Increasing the Effectiveness of Active Learning: Introducing Artificial Data Generation in Active Learning for Land Use/Land Cover Classification

Joao Fonseca[1], Georgios Douzas[1], Fernando Bacao[1*]

[1]NOVA Information Management School, Universidade Nova de Lisboa

*Corresponding Author

Postal Address: NOVA Information Management School, Campus de Campolide, 1070-312 Lisboa, Portugal

Telephone: +351 21 382 8610

TODO TODO

1 Introduction

The technological development of air and space borne sensors, as well as the increasing number of remote sensing missions have allowed the continuous collection of large amounts of high quality remotely sensed data. This data is often composed of multi and hyper spectral satellite imagery, essential for numerous applications, such as Land Use/Land Cover (LULC) change detection, ecosystem management [Nagai et al., 2020], agricultural management [Huang et al., 2018], water resource management [Wang and Xie, 2018], forest management, and urban monitoring [Khatami et al., 2016]. However, updating a LULC map is still a challenging task [Gavade and Rajpurohit, 2019, Wulder et al., 2018]. They can be updated using either one of the following strategies:

1. Photo-interpreted. Consists of evaluating a patch's LULC class based on orthophoto and satellite image interpretation [Costa et al., 2020]. This method guarantees a decent level of accuracy, as it is dependent on the interpreter's expertise and human error. Typically, it is an expensive, time-consuming task that requires the expertise of a photo-interpreter. This task is also frequently applied to obtain ground-truth labels for training and/or validating Machine Learning (ML) algorithms for related tasks [Vermote et al., 2020, Costantino et al., 2020].

2. Automated mapping. It is based on the usage of a ML method or a combination of methods in order to obtain an updated LULC map. The development of a reliable automated method is still a challenge among the ML and remote sensing community, since the efficacy of existing methods vary across applications and geographical areas [Gavade and Rajpurohit, 2019]. Typically, this method requires the existence of ground-truth data, which is frequently outdated or nonexistent for the required time frame [Nagai et al., 2020]. On the other hand, employing a ML method provides readily available and relatively inexpensive LULC maps. The increasing quality of state-of-the-art classification methods have motivated the application and adaptation of these methods in this domain [Maxwell et al., 2018].

3. Hybrid approaches. They employ photo-interpreted data to augment the training dataset and improve the quality of automated mapping [Růžička et al., 2020]. It attempts to accelerate the photo-interpretation process by selecting a smaller sample of the study area to be interpreted. The goal is to minimize the inaccuracies found in the LULC map by supplying high-quality ground-truth data to the automated method. The final (photo-interpreted) dataset consists of only the most informative samples, i.e., patches that are typically difficult to classify for a traditional automated mapping method [Liu et al., 2020].

The latter method is best know as Active Learning (AL). It is especially useful whenever there is an absence of ground-truth data and/or the mapping region does not contain updated LULC maps [Su et al., 2020]. In a context of limited sample-collection budget, the collection of the most informative samples capable of optimally increasing the classification accuracy of a LULC map is of particular interest [Su et al., 2020]. AL attempts to minimize the human-computer interaction involved in photo-interpretation by selecting the data points to include into the classification process. These data points are selected based on an uncertainty measure and represent the points close to the decision borders. Afterwards, they are passed on for photo-interpretation and added to the training dataset, while the points with the lowest uncertainty values are ignored for photo-interpretation and classification. This process is iterated until a convergence criterion is reached [Pasolli et al., 2016].

The relevant work developed within AL is described in detail in Section 2. The research attempts to address some of the challenges found in AL, mainly inherited from automated and photo-interpreted mapping: mapping inaccuracies and time consuming human-computer interactions. Mapping inaccuracies have different sources:

1. Human error. The involvement of photo-interpreters in the data labeling step carries an additional risk to the creation of LULC patches. The minimum mapping unit being considered, as well as the quality of the orthophotos and satellite images being used, are some of the factors that may lead to the overlooking of small-area LULC patches and label-noisy training data [Pelletier et al., 2017].

2. High-dimensional datasets. The amount of bands (i.e., features) present in multi and hyper spectral images introduce an increased level of complexity in the classification step [Stromann et al., 2020]. These datasets are often prone to the Hughes phenomenon, also known as the curse of dimensionality.

3. Class separability. Producing an LULC map considering classes with similar spectral signatures makes them difficult to separate [Alonso-Sarria et al., 2019]. A lower pixel resolution of the satellite images may also imply mixed-class pixels, which may lead to both lower class separability as well as higher risk of human error.

4. Existence of rare land cover classes. The varying morphologies of different geographical regions naturally implies an uneven distribution of land cover classes [Feng et al., 2018]. This is particularly

relevant in the context of AL: the data selection method is based on a given uncertainty measure over data points whose class label is unknown. Consequently, AL's iterative process of data selection may disregard wrongly classified land cover areas belonging to a minority class.

Research developed in the field of Active Learning typically focus on the reduction of human error by minimizing the human interaction with the process through the development of more efficient choosers and selection criteria within the generally accepted AL framework. Concurrently, the problem of rare land cover classes is rarely addressed. This is a frequent problem in the ML community, known as the Imbalanced Learning problem. This problem exists whenever there is an uneven between-class distribution in the dataset [Chawla et al., 2004]. Specifically, most classifiers are designed to optimize metrics such as overall accuracy, which are designed to work primarily with balanced datasets. Consequently, these metrics tend to introduce a bias towards the majority class by attributing an importance to each class proportional to its relative frequency [Maxwell et al., 2018]. As an example, such a classifier could achieve an overall accuracy of 99% on a binary dataset where the minority class represents 1% of the overall dataset and still be deemed useless. A number of methods have been developed to deal with this problem. They can be categorized into three different types of approaches [Fernández et al., 2013, Kaur et al., 2019]. Cost-sensitive solutions perform changes to the cost matrix in the learning phase. Algorithmic level solutions modify specific classifiers to reinforce learning on minority classes. Resampling solutions modify the dataset by removing majority samples and/or generating artificial minority samples. The latter is independent from the context and can be used alongside any classifier. We will focus on artificial data generation techniques, presented in Section 3.

