between the two vectors AREA3 and AREA4.

- 20'000 features generated from the CVA magnitude and angles on the patch 100x100 pixels.
- Storage of 2 aggregated features = average CVA magnitude & average CVA angle over the patch pixels

Then a Support Vector Machine is trained as follows:

- Parameters for the Support Vector Machine Classifier are searched in a stratified k-fold cross-validation scheme (k = 10 folds).
- Optimal parameters are used to train a Support Vector Machine Classifier and predict on a test fold (out-of-training) samples.

The following parameters and ranges were optimized during training phase (penalty parameter used for the change detection tests is highlighted bold):

- Penalty parameter C: [0.1, 1, **10**, 100]
- Class weights: 'balanced' which weights classes related to their frequencies, or 'equivalent'

7.4.3.2. Approach C3: Deep feature-based change detection

The change detection approach based on deep features is based on the following processing steps:

For each pair of patches AREA3&AREA4:

- Resampling of the image patch to the lowest resolution (25cm to 50cm for some AREA4 patches)
- Feed forward of the patches in the pre-trained CNN model, here ResNet50. More details on the ResNet-50 CNN model architecture can be found in the Appendix.
- A total of 2048 features per patch are extracted from the last convolutional layer of the CNN.
- Difference between the features from AREA4 and AREA3 patches.
- Storage of 2048 features extracted

Then a Support Vector Machine is trained as follows:

- Parameters for the Support Vector Machine Classifier are searched in a stratified k-fold cross-validation scheme (k = 10 folds).
- Optimal parameters are used to train a Support Vector Machine and predict on a test fold (out-of-training) samples.

As for Approach C1, optimal parameters for the Support Vector Machine Classifier are searched in similar ranges.

7.5. Experiments results

7.5.1. Classification

The results of classification using method A3 based on deep features are presented hereafter with accuracy measures per class.

The optimal parameters for the Random Forest Classifiers trained for land cover and land use prediction are similar and are the following:

Picterra Sàrl 2017, all rights reserved

- Number of decision trees: **500**
- Maximum depth of tree: **2000**
- Minimum samples per leave: **1**

The sensitivity to these parameters is low with only of variation in overall accuracy of 0.03. The general overall accuracies are for land cover: 0.693 +- 0.076 and for land use: 0.707 +- 0.038.

Here below are reported the precision, recall and the F1-score. These measures all lie in the range [0,1] with a F1-score of 1 being the optimal classifier making no false neither missed classification. The precision tells if a classifier is over-predicting a given class (false detections) and recall tells if a classifier is under-predicting a given class (missed detections). Additionally, the results for top-3 classification, which assess the presence of the correct class within the 3 top predictions, are reported. The result table cells are colored based on the F1-score values and distributed as described in Table 4. The range from 1 to 0.8 represents good classification performances, the range 0.8 to 0.55 represents classification performances which are better than a random predictor and could improve over time with more data interpreted, and finally the range 0.55 to 0 represents bad performances where the prediction confidence level is very low and automatic interpretation should be avoided.

Table 4 - Description of cell colors for the result tables

F1-score range	Description
[1 - 0.8]	Good classification performances
]0.8 – 0.55]	Moderate classification performances
]0.55 – 0]	Bad classification performances

Picterra Sàrl 2017, all rights reserved

The Table 5 reports the classification accuracy measures per AREA4 land cover class.

Table 5 - Classification results per land cover class for approach A3

Land cover classification results

AREA4 LC class	#Nb	Precision	Recall	F1-score
41	14490	0.82	0.97	0.89
21	22280	0.76	0.95	0.85
62	2574	0.93	0.79	0.85
51	4126	0.65	0.8	0.72
52	4237	0.74	0.7	0.72
45	1645	0.71	0.65	0.68
61	592	1	0.41	0.58
11	3938	0.44	0.66	0.52
32	1914	0.47	0.4	0.43
53	1952	0.5	0.33	0.4
31	1492	0.64	0.09	0.15
33	67	1	0.01	0.03
12	1372	0.42	0	0.01
44	1195	1	0	0.01
13	14	0	0	0
14	148	0	0	0
15	999	0	0	0
16	552	0	0	0
17	453	0	0	0
34	49	0	0	0
35	54	0	0	0
42	134	0	0	0
43	463	0	0	0
46	583	0	0	0
47	995	0	0	0
63	541	0	0	0
64	91	0	0	0

Land cover TOP-3 classification results

AREA4 LC class	#Nb	Precision	Recall	F1-score
62	2574	0.99	0.94	0.97
12	1372	1	0.93	0.96
41	14490	0.91	1	0.95
52	4237	0.96	0.94	0.95
21	22280	0.9	0.99	0.94
51	4126	0.9	0.99	0.94
44	1195	1	0.85	0.92
45	1645	0.93	0.85	0.89
53	1952	0.89	0.82	0.85
11	3938	0.74	0.95	0.83
15	999	1	0.68	0.81
63	541	1	0.67	0.81
32	1914	0.86	0.75	0.8
31	1492	0.98	0.59	0.74
61	592	1	0.57	0.73
33	67	1	0.57	0.72
34	49	1	0.55	0.71
47	995	1	0.44	0.61
16	552	0.99	0.22	0.35
46	583	1	0.21	0.35
43	463	1	0.2	0.33
35	54	1	0.11	0.2
17	453	1	0.08	0.15
14	148	1	0.02	0.04
64	91	1	0.02	0.04
42	134	1	0.01	0.01
13	14	0	0	0

The major LC classes: LC 41, 21, 62, representing 59% (39344/66950) of the total samples are among the highest F1-score scores with an average of 0.863. Moreover, when considering the top-3 predictions a larger number of land cover classes (10) have high accuracies and could be quickly interpreted by the expert using the top-3 predictions as recommendations.

Picterra Sàrl 2017, all rights reserved

The Table 6 reports the classification accuracy measures per AREA4 land use class.

Table 6 - Classification results per land use class for approach A3

Land use classification results

AREA4 LU class	#Nb	Precision	Recall	F1-score
301	17176	0.84	0.96	0.89
421	14763	0.83	0.92	0.87
401	334	1	0.72	0.84
221	11083	0.72	0.96	0.82
242	6285	0.62	0.71	0.66
103	1533	0.38	0.83	0.53
101	642	0.51	0.55	0.53
105	1092	0.41	0.3	0.35
222	3223	0.55	0.26	0.35
125	356	1	0.18	0.3
223	2546	0.47	0.21	0.29
122	1815	0.3	0.18	0.23
124	289	0.75	0.1	0.18
107	453	0.75	0.01	0.03
108	305	0.75	0.01	0.02
121	109	1	0.01	0.02
162	273	1	0	0.01
201	340	1	0	0.01
102	102	0	0	0
104	214	0	0	0
106	346	0	0	0
123	203	0	0	0
141	64	0	0	0
142	3	0	0	0
143	28	0	0	0
144	69	0	0	0
145	96	0	0	0
146	117	0	0	0
147	107	0	0	0
161	136	0	0	0
163	104	0	0	0
164	7	0	0	0
165	84	0	0	0
166	82	0	0	0
202	49	0	0	0
203	93	0	0	0
241	80	0	0	0
243	544	0	0	0
302	7	0	0	0
303	939	0	0	0
304	148	0	0	0
402	625	1	0	0
403	11	0	0	0
422	52	0	0	0
423	8	0	0	0
424	15	0	0	0

Land use top-3 classification results

AREA4 LU class	#Nb	Precision	Recall	F1-score
301	17176	0.93	1	0.96
421	14763	0.94	0.99	0.96
242	6285	0.89	0.99	0.94
221	11083	0.88	0.98	0.93
303	939	1	0.84	0.92
222	3223	0.89	0.88	0.88
401	334	1	0.75	0.86
105	1092	0.77	0.92	0.84
223	2546	0.86	0.81	0.83
125	356	1	0.63	0.78
101	642	0.68	0.79	0.73
103	1533	0.59	0.94	0.72
122	1815	0.71	0.7	0.7
124	289	0.95	0.42	0.58
243	544	1	0.29	0.45
145	96	1	0.23	0.37
121	109	1	0.22	0.36
163	104	1	0.22	0.36
304	148	1	0.2	0.34
422	52	1	0.19	0.32
107	453	0.97	0.13	0.23
202	49	1	0.12	0.22
165	84	1	0.12	0.21
162	273	1	0.11	0.2
402	625	1	0.11	0.2
123	203	1	0.09	0.17
141	64	1	0.09	0.17
201	340	1	0.09	0.17
104	214	1	0.07	0.13
108	305	1	0.07	0.13
144	69	1	0.04	0.08
146	117	1	0.04	0.08
166	82	1	0.04	0.07
241	80	1	0.03	0.05
106	346	1	0.01	0.02
147	107	1	0.01	0.02
203	93	1	0.01	0.02
161	136	1	0.01	0.01
102	102	0	0	0
142	3	0	0	0
143	28	0	0	0
164	7	0	0	0
302	7	0	0	0
403	11	0	0	0
423	8	0	0	0
424	15	0	0	0

The major LU classes: LU 301, 421, 401, 221, representing 68% (43356/66950) of the total samples are among the highest F1-score scores with an average of 0.855. Moreover, when considering the top-3 predictions a larger number of land use classes (5) have high accuracies and could be quickly interpreted by the expert using the top-3 predictions.

