REINFORCEMENT-BASED FRUGAL LEARNING FOR INTERACTIVE SATELLITE IMAGE CHANGE DETECTION

Sébastien Deschamps, Hichem Sahbi

REINFORCEMENT-BASED FRUGAL LEARNING FOR INTERACTIVE SATELLITE IMAGE CHANGE DETECTION

Sebastien Deschamps[1,2] *Hichem Sahbi*[1]

[1]Sorbonne University, UPMC, CNRS, LIP6, France [2]Theresis Thales, France

ABSTRACT

In this paper, we introduce a novel interactive satellite image change detection algorithm based on active learning. The proposed approach is iterative and asks the user (oracle) questions about the targeted changes and according to the oracle's responses updates change detections. We consider a probabilistic framework which assigns to each unlabeled sample a relevance measure modeling how critical is that sample when training change detection functions. These relevance measures are obtained by minimizing an objective function mixing diversity, representativity and uncertainty. These criteria when combined allow exploring different data modes and also refining change detections. To further explore the potential of this objective function, we consider a reinforcement learning approach that finds the best combination of diversity, representativity and uncertainty, through active learning iterations, leading to better generalization as corroborated through experiments in interactive satellite image change detection.

***Index Terms*—** active learning, reinforcement learning, satellite image change detection

1. INTRODUCTION

Satellite image change detection consists in finding occurrences of targeted (relevant) changes into a scene at a given instant w.r.t. the same scene acquired earlier [4, 6–8]. This includes appearance or disappearance of visual entities such as infrastructure destruction after natural hazards (earthquakes, tornadoes, etc.) [2, 3]. This task is very challenging as relevant changes are eclectic and satellite images are subject to multiple sources of irrelevant changes including illumination, artefacts, clouds, etc. Existing solutions either remove irrelevant variations in satellite images by correcting their radiometric effects [1, 10, 12–14] or consider them as a part of appearance modeling [9, 11, 15–20]. The latter consists in designing statistical or machine inference models [5, 37] that learn how to discriminate between relevant and irrelevant changes. Training these models requires enough labeled data covering all the sources of variability due to both the positive and negative classes. However, beside the scarceness of labeled data, the relevance of changes could be subjective and may vary from one user to another, and this makes the task of automatic change detection highly challenging.

Existing machine learning approaches that mitigate the scarceness of labeled data include few shot, self-supervised and active learning [21–24, 26–33]. Among these methods, active learning is particularly interesting and allows modeling the user's subjectivity (about targeted changes) more accurately. Active learning solutions are interactive approaches that show the most critical unlabeled data (a.k.a. displays) to the user/oracle, and ask the latter about the relevance of changes prior to update change detections [34]. Display section strategies usually rely on diversity, representativity and uncertainty [25]. Diversity allows exploring different modes of the unlabeled data, representativity seeks to select prototypical samples in those modes while uncertainty allows displaying the most ambiguous data that ultimately refine change detections. However, knowing a priori which sequence of display strategies (diversity, representativity and uncertainty) to apply through all the iterations of active learning is highly combinatorial. Besides, under the frugal learning regime, labeled validation sets are scarce in order to make the optimization of these strategies statistically meaningful.

In this paper, we devise a novel change detection algorithm that asks the oracle the most informative questions about targeted changes and according to the oracle's responses updates change detections. The proposed solution is probabilistic and assigns to each unlabeled sample a relevance measure which captures how critical is that sample when learning changes. These relevance measures are obtained as the optimum of an objective function that mixes diversity, representativity and ambiguity criteria. In order to tackle the combinatorial aspect of these criteria, we further rely on reinforcement learning (RL) which finds the "optimal" sequence of actions (diversity, representativity and ambiguity as well as their possible combination) that ultimately leads to high generalization. Experiments conducted on the challenging task of interactive satellite image change detection show the superiority and the outperformance of the proposed RL-based approach w.r.t. related work.