In this paper, we propose a novel AL framework to address two limitations commonly found in the literature: minimize human-computer interaction and reduce the class imbalance bias. This is done with the introduction of an additional component in the iterative AL procedure (the generator), used to generate artificial data to both balance and augment the training dataset. The introduction of this component is expected to reduce the number of iterations required until convergence of the predictor's quality.

This paper is organized as follows: Section 1 exposes the problem and its context, Sections 2 and 3 describe the state of the art in AL and Oversampling techniques, Section 4 exposes the proposed method, Section 5 covers the datasets, evaluation metrics, ML classifiers and experimental procedure, Section 6 presents the results and statistical analyses and Section 7 reports the conclusions drawn from our findings.

2 Active Learning Approaches

AL is used as the general definition of frameworks aiming to train a learning system in multiple steps, where a set of new data points are chosen and added to the training dataset each time [Růžička et al., 2020]. Typically, an AL framework is composed of 10 elements, out of which 4 are datasets, 2 are queries or estimations regarding the target class labels and 4 are components responsible for performing the tasks involved in AL [Sverchkov and Craven, 2017, Su et al., 2020, Růžička et al., 2020]:

1. Data source. In the context of LULC classification, the data source is usually a hyper/multi-spectral image, a Synthetic-aperture radar (SAR) image, or a composite image.

2. Unlabeled dataset. Consists of a sample of the original data source. It is used in combination with the chooser and the selection criterion to retrieve uncertainty estimates on each iteration.

3. Initial training sample. It is a small sample of the unlabeled dataset, used to initiate the first AL iteration. The size of the initial training sample normally varies between no observations at all and 10% [Li and Guo, 2013].

4. Augmented training dataset. This dataset is the concatenation of the labeled initial training sample along with the datasets labeled by the oracle in past iterations.

5. Uncertainty map. The dataset containing the highest uncertainty points/patches to be labeled by the oracle.

6. Oracle. An external entity to which the uncertainty map is presented to. The oracle is responsible for annotating unlabeled samples to be added to the augmented dataset. In remote sensing, the oracle is typically a photo-interpreter, as is the case in [Li et al., 2020]. Some of the research also refers to the oracle as the *supervisor* [Su et al., 2020, Shrivastava and Pradhan, 2021].

7. Chooser. Produces the class probabilities for each unlabeled sample. This is a classifier trained using the augmented dataset. It is used to estimate the class probabilities for each sample over the unlabeled dataset.

8. Selection criterion. It quantifies the chooser's uncertainty level for each sample belonging to the unlabeled dataset. It is typically based on the class probabilities assigned by the chooser. In some situations, the chooser and the selection criterion are grouped together under the concept *acquisition function* [Růžička et al., 2020] or *query function* [Su et al., 2020]. Some of the literature refers to the selection criterion by using the concept *sampling scheme* [Liu et al., 2020].

9. Predictor. The classifier used to infer the land cover classes for the final output map. Once a stopping criterion is met, the classifier is trained using the augmented dataset and the LULC classes are inferred from the data source.

10. Prediction output. In the context of LULC classification, the prediction output is the estimated LULC map raster.

Figure 1 schematizes the steps involved in a complete AL iteration. For a better context within the remote sensing domain, the prediction output is identified as the LULC map. This framework starts by collecting unlabeled data from the original data source. It is used to generate a random initial training sample and is labeled by the oracle. In practical applications, the oracle is frequently a group of photo-interpreters [Kottke et al., 2017]. The chooser is trained on the resulting dataset and is used to predict the class probabilities on the unlabeled dataset. They are fed into a selection criterion to estimate the prediction's uncertainty, out of which the samples with the highest uncertainty will be selected. This calculation is motivated by the absence of labels in the uncertainty dataset. Therefore, it is impossible to estimate the prediction's accuracy in a real case scenario. The iteration is completed when the selected points are tagged by the oracle and added to the training dataset (i.e., the augmented dataset).

A common challenge found in AL tasks is ensuring the consistency of AL over different initializations [Kottke et al., 2017]. There two factors involved in this phenomenon. On the one hand, the implementation of the same method over different initializations may result in significantly different accuracy curves. On the other hand, the lack of a robust selection criterion and/or chooser may also result in results' inconsistency across initializations. This phenomenon was observed and documented in a LULC classification context in [Tuia et al., 2011].

Selecting an efficient selection criterion is particularly important to find the samples closest to the decision border (i.e., samples difficult to classify) [Shrivastava and Pradhan, 2021]. Therefore, most of AL related studies focus on the design of the query/acquisition function [Su et al., 2020].