Picterra Sàrl 2017, all rights reserved

Here below are presented the ROC curves for the 3 land cover and land use classes with highest accuracies, in terms of F1-score, which are the classes best adapted for automatic classification.

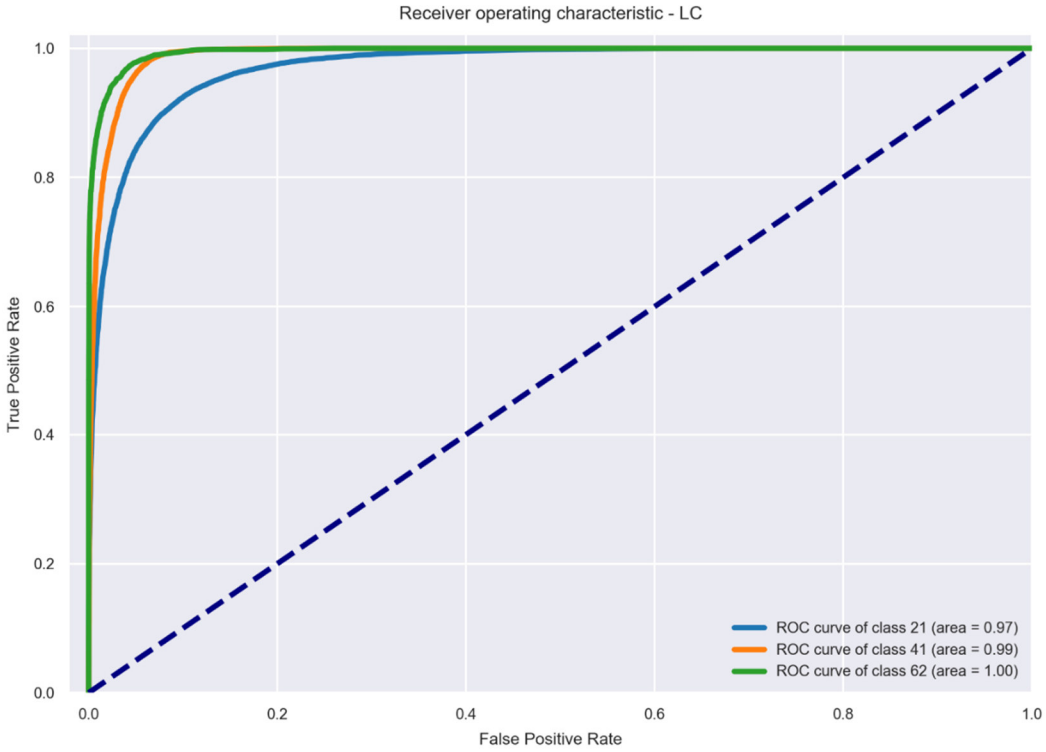


Figure 25 - ROC curves for the land cover classes with the top 3 F1-score

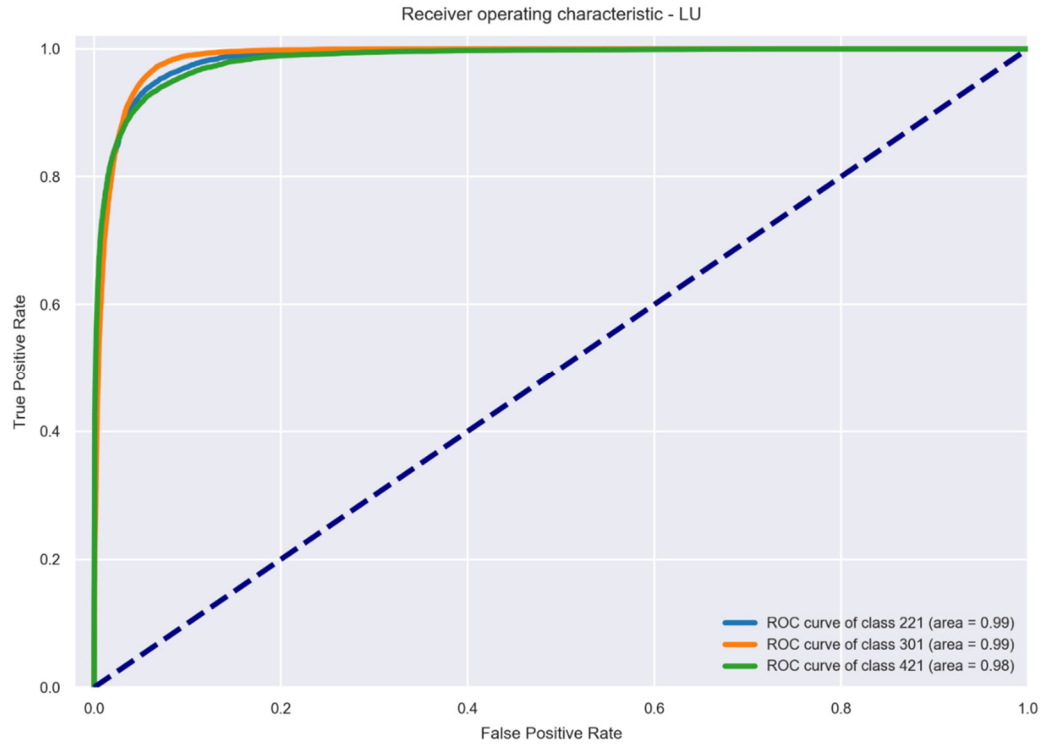


Figure 26 - ROC curves for the land use classes with the top 3 F1-score

General issue of imbalanced distribution of data points among the different classes for LC and LU has also been addressed and data-rebalancing using reweighted samples was exploited. It should be mentioned that no significant performance improvement was observed when using such data-rebalancing.

Picterra Sàrl 2017, all rights reserved

Regarding the distribution of the prediction scores, Figure 27 shows the average ranked scores for LC and LU classes. It can be observed that the scores are very quickly dropping with the first score above 60% and the second one around 20%. The top-3 scores represent by themselves 90% of the classifier confidence level.

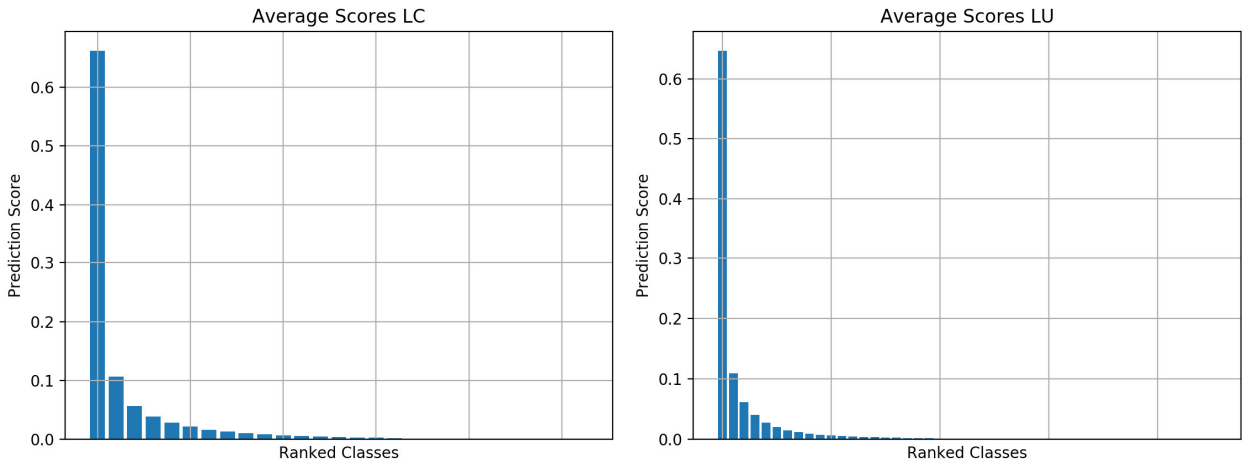


Figure 27 - Average ranked classification scores for land cover and land use classes

Observing directly the difference between the two highest scores, the second and third and finally third and fourth scores, it can be seen in Figure 28 that most of the scores have a large difference between the highest and second one (green bars), with only a small fraction having a difference of scores below 10% (0.1).