2. PROPOSED MODEL

Let $\mathcal{I}_r = \{p_1, \ldots, p_n\}$, $\mathcal{I}_t = \{q_1, \ldots, q_n\}$ denote two registered satellite images taken at two different time-stamps t_0,

t_1 respectively, and let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of aligned patch pairs with $\mathbf{x}_i = (p_i, q_i) \in \mathcal{I}_r \times \mathcal{I}_t$. Considering the labels of $\mathcal{X}$ initially unknown, our goal is to design a classifier $g(.)$ by interactively labeling a very *small* fraction of $\mathcal{X}$ (as change / no-change), and training the parameters of g. This interactive labeling and training is known as active learning.

Let $\mathcal{D}_t$ be a *display* (defined as a subset of $\mathcal{X}$) shown to an oracle[1] at any iteration t of active learning, and let $\mathcal{Y}_t$ be the underlying labels. The initial display $\mathcal{D}_t$ (with $t = 0$) is uniformly sampled at random, and used to train the subsequent classifiers by repeating the following steps till reaching high generalization or exhausting a labeling budget:

i) Get the labels of $\mathcal{D}_t$ as $\mathcal{Y}_t \leftarrow \text{oracle}(\mathcal{D}_t)$.

ii) Train $g_t(.)$ using $\bigcup_{\tau=1}^{t}(\mathcal{D}_\tau, \mathcal{Y}_\tau)$ where the subscript in $g_t(.)$ refers to the decision function at iteration t. In this paper, support vector machines (built on top of convolutional features) are used.

iii) Select the next display $\mathcal{D} \subset \mathcal{X} - \bigcup_{\tau=1}^{t} \mathcal{D}_\tau$ that possibly increases the generalization performances of the subsequent classifier $g_{t+1}(.)$. As the labels of $\mathcal{D}$ are unknown, one cannot combinatorially sample all the possible subsets $\mathcal{D}$, train the associated classifiers, and select the best display. Alternative display selection strategies (a.k.a display models) are usually related to active learning and seek to find the most representative display that eventually yields optimal decision functions [22]. In what follows, we introduce our main contribution: a novel display model which allows selecting the most representative samples to label by an oracle and ultimately lead to high generalization performances, in satellite image change detection, as corroborated later in experiments.

2.1. Display model

We consider a probabilistic framework which assigns for each sample $\mathbf{x}_i \in \mathcal{X}$ a membership degree μ_i that measures the probability of $\mathbf{x}_i$ belonging to the next display $\mathcal{D}_{t+1}$; consequently, $\mathcal{D}_{t+1}$ will correspond to the unlabeled data in $\{\mathbf{x}_i\}_i \subset \mathcal{X}$ with the highest memberships $\{\mu_i\}_i$. Considering $\mu \in \mathbb{R}^n$ (with $n = |\mathcal{X}|$) as a vector of these memberships $\{\mu_i\}_i$, we propose to find μ as the minimum of the following constrained optimization problem

$$\min_{\mu \geq 0, \|\mu\|_1 = 1} \eta\, \mathbf{tr}\big(\text{diag}(\mu'[\mathbf{C} \circ \mathbf{D}])\big) + \alpha\, [\mathbf{C}'\mu]' \log[\mathbf{C}'\mu] + \beta\, \mathbf{tr}\big(\text{diag}(\mu'[\mathbf{F} \circ \log \mathbf{F}])\big) + \mu' \log \mu, \tag{1}$$