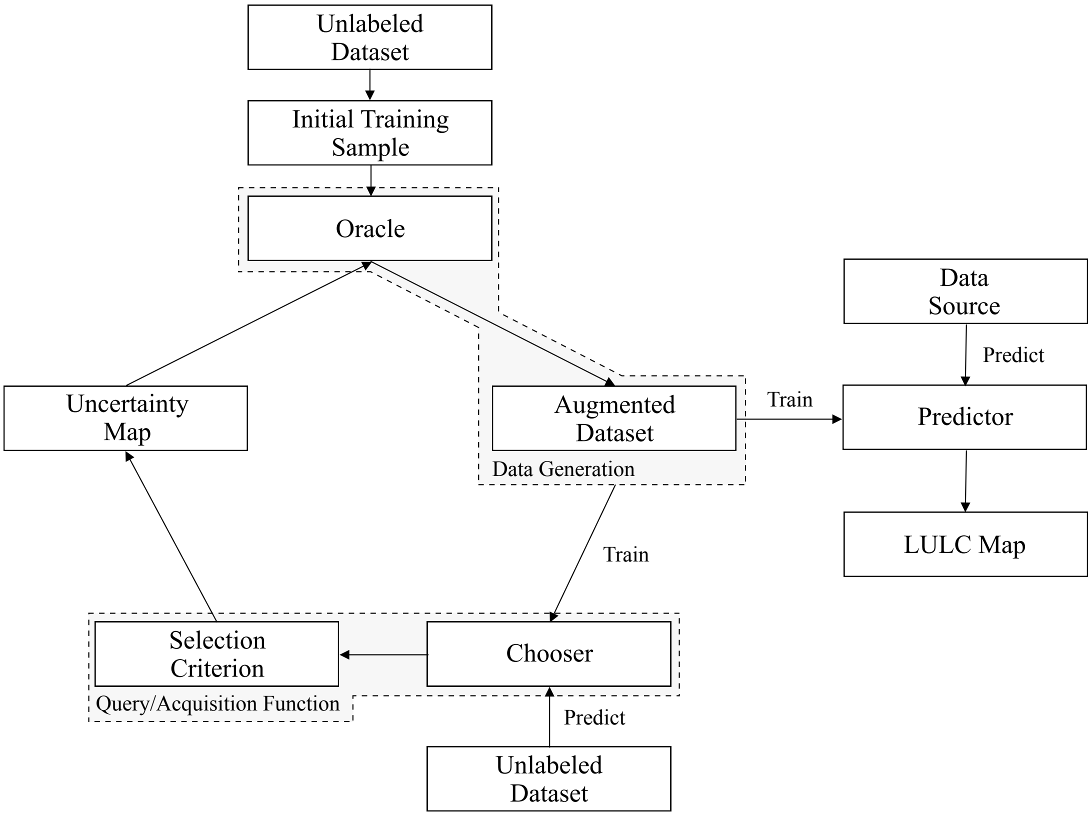

Figure 1: Typical AL framework.

2.1 Non-informed selection criteria

Only one non-informed selection criterion was found. Random sampling selects unlabeled samples without considering any external information produced by the chooser. Since the method for selecting the unlabeled samples is random, this method disregards the usage of a chooser and is comparatively worse than any other selection criterion. Although, random sampling is still a powerful baseline method [Cawley, 2011]. Generally, different AL initializations return high performance variability [Kottke et al., 2017]. When this happens, the analysis of the mean performances over multiple repetitions is not of interest. Instead, it is preferable to do pairwise comparison of different methods along with their corresponding variances.

2.2 Ensemble-based selection criteria

Ensemble disagreement is based on the class predictions of a set of classifiers. The disagreement between all the predictions for a given observation is a common measure for uncertainty, although computationally inefficient [Růžička et al., 2020, Pasolli et al., 2016]. This method was implemented successfully for complex applications like deep active learning [Růžička et al., 2020].

Multiview [Muslea et al., 2006] consists on the training of multiple independent classifiers using different views, which correspond to the selection of subsets of features or observations in the dataset. Therefore, it can be seen as a bootstrap aggregation (bagging) ensemble disagreement method. The set of classifications over a single observation is used to calculate the maximum disagreement metric, given by the number of votes assigned to the most frequent class [Shrivastava and Pradhan, 2021]. A lower value for this metric means a higher classification uncertainty. Multiview-based maximum disagreement has been successfully applied to hyper-spectral image classification in [Di and Crawford, 2012] and [Zhou et al., 2014].

An adapted disagreement criterion for an ensemble of k-nearest neighbors has been proposed in [Pasolli et al., 2016]. This method employs a k-nearest neighbors classifier and computes an instance's classification uncertainty based on the neighbors' class frequency using the maximum disagreement metric over varying values for k. As a result, this method is comparable to computing the dominant class' score over a weighted k-nearest neighbors classifier. This method was also used on a multimetric active learning framework [Zhang et al., 2016].

Another relevant ensemble-based selection criterion is the binary random forest-based query model [Su et al., 2020]. This method employs a one-versus-one ensemble method to demonstrate an efficient data selection method using the estimated probability of each binary random forest and determining the classification uncertainty based on the probabilities closest to 0.5 (i.e., the least separable pair of classes are used to determine the uncertainty value). Although, this study fails to compare the proposed method with other benchmark methods, such as random sampling.

2.3 Entropy-based criteria

A number of contributions have focused on entropy-based querying. The application of entropy is common among active deep learning applications [Aghdam et al., 2019], where the training of an ensemble of classifiers is often too expensive. The measure of entropy is formulated as follows:

$$H(x_i) = \sum_{\omega=1}^{N_i} p(y_i^* = \omega | x_i) \log_2 [p(y_i^* = \omega | x_i)] \tag{1}$$

The measurement of entropy H is based on the observed probability $p(y_i^* = \omega|x_i)$ of obtaining class ω as the predicted class label y_i^*, where N_i is the number classes predicted for observation x_i.