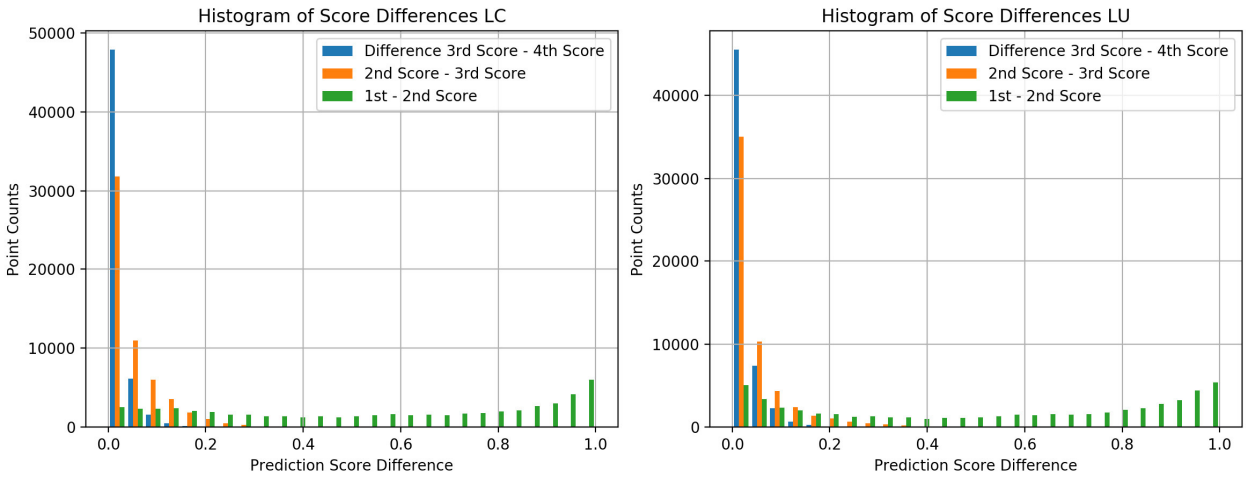


Figure 28 - Distribution of the difference between the highest classification scores for land cover and land use classes

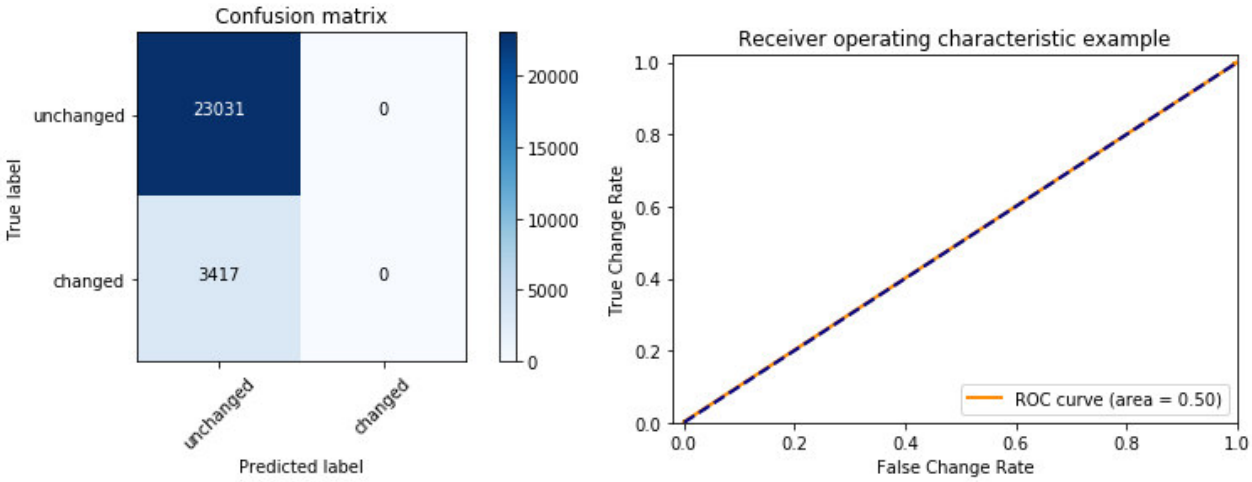
The difference of scores between the second and third are within a 20% range and finally the third and fourth scores are very small and therefore similar with a difference of around 1 to 5%.

7.5.2. Change detection

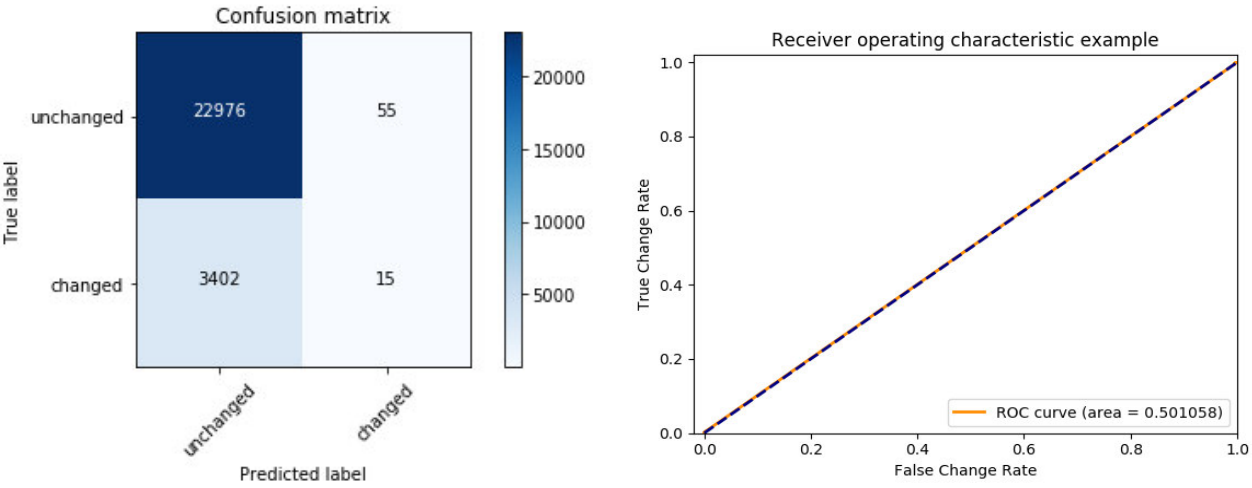
7.5.2.1. Approach C1: pixel-based feature change detection

The results of change detection using method C1 where generally not satisfactory. Training with features from all pixels (20’000 features) or from averaged pixel features (2 features) led to a classifier predicting always “unchanged”.

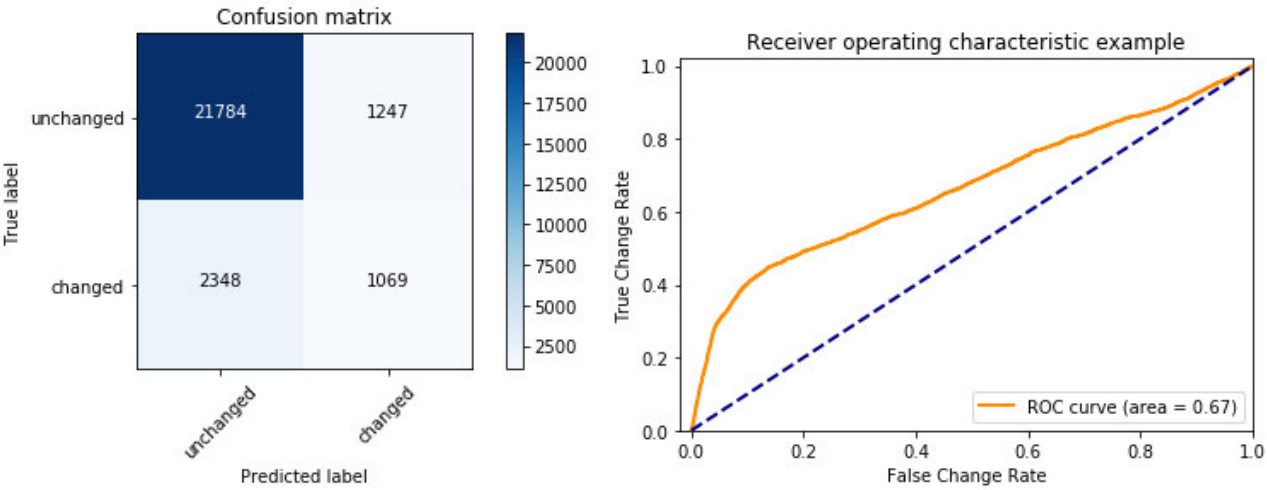
Picterra Sàrl 2017, all rights reserved



When training/validating/testing a classifier per land use classes (a total of 46 classifiers), the aggregated results are:



When training/validating/testing a classifier per land cover classes (a total of 26 classifiers), the aggregated results obtained are:



The best results with pixel-based features were therefore obtained by dividing the data by similar initial land cover (AREA3) and training individual classifiers (26 SVMs) specific to each of the initial land cover classes. The optimal parameters for the trained linear Support Vector Machine Classifiers are:

Picterra Sàrl 2017, all rights reserved

- Penalty C: 10
  Class weight: 'balanced' (related to class sample frequencies)

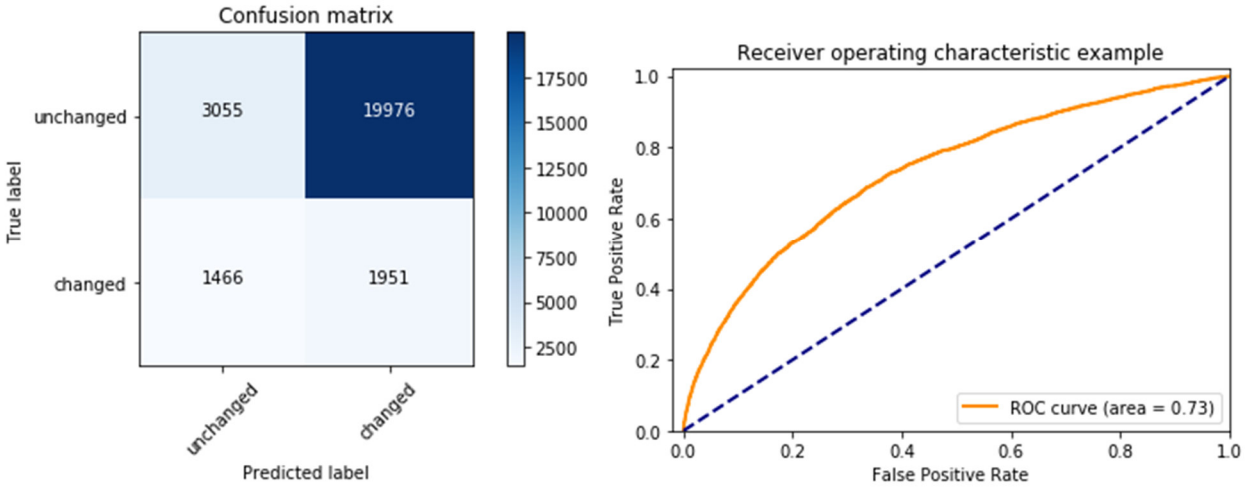
7.5.2.2. Approach C3: Deep feature-based change detection

The optimal parameters for the trained linear Support Vector Machine Classifiers are always:

- Penalty C: 10
- Class weight: 'balanced' (related to class sample frequencies)

The sensitivity to the parameter C is rather low, with a variation of the overall accuracy of ~0.01 in the range C=[1, 100].