here $\circ$, $'$ are respectively the Hadamard product and the matrix transpose, $\|.\|_1$ is the ℓ_1 norm, $\log$ is applied entry-wise, and diag maps a vector to a diagonal matrix. In the above objective function (i) $\mathbf{D} \in \mathbb{R}^{n \times K}$ and $\mathbf{D}_{ik} = d_{ik}^2$ is the euclidean distance between $\mathbf{x}_i$ and k^{th} cluster centroid of a partition of $\mathcal{X}$ obtained with K-means clustering, (ii) $\mathbf{C} \in \mathbb{R}^{n \times K}$ is the indicator matrix with each entry $\mathbf{C}_{ik} = 1$ iff $\mathbf{x}_i$ belongs to the k^{th} cluster (0 otherwise), and (iii) $\mathbf{F} \in \mathbb{R}^{n \times 2}$ is a scoring matrix with $(\mathbf{F}_{i1}, \mathbf{F}_{i2}) = (\hat{g}_t(\mathbf{x}_i), 1 - \hat{g}_t(\mathbf{x}_i))$ and $\hat{g}_t \in [0, 1]$ being a normalized version of g_t. The first term of this objective function (rewritten as $\sum_i \sum_k 1_{\{\mathbf{x}_i \in h_k\}} \mu_i d_{ik}^2$) measures the *representativity* of the selected samples in $\mathcal{D}$; in other words, it captures how close is each $\mathbf{x}_i$ w.r.t. the centroid of its cluster, so this term reaches its smallest value when all the selected samples coincide with these centroids. The second term (rewritten as $\sum_k [\sum_{i=1}^n 1_{\{\mathbf{x}_i \in h_k\}} \mu_i] \log[\sum_{i=1}^n 1_{\{\mathbf{x}_i \in h_k\}} \mu_i]$) measures the *diversity* of the selected samples as the entropy of the probability distribution of the underlying clusters; this measure is minimized when the selected samples belong to different clusters and vice-versa. The third criterion (equivalent to $\sum_i \sum_c^{nc} \mu_i \mathbf{F}_{ic} \log \mathbf{F}_{ic}$) captures the *ambiguity* in $\mathcal{D}$ measured as the entropy of the scoring function; this term reaches its smallest value when data are evenly scored w.r.t. different categories. Finally, the fourth term is related to the *cardinality* of $\mathcal{D}$, measured by the entropy of the distribution μ; this term also acts as a regularizer. Considering $\mathbf{1}_{nc}$, $\mathbf{1}_K$ as vectors of nc and K ones respectively (with $nc = 2$ in practice), one may show that the solution of Eq. 1 is given as $\mu^{(\tau+1)} := \hat{\mu}^{(\tau+1)} / \|\hat{\mu}^{(\tau+1)}\|_1$, with $\hat{\mu}^{(\tau+1)}$ being

$$\exp\Big(-\big[\eta(\mathbf{D} \circ \mathbf{C})\mathbf{1}_K + \alpha \mathbf{C}(\log[\mathbf{C}'\mu^{(\tau)}] + \mathbf{1}_K) + \beta(\mathbf{F} \circ \log \mathbf{F})\mathbf{1}_{nc}\big]\Big). \tag{2}$$

As shown later in experiments, the setting of the hyper-parameters α, β, η is crucial for the success of the display model. For instance, putting more emphasis on diversity (i.e., high α) results into high exploration of class modes while a high focus on ambiguity (i.e., large β) locally refines the trained decision functions. A suitable balance between exploration and local refinement of the learned decision functions should be achieved by selecting the best configuration of these hyper-parameters. Nevertheless, since labeling is sparingly achieved by the oracle, no sufficiently large validation sets could be made available beforehand to accurately set these hyper-parameters.

2.2. RL-based display model

Let Λ_α, Λ_β, Λ_η denote the parameter spaces associated to α, β, η respectively, and let Λ be the underlying Cartesian product. For any instance $\lambda \in \Lambda$ (at a given iteration $t+1$), one may obtain a display (now rewritten as $\mathcal{D}_{t+1}^{\lambda}$) by solving Eq. 1. In order to find the best configuration λ^* that yields an "optimal" display, we model hyper-parameter selection as a Markov Decision Process (MDP). An MDP based RL corresponds to a tuple $\langle \mathcal{S}, \mathcal{A}, R, T, \delta \rangle$ with $\mathcal{S}$ being a state set, $\mathcal{A}$ an action set, $R : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ an immediate reward function, $T : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$ a transition function and δ a discount factor [35]. RL consists in running a sequence of actions from $\mathcal{A}$ with the goal of maximizing an expected discounted reward by following a stochastic policy, $\pi : \mathcal{S} \mapsto \mathcal{A}$; this leads to the

[1] The oracle is defined as an expert annotator providing labels (changes / no-changes) for any given subset of images.