Entropy query-by-bagging (EQB), also defined as maximum entropy [Liu et al., 2020], is an ensemble approach of the entropy selection criterion, originally proposed in [Tuia et al., 2009]. This strategy uses the set of predictions produced by the ensemble classifier to calculate those many entropy measurements. The estimated uncertainty measure for one sample is given by the maximum entropy within that set. EQB was observed to be an efficient selection criterion. Specifically, [Shrivastava and Pradhan, 2021] applied EQB on hyper-spectral remote sensing imagery using Support Vector Machines (SVM) and Extreme Learning Machines (ELM) as choosers, achieving optimal results when combining EQB with ELM. Another study successfully implemented this method on an active deep learning application [Liu et al., 2020]. Another study improved over this method with a normalized EQB selection criterion [Copa et al., 2010].

2.4 Other relevant criteria

Margin Sampling is a SVM-specific criterion, based on the distance of a given point to the SVM's decision boundary [Shrivastava and Pradhan, 2021]. This method is less popular than the remaining methods because it is limited to one type of chooser (SVMs). One extension of this method is the multiclass level uncertainty [Shrivastava and Pradhan, 2021], calculated by subtracting the observation's distance to the decision boundaries of the two most probable classes [Demir et al., 2011].

The Mutual Information-based (MI) criterion selects the new training samples by maximizing the mutual information between the classifier and class labels in order to select samples from regions that are difficult to classify. Although this method is commonly used, it is frequently outperformed by the breaking ties selection criterion [Li et al., 2011, Liu et al., 2018].

The breaking ties (BT) selection criterion was originally introduced in [Luo et al., 2003]. It is formulated as follows:

$$BT(x_i) = \arg\min_{x_i, i \in S_u} \{\max_{\omega \in N} p(y_i^* = \omega | x_i) - \max_{\omega \in N \setminus \{\omega^+\}} p(y_i^* = \omega | x_i)\} \tag{2}$$

Which is the subtraction of the probabilities of the two most likely classes. Another related method is Modified Breaking Ties scheme (MBT), which aims at finding the samples containing the largest probabilities for the dominant class [Liu et al., 2018, Li et al., 2013]

The last type of selection criteria identified is the loss prediction method [Yoo and Kweon, 2019]. This method replaces the selection criterion with a second predictor whose goal is to estimate the chooser's loss for a given prediction. This allows the new classifier to estimate the prediction loss on unlabeled observations and select the ones with the highest predicted loss.

Some of the literature fail to specify the strategy employed, although inferring it is generally intuitive. For example, [Ertekin et al., 2007] successfully used AL to address the imbalanced learning problem. They employed an ensemble of SVMs as the chooser and predictor to employ an ensemble-based selection criterion. All of the research found related to this topic focused on the improvement of AL through modifications on the selection criterion, chooser or predictor. None of these publications proposed significant variations to the typical AL framework.

3 Artificial Data Generation Approaches

The generation of artificial data is a common approach to address imbalanced learning tasks [Kaur et al., 2019], as well as improving the effectiveness of supervised learning tasks [DeVries and Taylor, 2017]. In recent years some sophisticated data generation approaches were found. Although, the scope of this work is to propose the integration of a generator within the AL framework. Due to the complexity and computational cost of network-based approaches (e.g., Generative Adversarial Networks), we will only focus on heuristic data generation approaches.

Heuristic data resampling methods employ local and/or global information to generate new, relevant, non-duplicated instances. These methods are most commonly used to populate minority classes and balance the between-class distribution of a dataset. The Synthetic Minority Oversampling Technique (SMOTE) [Chawla et al., 2002] was the first heuristic oversampling algorithm to be proposed. The simplicity and effectiveness of this method contributes to its prevailing popularity. It generates a new instance $\overrightarrow{z}$ through a linear interpolation of a randomly selected minority-class observation $\overrightarrow{x}$ and one of its randomly selected k-nearest neighbors $\overrightarrow{y}$ such that $\overrightarrow{z} = \alpha\overrightarrow{x} + (1-\alpha)\overrightarrow{y}$ where α is a random float between 0 and 1, as shown in Figure 2.

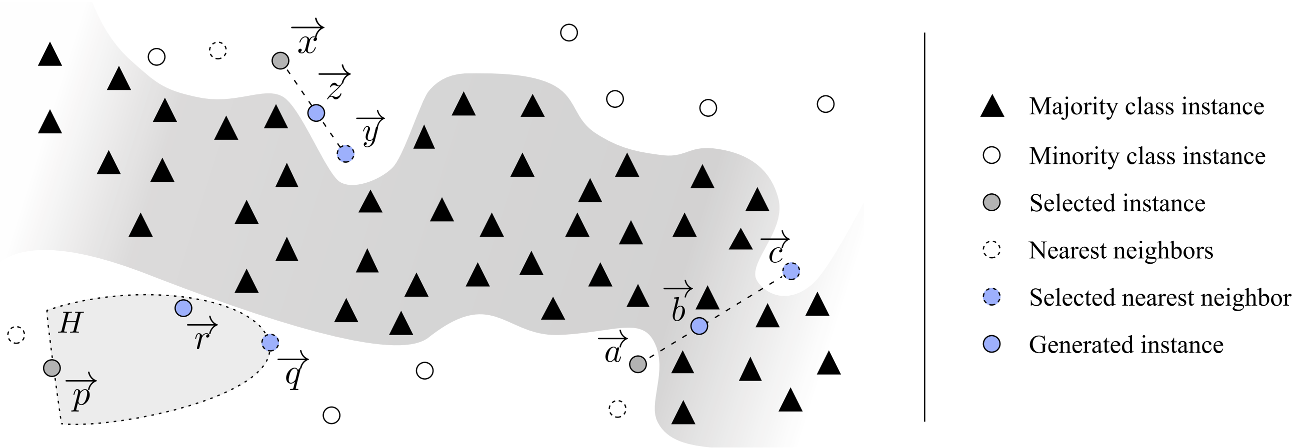

Figure 2: Examples of SMOTE and G-SMOTE generation process.