For a unique model trained on samples from all classes, the average change detection overall accuracy is 0.777 +- 0.084. The Confusion matrix and aggregated ROC curve is presented below:



A finer interpretation of the results can be done with the reported overall accuracy (true non-change + true change / all samples) and the AUC measures per AREA3 class. The AUC provides a good measure of the trade-off between false detection and missed detection of changes, it lies in the range [0,1] with 1 being an optimal classifier making no false detection neither missed detection.

Picterra Sàrl 2017, all rights reserved

The Table 7 reports the change detection accuracy measures per AREA3 land cover class. The table cells are colored based on the AUC values and distributed as described in Table 4.
Please note that in cases where only changes or non-changes appear, a minus sign (-) indicates that the AUC cannot be measured.

Table 7 - Change detection results per class for approach C3

		Unique Change Detection Model		Change Detection Model per LC class		Change Detection Model per LU class	
AREA3 class	#Nb	OA	AUC	OA	AUC	OA	AUC
LC 11	1629	0.77470841	0.59904271	0.947207	0.722461	0.932474	0.752996
LC 12	648	0.780864	0.550241	0.95679	0.694831	0.921296	0.700289
LC 13	5	0.8	-	0.8	1	0.8	1
LC 14	73	0.616438	0.561975	0.753425	0.731092	0.684932	0.415966
LC 15	415	0.691566	0.467341	0.785542	0.552835	0.80241	0.60778
LC 16	275	0.647273	0.496108	0.807273	0.660053	0.785455	0.474654
LC 17	240	0.6625	0.548234	0.7	0.623544	0.766667	0.603552
LC 21	10629	0.845235	0.696021	0.839213	0.684781	0.902719	0.846027
LC 31	751	0.667111	0.684253	0.876165	0.892354	0.858855	0.884732
LC 32	357	0.647059	0.712654	0.759104	0.837758	0.719888	0.797454
LC 33	23	0.521739	0.366071	0.73913	0.866071	0.73913	0.839286
LC 35	12	0.416667	0.35	0.75	0.4	0.583333	0.3
LC 41	5916	0.7762	0.719719	0.922752	0.874802	0.905003	0.84914
LC 42	55	0.672727	0.629845	0.781818	0.616279	0.709091	0.577519
LC 43	277	0.689531	0.548294	0.837545	0.718649	0.808664	0.572845
LC 44	484	0.543388	0.566785	0.671488	0.724198	0.778926	0.811187
LC 45	235	0.680851	0.623857	0.795745	0.646	0.748936	0.570714
LC 46	309	0.660194	0.591474	0.796117	0.637775	0.721683	0.625988
LC 47	682	0.608504	0.658493	0.664223	0.735958	0.760997	0.806418
LC 51	955	0.908901	0.464774	0.993717	0.531546	0.943455	0.556782
LC 52	1065	0.840376	0.719857	0.895775	0.836649	0.899531	0.814008
LC 53	371	0.754717	0.662253	0.816712	0.680299	0.816712	0.668846
LC 61	169	0.715976	0.673567	0.934911	0.901805	0.928994	0.890127
LC 62	563	0.817052	0.688798	0.909414	0.932923	0.840142	0.778989
LC 63	304	0.921053	0.908084	0.986842	0.717608	0.963816	0.79845
LC 64	6	0.666667	-	1	-	1	-

Interestingly the major LC classes: LC 21, 41, 52, representing 67% (17610/26448) of the total samples are among the highest AUC scores with an average of 0.798 (per LC3 change detection models) or 0.836 (per LU3 change detection models).

Picterra Sàrl 2017, all rights reserved

The Table 8 reports the change detection accuracy measures per AREA3 land use class.

Table 8 - Change detection results per class for approach C3

		Single Change Detection Model		Change Detection Model per LC class		Change Detection Model per LU class	
AREA3 class	#Nb	OA	AUC	OA	AUC	OA	AUC
LU 101	243	0.765432	0.636578	0.868313	0.697848	0.860082	0.707688
LU 102	41	0.780488	0.705556	0.804878	0.566667	0.878049	0.361111
LU 103	672	0.700893	0.538872	0.824405	0.700877	0.751488	0.527004
LU 104	82	0.719512	0.591667	0.804878	0.765476	0.817073	0.569048
LU 105	521	0.754319	0.536336	0.886756	0.665175	0.886756	0.540511
LU 106	197	0.705584	0.490125	0.86802	0.722303	0.857868	0.524326
LU 107	195	0.702564	0.499058	0.897436	0.782172	0.887179	0.492781
LU 108	181	0.79558	0.59688	0.917127	0.670772	0.955801	0.662562
LU 121	54	0.611111	0.506818	0.759259	0.643182	0.833333	0.904545
LU 122	823	0.792224	0.528322	0.948967	0.643154	0.972053	0.620057
LU 123	92	0.73913	0.522321	0.923913	0.639881	0.902174	0.622024
LU 124	133	0.774436	0.707154	0.887218	0.727273	0.879699	0.45082
LU 125	4	0.75	0	0.75	0.666667	0.75	1
LU 141	4	0.5	0.5	0.75	0.5	0.25	0
LU 142	2	1	1	1	1	0.5	0.5
LU 143	12	0.333333	0.407407	0.75	0.592593	0.75	0.814815
LU 144	17	0.588235	0.666667	0.764706	0.8	0.882353	1
LU 145	77	0.636364	0.631081	0.584416	0.622297	0.649351	0.702027
LU 146	30	0.3	0.0344828	0.7	0.827586	1	1
LU 147	33	0.515152	0.479339	0.545455	0.68595	0.636364	0.657025
LU 161	98	0.632653	0.5	0.795918	0.653409	0.877551	0.619318
LU 162	112	0.660714	0.341134	0.857143	0.59766	0.910714	0.715572
LU 163	1	1	-	1	-	1	-
LU 164	4	0.75	1	0.5	0.5	0.75	1
LU 165	21	0.619048	0.481481	0.904762	0.648148	0.809524	0.833333
LU 166	34	0.647059	0.841379	0.852941	0.924138	0.852941	0.317241
LU 201	199	0.557789	0.540579	0.61809	0.59757	0.728643	0.764668
LU 203	29	0.517241	0.5	0.517241	0.590909	0.413793	0.353535
LU 221	4882	0.891438	0.775074	0.871979	0.729189	0.937116	0.905414
LU 222	1189	0.751892	0.641538	0.75021	0.63354	0.809924	0.755358
LU 223	1429	0.652904	0.465957	0.669699	0.518915	0.789363	0.756492
LU 241	40	0.975	-	0.925	-	1	-
LU 242	3341	0.85663	0.73122	0.881473	0.783769	0.895241	0.758766
LU 243	297	0.888889	0.807692	0.902357	0.801017	0.956229	0.900509
LU 301	6635	0.763075	0.69817	0.904748	0.862014	0.888169	0.794287
LU 302	8	0.25	-	0.75	-	0	-
LU 303	416	0.576923	0.349633	0.899038	0.589242	0.973558	0.615788
LU 304	672	0.703869	0.504307	0.869048	0.778312	0.894345	0.686031
LU 401	16	0.5625	0.642857	0.875	0.714286	0.875	0.392857
LU 402	284	0.711268	0.660763	0.823944	0.771055	0.823944	0.753444
LU 403	9	0.777778	0.875	0.666667	0.5	0.888889	1
LU 421	3314	0.840676	0.695254	0.911285	0.796369	0.882016	0.707183
LU 423	4	0.75	-	0.5	-	1	-
LU 424	1	1	-	1	-	0	-

The major LU classes: LU 221, 242, 301, 421, representing 69% (18172/26448) of the total samples are among the highest AUC scores with an average of 0.793 (per LC3 change detection models) or 0.791 (per LU3 change detection models).

Picterra Sàrl 2017, all rights reserved

Based on the results obtained using a change detection model per LU class, two class groups can be defined:

- Class group 1 “for expert interpretation”:
  - Land cover: LC 14, LC 16, LC 35
  - Land use: LU 102, LU 107, LU 124, LU 141, LU 142, LU 166, LU 203, LU 302, LU 401
- Class group 2 “for automatic change detection”:
  - Land cover: LC 21, LC 31, LC 33, LC 41, LC 44, LC 47, LC 52, LC 61
  - Land use: LU 121, LU 125, LU 143, LU 144, LU 146, LU 164, LU 165, LU 221, LU 241, LU 242, LU 243, LU 301, LU 403, LU 423

The classes not mentioned above can be placed in the two groups after a performance assessment of the automatic change detection system during and after the development of the system.