true state-action value as

$$Q(s,a) = E_\pi \left[\sum_{k=0}^{\infty} \delta^k r_k | S_0, = s, A_0 = a \right], \quad (3)$$

here E_π denotes the expectation w.r.t. π, r_k is the immediate reward at the k^{th} step of RL, S_0 an initial state, A_0 an initial action and $\delta \in [0,1]$ is a discount factor that balances between immediate and future rewards. The goal of the optimal policy is to select actions that maximize the discounted cumulative reward; i.e., $\pi_*(s) \leftarrow \arg\max_a Q(s,a)$. One of the most used methods to solve this type of RL problems is Q-learning [36], which directly estimates the optimal value function and obeys the fundamental identity, the Bellman equation

$$Q_*(s,a) = E_\pi \left[R(s,a) + \delta \max_{a'} Q_*(s',a') | S_0 = s, A_0 = a \right], \quad (4)$$

with $s' = T(s,a)$ and $R(s,a)$ is again the immediate reward. We consider in our hyper-parameter optimization, a stateless version, so $Q(s,a)$ and $R(s,a)$ are rewritten simply as $Q(a)$, $R(a)$ respectively. In this configuration, the parameter space Λ is equal to $\{0,1\}^3 \backslash (0,0,0)$ so the underlying action set $\mathcal{A}$ corresponds to 7 possible binary (zero / non-zero) settings of α, β, η. We consider an adversarial immediate reward function R that scores a given action (and hence the underlying configuration $\lambda \in \Lambda$) proportionally to the error rates of $g_t(\mathcal{D}^\lambda_{t+1})$; put differently, the display $\mathcal{D}^\lambda_{t+1}$ is selected in order to challenge (the most) the current classifier g_t, leading to a better estimate of g_{t+1}. With this RL-based design, better change detection performances are observed as shown subsequently in experiments.

3. EXPERIMENTS

Dataset and setting. We evaluate the accuracy of our RL-based interactive change detection algorithm using the Jefferson dataset. The latter consists of $2,200$ non-overlapping (30×30 RGB) patch pairs taken from (bi-temporal) GeoEye-1 satellite images of $2,400 \times 1,652$ pixels with a spatial resolution of 1.65m/pixel. These patch pairs pave a large area from Jefferson (Alabama) in 2010 and in 2011. These images show several damages caused by tornadoes (building destruction, debris on roads, etc) as well as no-changes including irrelevant ones (clouds, etc). In this dataset $2,161$ patch pairs correspond to negative data and only 39 pairs to positive, so $< 2\%$ of these data correspond to relevant changes and this makes their detection very challenging. In our experiments, half of the patch pairs are used for training and the remaining ones for testing. We measure the accuracy of change detection using the equal error rate (EER); the latter is a balanced generalization error that evenly weights errors in the positive and negative classes. Smaller EERs imply better performances.

Ablation study and impact of RL. In the first set of experiments, we show an ablation study of our display model and thereby the impact of ambiguity, representativity and diversity criteria when taken individually and combined. From these results, we observe the positive impact of diversity at the early iterations of active learning, while the impact of ambiguity comes later in order to further refine the learned change detection functions. However, none of the settings (rows) in table 1 obtains the best performance through all the iterations of active learning. Considering these observed ablation performances, a better setting of the α, β and η should be cycle-dependent using reinforcement learning (as described in section 2.2), and as also corroborated through performances shown in table 1. Indeed, it turns out that this adaptive setting outperforms the other combinations (including "all", also referred to as "flat"), especially at the late iterations of change detection.

Iter	1	2	3	4	5	6	7	8	9	10	AUC
Samp%	1.45	2.90	4.36	5.81	7.27	8.72	10.18	11.63	13.09	14.54	
rep	48.05	26.21	12.72	10.48	9.88	9.70	8.52	8.85	8.61	8.82	15.18
div	48.05	31.24	23.45	30.41	44.81	24.12	13.22	17.02	6.88.	7.98	24.71
amb	48.05	46.68	38.73	29.91	14.74	20.11	8.33	7.41	7.37	5.53	22.68
rep+div	48.05	26.21	33.35	25.10	21.55	11.71	2.84	1.65	1.59	1.43	17.34
rep+amb	48.05	26.21	12.62	10.81	9.82	9.70	8.53	9.23	8.60	8.82	15.23
div+amb	48.05	41.69	28.82	23.08	23.41	23.42	19.82	13.10	8.16	6.97	23.65
all (flat)	48.05	26.21	33.35	25.52	23.70	14.59	2.74	1.54	1.67	1.48	17.88
RL-based	48.05	31.75	10.36	14.83	13.36	14.70	1.06	1.06	1.10	**1.01**	**13.72**