The implementation of SMOTE for LULC classification tasks has been found to improve the quality of the predictors used [Jozdani et al., 2019, Bogner et al., 2018]. Despite its popularity, its drawbacks motivated the development of other oversampling methods [Douzas and Bacao, 2019]:

1. Generation of noisy samples due to the selection of k-nearest neighbors and initial observation. The selection of a sample and/or neighboring sample located inside a majority class region may produce artificial samples within that region and amplify noisy data. Borderline-SMOTE [Han et al., 2005] is a modification of SMOTE in which only the minority examples near the borderline are over-sampled. This method avoids the generation of noisy samples by disregarding minority samples located in a majority class region as well as samples distant from the decision borders. The Adaptive Synthetic Sampling approach (ADASYN) [He et al., 2008] uses a density distribution ratio to address this limitation and focus the artificial data generation on minority class regions that are more difficult to classify.

2. Generation of noisy instances due to the use of observations from two different minority class clusters. Choosing a minority sample $\overrightarrow{d}$ and one of its nearest neighbors $\overrightarrow{b}$ belonging to a different minority cluster may lead to the generation of a sample $\overrightarrow{c}$ located within the two classes, as shown in Figure 2. K-means SMOTE [Douzas et al., 2018] and Self-Organizing map oversampling (SOMO) [Douzas and Bacao, 2017] reduce this effect by oversampling minority class samples within the same clusters.

3. Generation of nearly duplicated instances. The linear interpolation of parent samples that are close to each other produces an artificial sample with similar properties as its parents. Geometric SMOTE (G-SMOTE) [Douzas and Bacao, 2019] introduces a modification of the SMOTE algorithm in the data generation mechanism to produce artificial samples with higher variability.

The G-SMOTE algorithm is introduced as a generalization of the vanilla SMOTE. Instead of generating artificial data as a linear combination of the parent samples, it is done within a deformed, truncated hyper-spheroid. G-SMOTE generates an artificial sample $\overrightarrow{r}$ within a hyper-spheroid H, formed by selecting a minority sample $\overrightarrow{p}$ and one of its nearest neighbors $\overrightarrow{q}$, as shown in Figure 2. The truncation and deformation parameters define the shape of the spheroid's geometry. The method also modifies the selection strategy for the k-nearest neighbors, accepting the generation of artificial samples using observations from different classes. G-SMOTE has shown superior performance when compared with other oversampling methods for LULC classification tasks, regardless of the classifier used [Douzas et al., 2019].

4 Proposed method

Within the literature identified, most of the work developed in the AL domain revolved around improving the quality of the chooser, predictor and/or selection criterion. Although these methods allow earlier convergence of the AL iterative process, the impact of these methods are only observed between iterations. Consequently, none of these contributions focused on the definition of decision borders within iterations. The method proposed in this paper modifies the AL framework by introducing an artificial data generation step within AL's iterative process. We define this component as the generator, as defined in Figure 3.

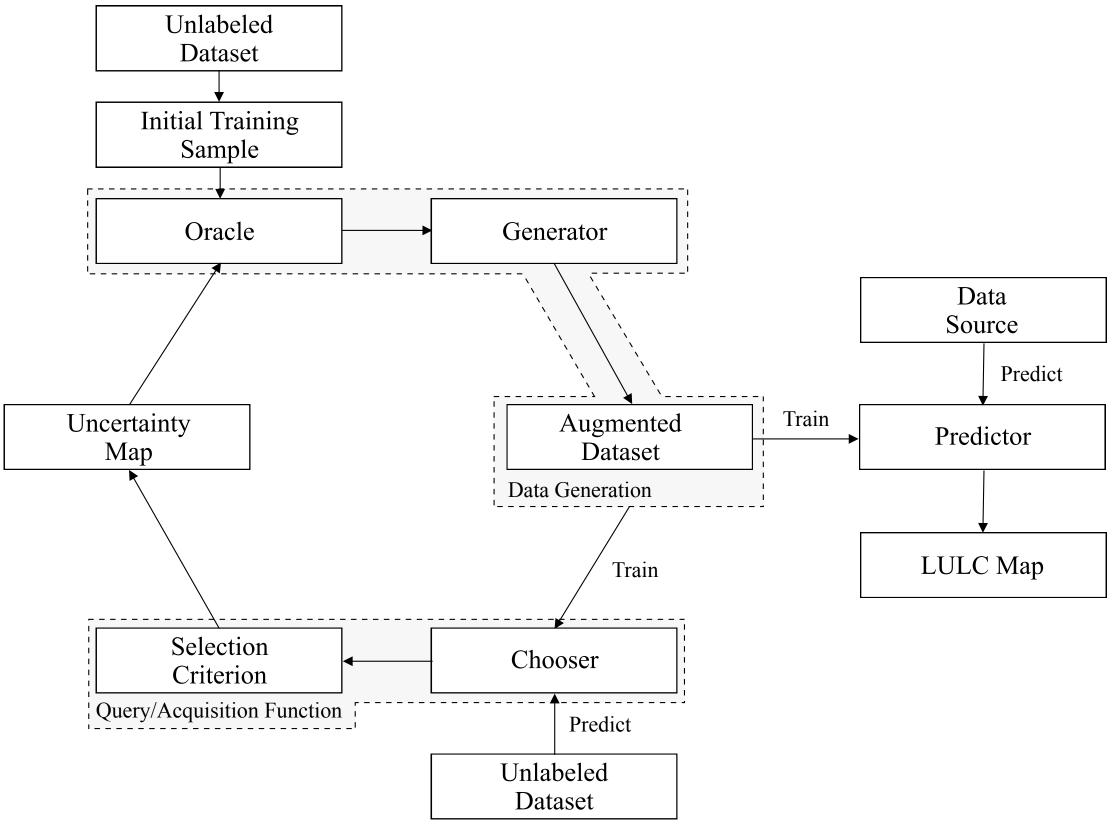

Figure 3: Proposed AL framework.