7.5.3. Computational considerations

The computational time for the different steps, classification and change detection, is reported here below along with the workstation configuration. It should be noted that the reported times are rough estimations based on the computations performed within the test environment.

For the prediction of a sample point:

- Feature extraction: ~0.0015 s / patch (extracted from ResNet-50)
- Change detection prediction: ~0.0015 s / patch (SVM prediction)
- Classification prediction: ~0.003 s / patch (2x due to land cover and land use classifications)
- Total: ~0.006 s / patch

Workstation configuration overview:

- HP Z640 Workstation
- Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, 2201 Mhz
- 10 Core(s), 20 Logical Processor(s)
- Physical RAM memory: 16Gb
- GPU Nvidia GTX1080 Ti, RAM 11Gb

Software test environment overview:

- Windows 10 x64
- Python 2.7
- Toolboxes: Numpy, Scipy, OpenCV, Scikit-Learn, Scikit-Image, Microsoft Cognitive Toolkit, GDAL/OGR

Picterra Sàrl 2017, all rights reserved

8. Conclusions

8.1. Conclusions on the state-of-the-art review

The review of the Arealstatistik class catalogue could provide a view on the different types of classes, the extent and the criteria to be considered in order to discriminate them. This results in considering in average an extent of 50m around the sample point, while some classes could require less than a meter and some up to several hundreds of meters. Moreover, the classes could be grouped into three types: “single”, “mixed” or “multiples”, with the latter being a type of class made of a set of uniform elements (e.g. several trees, buildings or roads) which will benefit from dense labeling classification methods such as A4 “Dense land cover/use classification using aerial imagery and supervised deep features”.

The state-of-the-art review presented the most relevant machine learning and computer vision methods, which can serve either the land cover/use classification or change detection. Among the different learning strategies, supervised learning is the most promising in the BFS Arealstatistik context where training data is available in very large amounts. Semi-supervised learning exploiting unlabeled data could provide an increased accuracy but is generally less relevant when the training dataset is large enough (>1000 samples/class). Data augmentation, which consists in applying transformation such as rotation to the data to synthesize additional training data and increase the robustness of the classifier, cannot be generally applied in the Arealstatistik context, due to the interpretation area which is not centered around the sample point. Active learning is also less relevant when dealing with large training sets, but could adapt efficiently the trained classifier or change detector to the new survey data.

The type of features to be extracted from the image can be separated in two main families: the hand-crafted & unsupervised features, and the supervised features. The latter ones have recently shown tremendous improvements in the classification accuracies of land use/cover aerial scenes compared to the hand-crafted features. This came with the recent development of deep learning architectures specific to image data: Convolutional Neural Networks (CNNs). Networks pre-trained on millions of images from everyday scenes are available and can work extremely well on different type of images such as aerial images.

The general families of classifiers include these CNNs, the Random Forest and the Support Vector Machines. The latter as strong guarantees regarding bias and variance, however it lacks flexibility regarding the feature types and large number of training samples. The Random Forest are more adapted to large training sets and has a certain flexibility to handle multiple data sources. CNNs are combining the feature extraction and the classification and are well adapted to large training sets, however they lack flexibility to combine multiple data.

Datasets used for benchmarking classification methods are reviewed and their land cover/use classes are compared to BFS Arealstatistik classes. They can be linked to land cover and land use classes and even to certain specifications (e.g. church, school). Only few classes could not be directly linked such as harbor, beach, island, palace, desert, and cloud. In addition, datasets with aerial imagery and digital surface models used for benchmarking dense labeling classification methods are presented.

Regarding change detection, the state-of-the-art is more heterogeneous and without common datasets to benchmark methods on aerial image change detection. Certain hand-crafted features extracted from objects (image segments) or a set of pixels and classification methods allow to be robust to illumination changes, sensor specificities and varying viewpoints. Some early works have been published using deep learning features, which are robust to irrelevant changes such as illumination variations, shadows and present a high potential for reducing interpretation.

The different approaches, or families of methods, have been assessed with five criteria: “reduced interpretation”, “accuracy”, “robustness”, “sustainability” and “adaptation cost” scored between one and five.

Picterra Sàrl 2017, all rights reserved

The most promising approach is A3 “Land cover/use classification using aerial imagery and supervised deep features” providing a good balance between the five criteria. Then,the change detection approach C3 “Land cover/use change detection using aerial imagery and unsupervised deep features” follows, providing high potential for change detection. Last change detection algorithm based on traditional methods using pixel features shows lower potential when compared to first two methods. Nevertheless, it might be more adequate to control the missed detection rate.

Finally, the complementary data that could support the Arealstatistik, are among the new satellite constellations which provide very high spatial resolution and revisit rates (daily) and low costs, and the forthcoming high-altitude platform systems (haps) which will provide regional real-time monitoring. Moreover, the crowdsourced geospatial data (CGD) and volunteered geographic information (VGI) fields are getting a massive amount of data with the combination of social media with geo-tagged images, mapping from users as in OpenStreetMap and the large imaging campaigns such as google street view and google maps. Researchers are exploring the combined usage of crowdsourced data and images to improved classification models (e.g. OpenStreetMap to pretrain CNNs) or the comparisons of geo-tagged panoramas and aerial imagery for change detection purposes.

8.2.Conclusions on the experiments performed

From the outputs of the state-of-the-art review, a certain number of experiments were tested to evaluate quantitatively different approaches and assess the potential for automation of Arealstatistik interpretation. The tests exploited a set of aerial data from Swisstopo over 66’950 samples points for classification and 26’448 sample points for change detection.

The automatic classification performed with approach A3 based on deep features demonstrated that high accuracies could be reached on the most represented classes: LC 41, 21, 62 and LU 301, 421 401, 221. These classes represent around 60% of the entire point classes. Additionally, it could be shown that a larger number of land cover and land use classes were correctly classified when considering the top-3 class predicted. This could support the expert in performing a fast interpretation from the 3 recommended classes.

The automatic change detection approaches C1 and C3 were assessed and compared. Method C1 based on pixel features fails to detect changes accurately, while approach C3 based on deep features is suitable for performing change detection. The different tests highlighted that training models per class group could improve the detection performances and that again the major classes: LC 21, 41, 52 (67% of the points) and LU 221, 242, 301, 421 (69% of the points) can benefit from automatic change detection. Finally, two class groups could be defined: classes, which requires expert interpretation and classes, which are suitable for automatic change detection. It can be concluded that around 50% of the interpretation workload can be saved, as described in Figure 29:

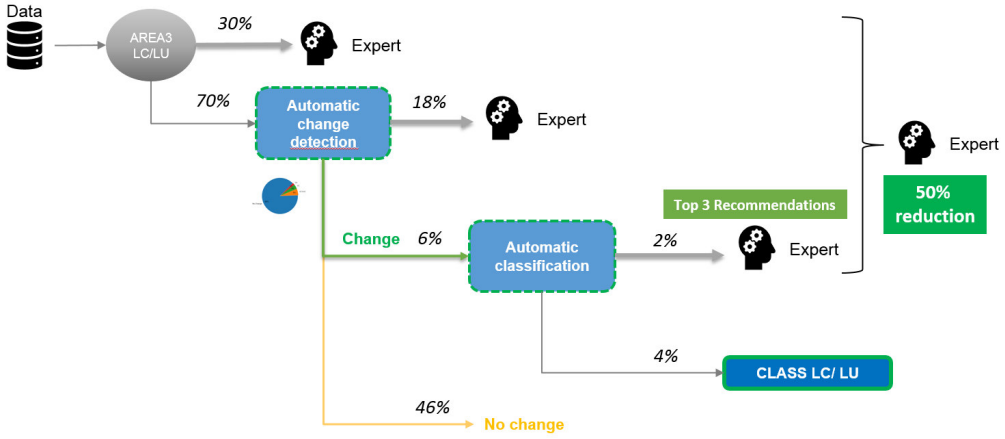


Figure 29 - General workflow and percentage of interpretation made by expert or automation. Change detection confidence level set to 75%, putting apart still 18% of the total data to the expert. Classification of the changes detected can be performed partially by the expert with top-3 recommendations (25%) and by the system automatically.

Picterra Sàrl 2017, all rights reserved

8.3. Recommendations

In the light of this study, the following elements can be highlighted and recommended:

- Machine learning specifically deep learning and computer vision tools are allowing high accuracy aerial image scene classification and can provide more automation in the interpretation of the Arealstatistik.
- All auxiliary data should be exploited within the classification models and the additional usage of the surface topology (Digital Surface Model) would improve the classification and change detection performances according to state-of-the-art. DSM could be derived from Federal and Cantonal LiDAR surveys (Swisstopo point clouds available have 5 to 10 points / m2) or directly using Socet GXP "Automatic Terrain Generation (ATG) from stereo imagery".
- Aerial data from the same season should be used between surveys and summer time is recommended for a better detection of trees and a better discrimination of classes. However, it is unlikely that the season could be controlled in terms of data and this type of variability will have to be learned by the system.
- Satellite data should be considered for an improved discrimination of time-varying classes (e.g. agricultural land) and the identification of non-permanent changes.
- A workflow combining expert interpretation and automation is recommended to let the system improve over time and maximize the accuracy of the interpretation with expert interpretation on focused difficult classes and machine learning predictions on large mainstream classes.