This method focuses on the capacity of artificial data to introduce more data variability into the augmented dataset and facilitate the chooser's training phase with a more consistent definition of the decision boundaries at each iteration. The artificial data is only used to train the classifiers involved in the process (chooser and predictor) and is discarded every time it is used to train the chooser. The remaining steps in the AL framework remain unchanged. This method is addressed towards the limitations found in the previous sections:

1. The lack of consistency of the chooser's performance over different initializations should be mitigated with the introduction of artificial data early on in the iterative process.

2. The convergence of the predictor's performance should be anticipated with the clearer definition of the decision boundaries across iterations.

3. Annotation cost is expected to reduce as the need for labeled observations reduces along with the early convergence of the classification performance.

4. The class imbalance bias observed in typical classification tasks, as well as in AL is mitigated by balancing the class frequencies at each iteration.

Although the performance of this method is shown within a LULC classification context, the proposed framework is completely domain agnostic. The high dimensionality of remotely sensed imagery make its classification particularly challenging when the availability of labeled data is scarce and/or comes at a high cost, being subjected to the curse of dimensionality. Consequently, it is a relevant and appropriate domain to test this method.

5 Methodology

5.1 Datasets

Dataset	Features	Instances	Min. Instances	Maj. Instances	IR	Classes
Botswana	145	1500	106	164	1.55	11
Salinas A	224	1500	109	428	3.93	6
Kennedy Space Center	176	1500	69	281	4.07	11
Indian Pines	220	1500	31	366	11.81	12
Salinas	224	1500	25	312	12.48	16
Pavia University	103	1500	33	654	19.82	9
Pavia Centre	102	1500	27	668	24.74	9

Table 1: Description of the datasets used for this experiment.

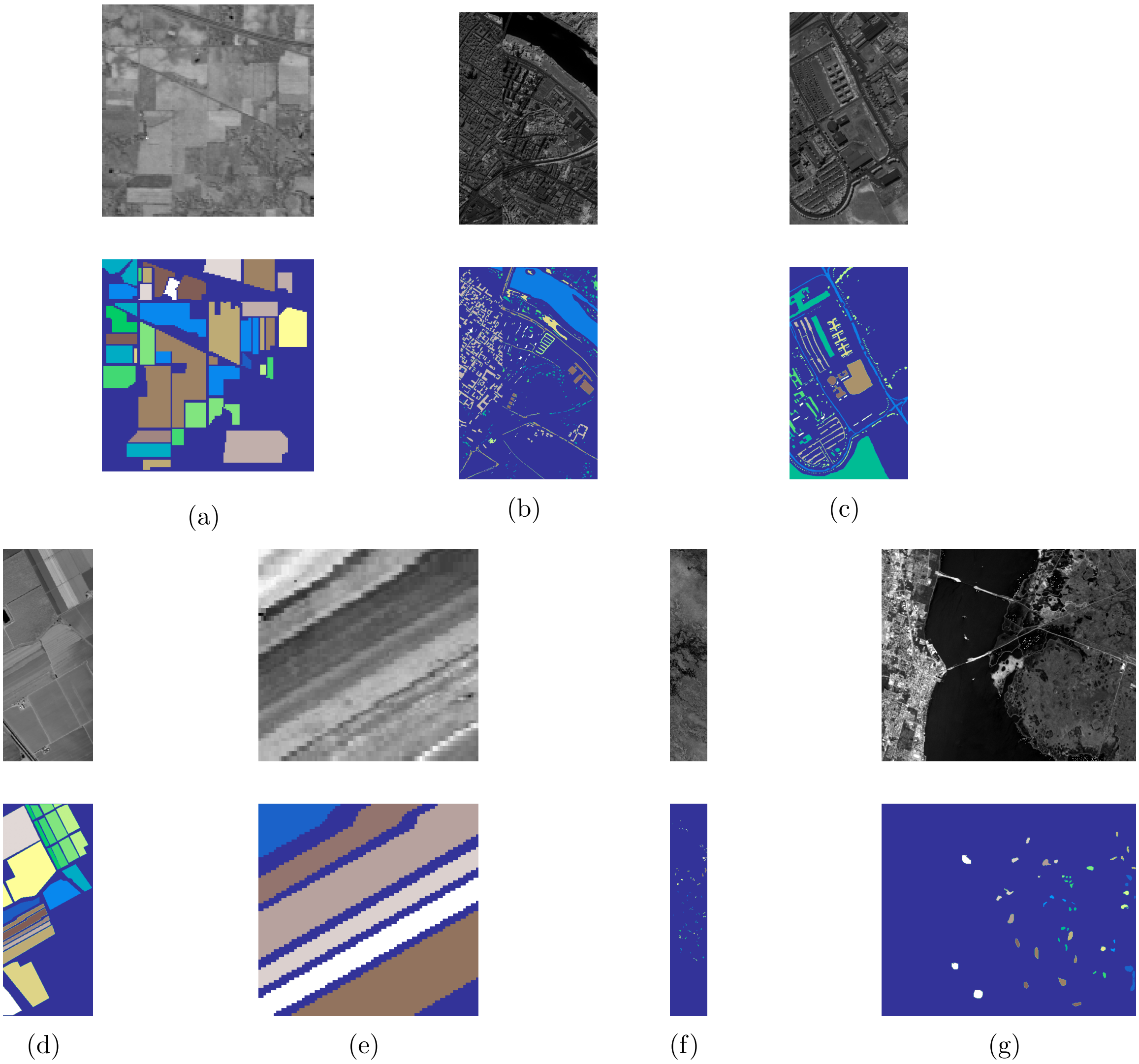

Figure 4: Gray scale visualization of a band (top row) and ground truth (bottom row) of each scene used in this study. (a) Indian Pines, (b) Pavia Centre, (c) Pavia University, (d) Salinas, (e) Salinas A, (f) Botswana, (g) Kennedy Space Center