The following elements can be considered as potential extension of the Arealstatistik mandate:

- Additional points could be predicted (denser interpretation grid), exploiting the parallel processing on GPUs for class predictions. Here are rough estimates for three different amounts of additional points. A point takes ~0.006 s for being predicted, if a point is every:
  - 100m: current predictions are 4'128'500 ~ 7.5 hours.
  - 50m: 4x times the current predictions is 4x 4'128'500 = 16'514'000 ~ 30 hours.
  - 10m: 100x times the current predictions is 100x 4'128'500 = 412'850'000 ~ 30 days.
  - 5m: 400x times the current predictions is 400x 4'128'500 = 1'651'400'000 ~ 120 days.
- Coarser class groups could be predicted in a dense manner (every pixel) using dense labeling techniques without a computational burden, probably at the level of land cover class groups (1 to 6).
- Automatic detection of class specifications (e.g. schools, church, etc.) could be investigated and potentially provided with a certain accuracy tag

Picterra Sàrl 2017, all rights reserved

8.3.1. General system architecture

The envisioned general system architecture is depicted in Figure 30. First, the customized prediction models for both change detection and classification are trained using the image and auxiliary data acquired at time *t-1* (in this test scenario t-1 refers to AREA 3 data). This process is illustrated with gray arrows. After the trained model, considering the defined quality thresholds, is created and saved, the image and auxiliary data acquired at time *t* are processed (follow the black arrows) as follows:

1. Based on the performance of created model on data from time *t-1*, all data points are generally divided to two groups ("Complex Analysis"), one for direct expert interpretation and second for automatic prediction and labeling.
2. The aerial and auxiliary data (GWR, TLM, etc.) are preprocessed to match input data format suitable for selected model.
3. A suitable model is:
   3.1. trained or fine-tuned.
   3.2. selected among available existing models.
4. Model serving unit selects model suitable for required tasks (change detection, LC/LU classification). Prediction serving unit performs prediction on unlabeled data and returns a predicted class for a given image data.
5. Predicted classes on subset of selected data points are subjects to quality control. Based on the predefined quality thresholds, the output is either stored to data point database as a new label for new data or a request to generate top 3 recommendation, or direct expert interpretation is generated.

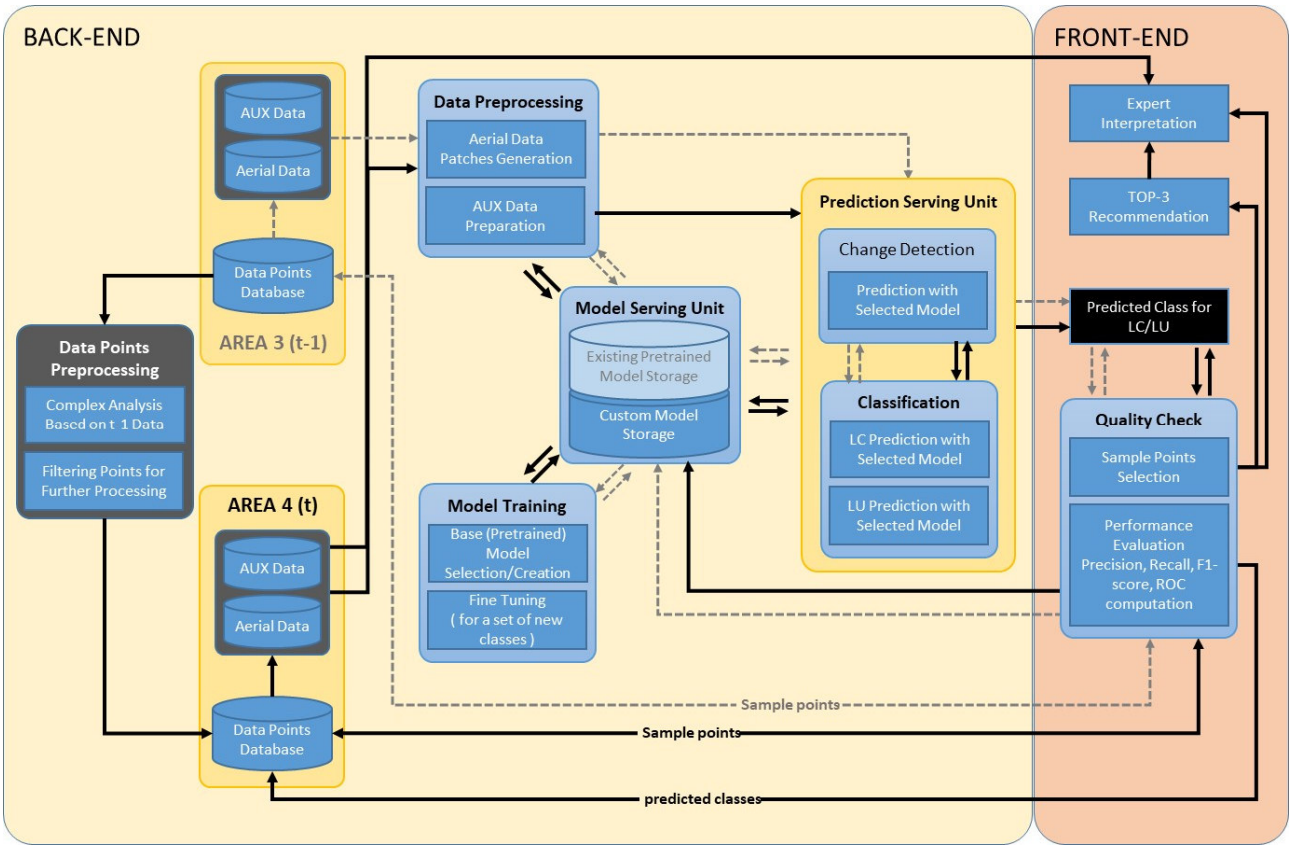


Figure 30 – General system architecture

Picterra Sàrl 2017, all rights reserved

8.3.2. System development planning

Effort to implement the entire system into BFS Arealstatistik operational workflow is envisioned as illustrated in Figure 31. The whole project can be divided into 5 main parts, which are all covered by continuous project management. While the Figure 31 shows the individual engineering steps in a sequential manner, Figure 32 presents the Gantt-chart showing the possible work parallelization under an assumption of hiring 2.5 project engineers to develop and implement the automatic prediction system.

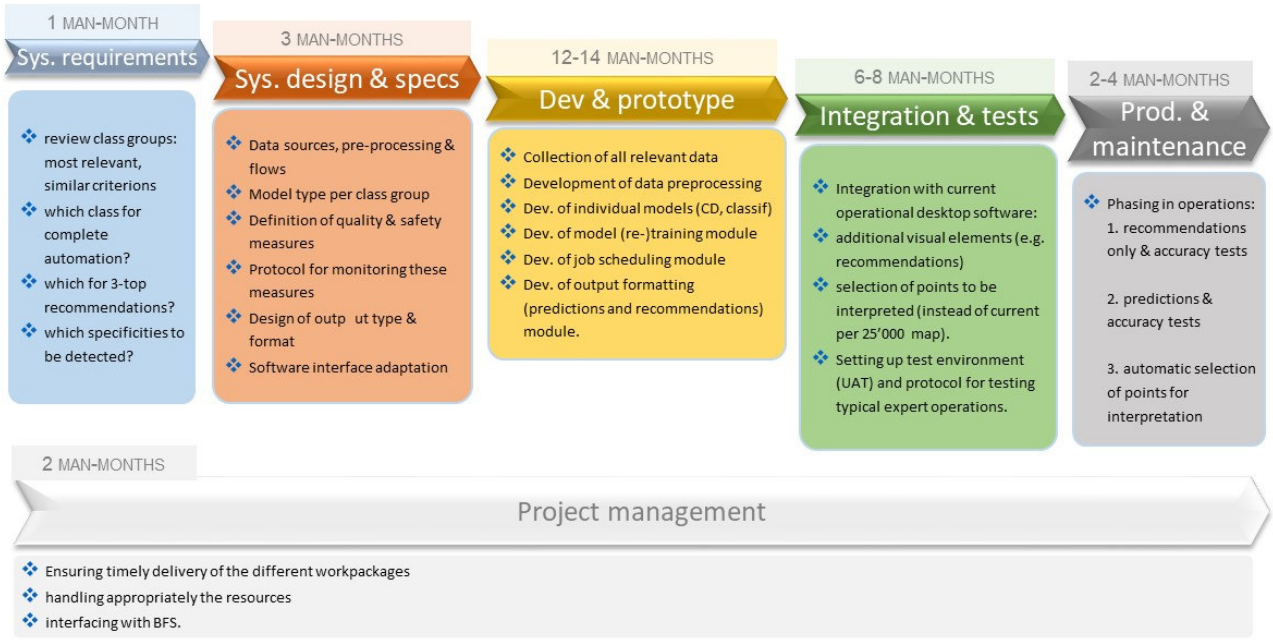


Figure 31 – Estimated implementation workload

The effort for implementation into BFS Arealstatistik operational workflow with respect to the minimum duration of each workpackage and the part of work foreseen to be externalized is presented here below:

Workpackage	Man-month(s)	Minimum duration	Part externalized
System requirements	1	1	10%
System design and specifications	3	2	40%
Development and prototyping	12-14	6	60% (dev. of data preprocessing, individual models, model re-training, jobs scheduling)
Integration & tests in operational workflow	6-8	6	40%
Deployment in production and maintenance	2-4	2	20%
Project management	2	2	30%
Total	**26-32**	**19**	**~48%**

Picterra Sàrl 2017, all rights reserved

ID	Work Package	Start	End	Duration	Year 1											
					M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
1	**Project Management**	M1	M12	2												
2	**Sys. Requirements**	M1	M1	1												
3	**Sys. design & specs**	M1	M2	2												
4	**Dev & prototype**	M2	M7	6												
	Collection of all relevant data															
	Development of data preprocessing															
	Dev. of individual models (CD, classif)															
	Dev. of model (re-)training module															
	Dev. of job scheduling module															
	Dev. of output formatting (predictions and recommendations) module.															
5	**Integration & tests**	M6	M11	6												
	Integration with current operational desktop software:															
	additional visual elements (e.g. recommendations)															
	selection of points to be interpreted (instead of current per 25'000 map).															
	Setting up test environment (UAT) and protocol for testing typical expert operations.															
6	**Prod. & maintenance**	M10	M12	3												
	Total man-months requirement per actual calendar month				2	2	2	2	2	4	5	3	3	2	2	1
					0.5			0.5				0.5				0.5
	Total MM		**30**		project engineering											
			2		project management											

Figure 32 – Gantt chart illustrating the system development planning and work parallelization

The foreseen development would take approximatively 12 months with around 48% of the work externalized, mainly on the development and prototyping work (data preprocessing, individual models, model re-training and jobs scheduling).

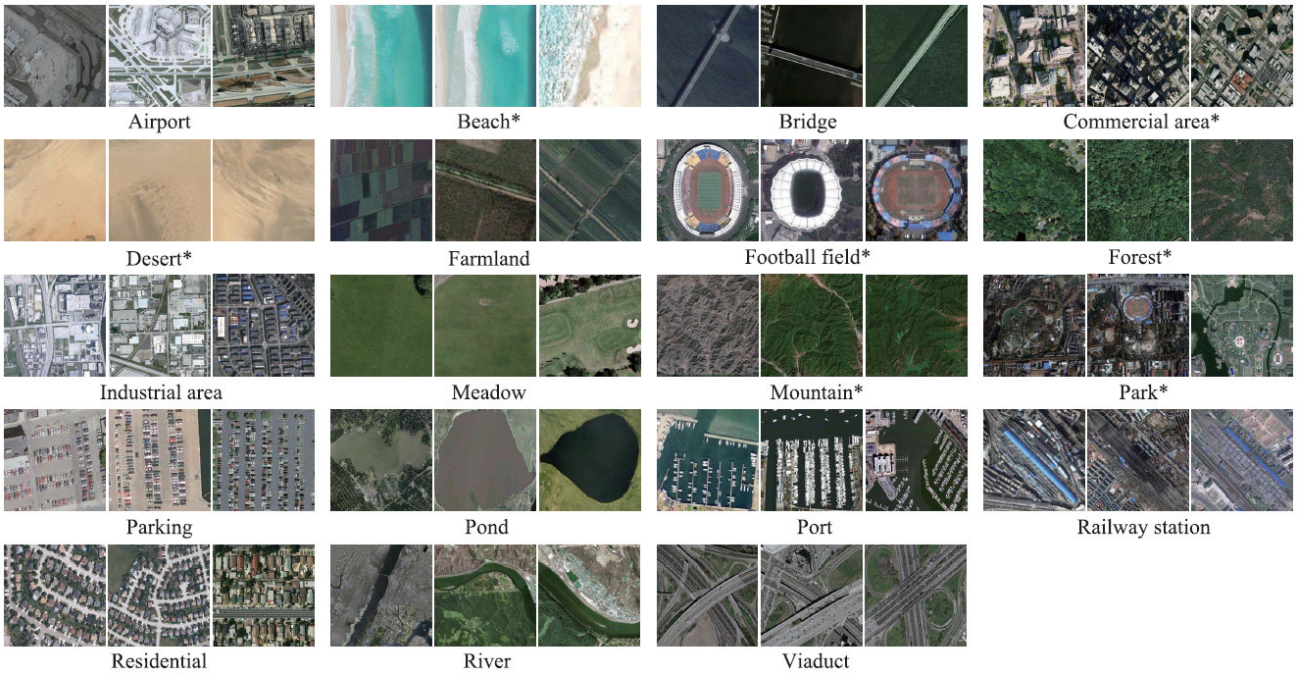
Picterra Sàrl 2017, all rights reserved

Appendix

Image samples of the different benchmark datasets.

UC MERCED Dataset

WHU-RS19 Dataset



Picterra Sàrl 2017, all rights reserved

SIRI-WHU Dataset

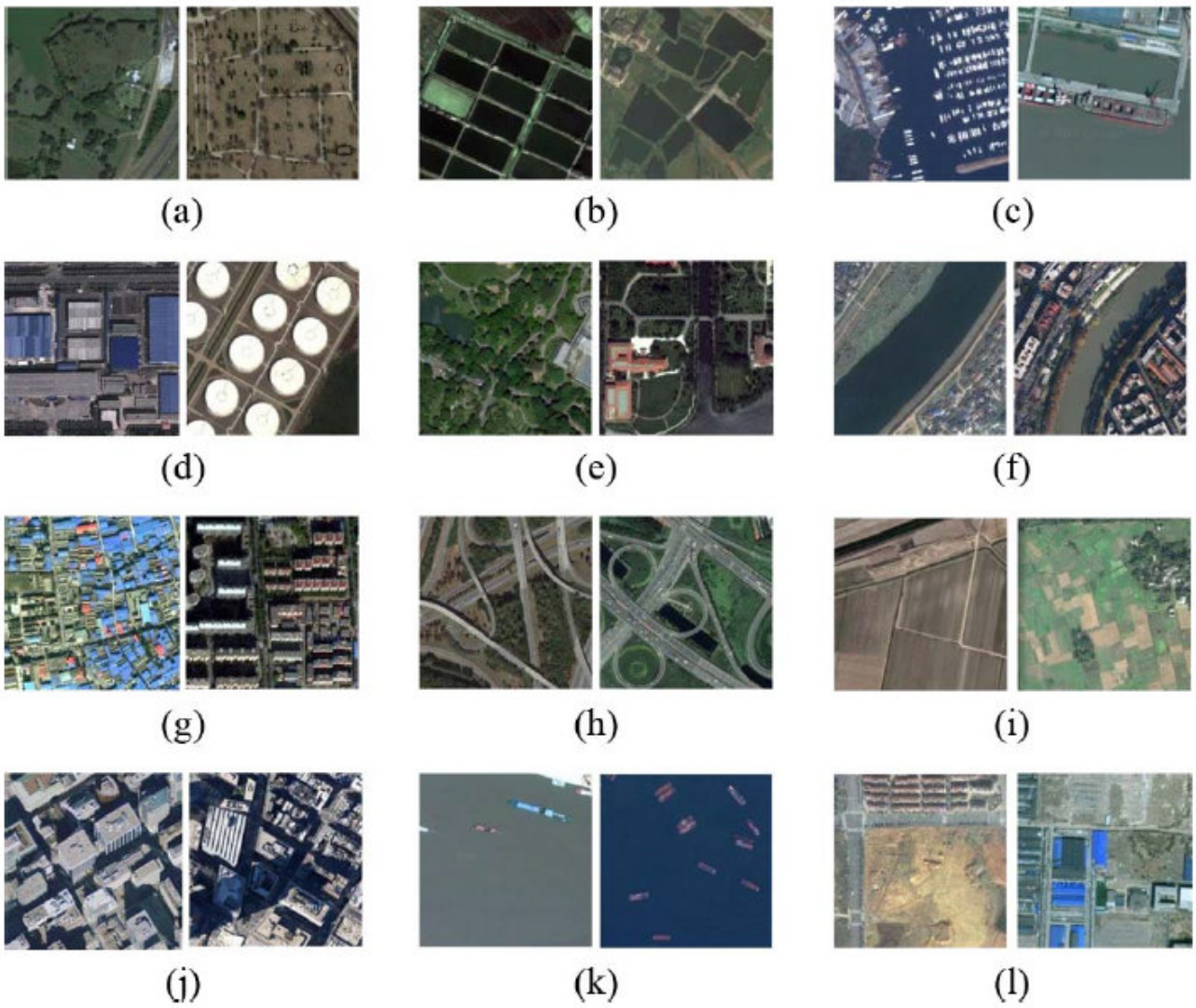


Fig. 9. Google data set. (a)–(l) Meadow, pond, harbor, industrial, park, river, residential, overpass, agriculture, commercial, water, and idle land, respectively.

Picterra Sàrl 2017, all rights reserved

RSSCN7 Dataset

Grass. Farm. Industry. River. Forest. Resident. Parking.