5.2 Evaluation Metrics

1. Area under the learning curve (sums up all the performance values for each time point)

2. Deficiency

3. Data Utilization Rate

5.3 Machine Learning Algorithms

5.4 Experimental Procedure

A common practice in methodological evaluations is the implementation of an offline experiment [Kagy et al., 2019]. It consists of using an existing set of labeled data as a proxy for the population of unlabeled samples. Because the dataset is already fully labeled, the oracle's typical annotation process involved in each iteration is done at zero cost.

5.5 Software Implementation

6 Results

6.1 Statistical Analysis

7 Conclusion

References

[Aghdam et al., 2019] Aghdam, H. H., Gonzalez-Garcia, A., Lopez, A., and Weijer, J. (2019). Active learning for deep detection neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-Octob, pages 3671–3679.

[Alonso-Sarria et al., 2019] Alonso-Sarria, F., Valdivieso-Ros, C., and Gomariz-Castillo, F. (2019). Isolation forests to evaluate class separability and the representativeness of training and validation areas in land cover classification. *Remote Sensing*, 11(24):3000.

[Bogner et al., 2018] Bogner, C., Seo, B., Rohner, D., and Reineking, B. (2018). Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea. *PLoS ONE*, 13(1).

[Cawley, 2011] Cawley, G. (2011). Baseline Methods for Active Learning. *Proceedings of Active Learning and Experimental Design workshop In conjunction with AISTATS*, 16:47–57.

[Chawla et al., 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.

[Chawla et al., 2004] Chawla, N. V., Japkowicz, N., and Kotcz, A. (2004). Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explorations Newsletter*, 6(1):1–6.

[Copa et al., 2010] Copa, L., Tuia, D., Volpi, M., and Kanevski, M. (2010). Unbiased query-by-bagging active learning for VHR image classification. In Bruzzone, L., editor, *Image and Signal Processing for Remote Sensing XVI*, volume 7830, page 78300K. SPIE.

[Costa et al., 2020] Costa, H., Benevides, P., Marcelino, F., and Caetano, M. (2020). Introducing automatic satellite image processing into land cover mapping by photo-interpretation of airborne data.

The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42:29–34.

[Costantino et al., 2020] Costantino, D., Pepe, M., Dardanelli, G., and Baiocchi, V. (2020). USING OPTICAL SATELLITE AND AERIAL IMAGERY FOR AUTOMATIC COASTLINE MAPPING. *Geographia Technica*, pages 171–190.

[Demir et al., 2011] Demir, B., Persello, C., and Bruzzone, L. (2011). Batch-mode active-learning methods for the interactive classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 49(3):1014–1031.

[DeVries and Taylor, 2017] DeVries, T. and Taylor, G. W. (2017). Dataset augmentation in feature space. In *5th International Conference on Learning Representations, ICLR 2017 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR.

[Di and Crawford, 2012] Di, W. and Crawford, M. M. (2012). View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5 PART 2):1942–1954.

[Douzas and Bacao, 2017] Douzas, G. and Bacao, F. (2017). Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning. *Expert Systems with Applications*, 82:40–52.

[Douzas and Bacao, 2019] Douzas, G. and Bacao, F. (2019). Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE. *Information Sciences*, 501:118–135.

[Douzas et al., 2019] Douzas, G., Bacao, F., Fonseca, J., and Khudinyan, M. (2019). Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm. *Remote Sensing*, 11(24):3040.

[Douzas et al., 2018] Douzas, G., Bacao, F., and Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*, 465:1–20.

[Ertekin et al., 2007] Ertekin, S., Huang, J., and Giles, C. L. (2007). Active learning for class imbalance problem. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07*, pages 823–824, New York, New York, USA. ACM Press.

[Feng et al., 2018] Feng, W., Huang, W., Ye, H., and Zhao, L. (2018). Synthetic minority over-sampling technique based rotation forest for the classification of unbalanced hyperspectral data. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 2018-July, pages 2651–2654. Institute of Electrical and Electronics Engineers Inc.

[Fernández et al., 2013] Fernández, A., López, V., Galar, M., del Jesus, M. J., and Herrera, F. (2013). Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowledge-Based Systems*, 42:97–110.

[Gavade and Rajpurohit, 2019] Gavade, A. B. and Rajpurohit, V. S. (2019). Systematic analysis of satellite image-based land cover classification techniques: literature review and challenges. *International Journal of Computers and Applications*, pages 1–10.

[Han et al., 2005] Han, H., Wang, W. Y., and Mao, B. H. (2005). Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. *Lecture Notes in Computer Science*, 3644(PART I):878–887.

[He et al., 2008] He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the International Joint Conference on Neural Networks*, pages 1322–1328. IEEE.

[Huang et al., 2018] Huang, Y., xin CHEN, Z., YU, T., zhi HUANG, X., and fa GU, X. (2018). Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, 17(9):1915–1931.