RSC11 Dataset

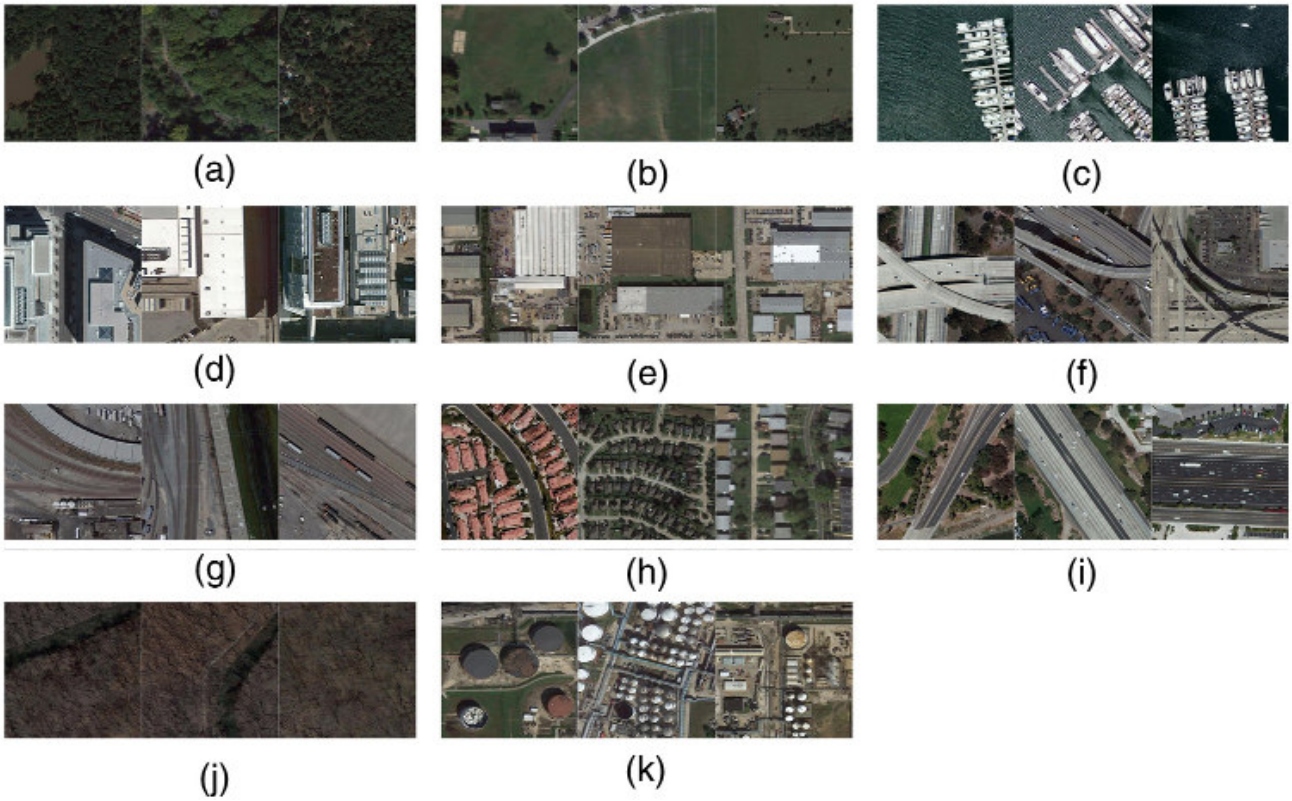
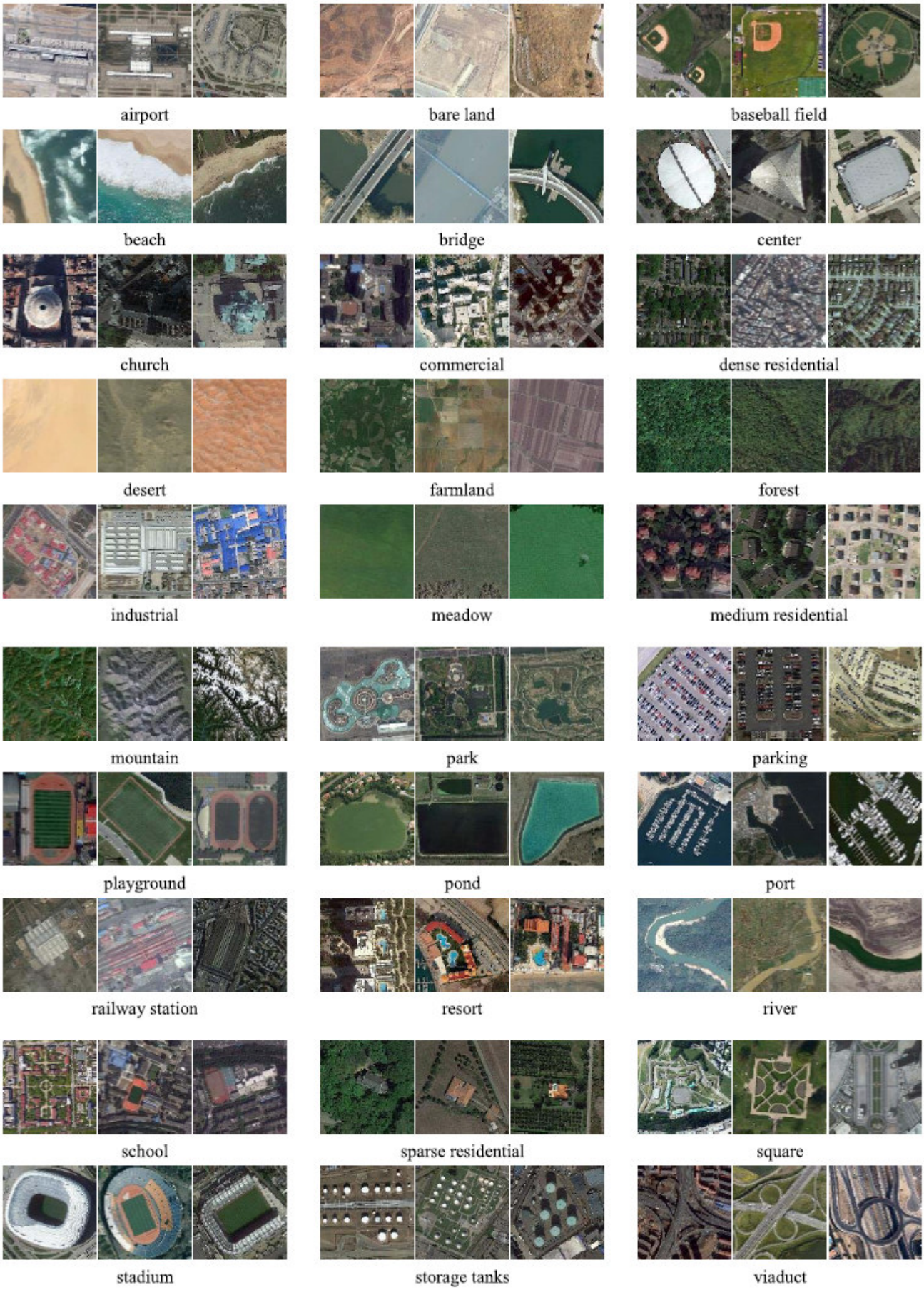


Fig. 7 Examples of dataset 2: (a) dense forest, (b) grassland, (c) harbor, (d) high buildings, (e) low buildings, (f) overpass, (g) railway, (h) residential area, (i) roads, (j) sparse forest, and (k) storage tanks.

Picterra Sàrl 2017, all rights reserved

AID Dataset



Picterra Sàrl 2017, all rights reserved

NWPU - RESISC45 Dataset

Picterra Sàrl 2017, all rights reserved

RESNET-50 CONVOLUTIONAL NEURAL NETWORK

Here are presented some information and references related to the type of CNN Networks used in the experiments of this study.

The “ResNet50” CNN model is a “Deep Residual Networks» (ResNets), a neural network structured in a different manner than traditional CNNs. The idea is to bring back some original information after having performed the typical convolutions. This architecture allowed to win the ImageNet and COCO competitions in 2015. The Residual Networks and their residual blocks, schematized in Figure 33, are structure as follows:

- Input data is sent to a convolutional layer
- then followed by a “batch normalization”, which normalizes these outputs over a certain number of convolutional windows
- then a rectified linear unit (ReLU) is applied, which preserves only positive values (and set negative ones to 0).
- After another set of convolution and batch normalization is performed.
- A set of skip connections are bypassing or shortcutting these transformations and added to the outputs before being fed to a ReLU.

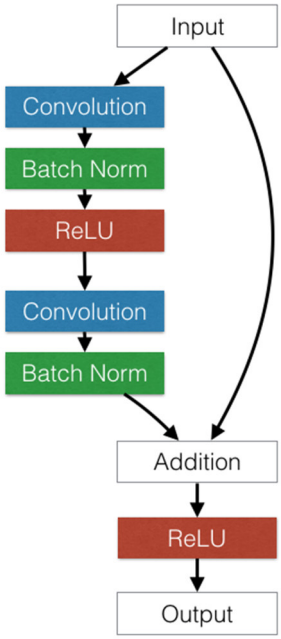


Figure 33 - Schematic of Residual block from http://torch.ch/blog/2016/02/04/resnets.html

The ResNet-50 model is made of 50 of these Residual blocks stacked one after the other.

The ResNet-50 model weights used in the study experiments were obtained from a pre-training of the model on:

- ImageNet dataset of millions of images annotated within a set of 1000 classes.
- Stochastic Gradient Descent (SGD) and momentum training strategy were used.

Some references related to ResNets:

- He, Kaiming, et al. “Deep Residual Learning for Image Recognition.” *arXiv preprint* arXiv:1512.03385 (2015).
- Ioffe, Sergey, and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” *arXiv preprint* arXiv:1502.03167 (2015).

Picterra Sàrl 2017, all rights reserved