[Jozdani et al., 2019] Jozdani, S. E., Johnson, B. A., and Chen, D. (2019). Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. *Remote Sensing*, 11(14):1713.

[Kagy et al., 2019] Kagy, J.-F., Kayadelen, T., Ma, J., Rostamizadeh, A., and Strnadova, J. (2019). The practical challenges of active learning: Lessons learned from live experimentation.

[Kaur et al., 2019] Kaur, H., Pannu, H. S., and Malhi, A. K. (2019). A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys*, 52(4):1–36.

[Khatami et al., 2016] Khatami, R., Mountrakis, G., and Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, 177:89–100.

[Kottke et al., 2017] Kottke, D., Calma, A., Huseljic, D., Krempl, G., and Sick, B. (2017). Challenges of reliable, realistic and comparable active learning evaluation. In *CEUR Workshop Proceedings*, volume 1924, pages 2–14.

[Li et al., 2011] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2011). Hyperspectral image segmentation using a new bayesian approach with active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10 PART 2):3947–3960.

[Li et al., 2013] Li, J., Bioucas-Dias, J. M., and Plaza, A. (2013). Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2):844–856.

[Li et al., 2020] Li, J., Huang, X., and Chang, X. (2020). A label-noise robust active learning sample collection method for multi-temporal urban land-cover classification and change analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163(January):1–17.

[Li and Guo, 2013] Li, X. and Guo, Y. (2013). Active learning with multi-label svm classification. In *In IJCAI*, pages 1479–1485.

[Liu et al., 2020] Liu, S.-J., Luo, H., and Shi, Q. (2020). Active Ensemble Deep Learning for Polarimetric Synthetic Aperture Radar Image Classification. *IEEE Geoscience and Remote Sensing Letters*, pages 1–5.

[Liu et al., 2018] Liu, W., Yang, J., Li, P., Han, Y., Zhao, J., and Shi, H. (2018). A novel object-based supervised classification method with active learning and random forest for PolSAR imagery. *Remote Sensing*, 10(7).

[Luo et al., 2003] Luo, T., Kramer, K., Goldgof, D., Hall, L. O., Samson, S., Remsen, A., and Hopkins, T. (2003). Learning to recognize plankton. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 888–893.

[Maxwell et al., 2018] Maxwell, A. E., Warner, T. A., and Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9):2784–2817.

[Muslea et al., 2006] Muslea, I., Minton, S., and Knoblock, C. A. (2006). Active learning with multiple views. *Journal of Artificial Intelligence Research*, 27:203–233.

[Nagai et al., 2020] Nagai, S., Nasahara, K. N., Akitsu, T. K., Saitoh, T. M., and Muraoka, H. (2020). Importance of the Collection of Abundant Ground-Truth Data for Accurate Detection of Spatial and Temporal Variability of Vegetation by Satellite Remote Sensing. In *Biogeochemical Cycles: Ecological Drivers and Environmental Impact*, pages 223–244. American Geophysical Union (AGU).

[Pasolli et al., 2016] Pasolli, E., Yang, H. L., and Crawford, M. M. (2016). Active-metric learning for classification of remotely sensed hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(4):1925–1939.

[Pelletier et al., 2017] Pelletier, C., Valero, S., Inglada, J., Champion, N., Sicre, C. M., and Dedieu, G. (2017). Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing*, 9(2):173.

[Růžička et al., 2020] Růžička, V., D'Aronco, S., Wegner, J. D., and Schindler, K. (2020). Deep active learning in remote sensing for data efficient change detection. *arXiv preprint arXiv:2008.11201*.

[Shrivastava and Pradhan, 2021] Shrivastava, V. K. and Pradhan, M. K. (2021). Hyperspectral Remote Sensing Image Classification Using Active Learning. In *Studies in Computational Intelligence*, volume 907, pages 133–152. Springer.

[Stromann et al., 2020] Stromann, O., Nascetti, A., Yousif, O., and Ban, Y. (2020). Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine. *Remote Sensing*, 12(1):76.

[Su et al., 2020] Su, T., Zhang, S., and Liu, T. (2020). Multi-spectral image classification based on an object-based active learning approach. *Remote Sensing*, 12(3):504.

[Sverchkov and Craven, 2017] Sverchkov, Y. and Craven, M. (2017). A review of active learning approaches to experimental design for uncovering biological networks. *PLOS Computational Biology*, 13(6):e1005466.

[Tuia et al., 2011] Tuia, D., Pasolli, E., and Emery, W. J. (2011). Using active learning to adapt remote sensing image classifiers. *Remote Sensing of Environment*, 115(9):2232–2242.

[Tuia et al., 2009] Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., and Emery, W. J. (2009). Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2218–2232.

[Vermote et al., 2020] Vermote, E. F., Skakun, S., Becker-Reshef, I., and Saito, K. (2020). Remote sensing of coconut trees in tonga using very high spatial resolution worldview-3 data. *Remote Sensing*, 12(19):3113.

[Wang and Xie, 2018] Wang, X. and Xie, H. (2018). A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management. *Water (Switzerland)*, 10(5):608.

[Wulder et al., 2018] Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., and Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, 39(12):4254–4284.

[Yoo and Kweon, 2019] Yoo, D. and Kweon, I. S. (2019). Learning loss for active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Zhang et al., 2016] Zhang, Z., Pasolli, E., Yang, H. L., and Crawford, M. M. (2016). Multimetric Active Learning for Classification of Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 13(7):1007–1011.

[Zhou et al., 2014] Zhou, X., Prasad, S., and Crawford, M. (2014). Wavelet domain multi-view active learning for hyperspectral image analysis. In *Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing*, volume 2014-June. IEEE Computer Society.