Table 1. This table shows an ablation study of our display model. Here rep, amb and div stand for representativity, ambiguity and diversity respectively. These results are shown for different iterations t (Iter) and the underlying sampling rates (Samp) defined as $(\sum_{k=0}^{t-1} |\mathcal{D}_k|/(|\mathcal{X}|/2)) \times 100$. The AUC (Area Under Curve) corresponds to the average of EERs across iterations.

Extra comparison. Figure. 1 shows extra comparisons of our RL-based display model w.r.t. different related display sampling techniques including *random, MaxMin and uncertainty*. Random picks data from the unlabeled set whereas MaxMin greedily selects a sample $\mathbf{x}_i$ in $\mathcal{D}_{t+1}$ from the pool $\mathcal{X} \backslash \cup_{k=0}^{t} \mathcal{D}_k$ by maximizing its minimum distance w.r.t $\cup_{k=0}^{t} \mathcal{D}_k$. We also compare our method w.r.t. uncertainty which consists in selecting samples in the display whose scores are the closest to zero (i.e., the most ambiguous). Finally, we also consider the fully supervised setting as an upper bound on performances; this configuration relies on the whole annotated training set and builds the learning model in one shot.

The EERs in figure 1 show the positive impact of the proposed RL-based display model against the related sampling strategies for different amounts of annotated data. The comparative methods are effective either at the early iterations of active learning (such as MaxMin and random which capture the diversity of data without being able to refine decision functions) or at the latest iterations (such as uncertainty which locally refines change detection functions but suffers from the lack of diversity). In contrast, our proposed RL-based design adapts the choice of these criteria as active learning cycles evolve, and thereby allows our interactive change detection to reach lower EERs and to overtake all the other strategies at the end of the iterative process.

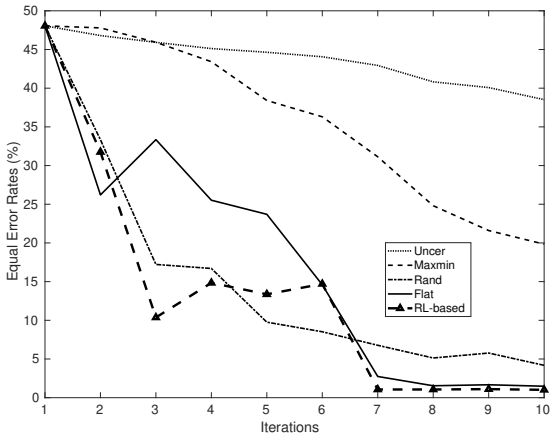

Fig. 1. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table 1 (Samp). Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER of 0.94%. See again section 3 for more details.

4. CONCLUSION

We introduce in this paper a satellite image change detection algorithm based on active and reinforcement learning. The strength of the proposed method resides in its ability to find and adapt display selection criteria to the active learning iterations, thereby leading to more informative subsequent displays and more accurate decision functions. Extensive experiments conducted on the challenging task of change detection shows the accuracy and the out-performance of the proposed interactive method w.r.t. the related work.

5. REFERENCES

[1] N. Bourdis, D. Marraud and H. Sahbi. "Camera pose estimation using visual servoing for aerial video change detection." IEEE IGARSS 2012.

[2] D. Brunner, G. Lemoine, and L. Bruzzone, Earthquake damage assessment of buildings using vhr optical and sar imagery, IEEE Trans. Geosc. Remote Sens., vol. 48, no. 5, pp. 2403–2420, 2010.

[3] H. Gokon, J. Post, E. Stein, S. Martinis, A. Twele, M. Muck, C. Geiss, S. Koshimura, and M. Matsuoka, A method for detecting buildings destroyed by the 2011 tohoku earthquake and tsunami using multitemporal terrasar-x data, GRSL, vol. 12, no. 6, pp. 1277–1281, 2015.

[4] J. Deng, K. Wang, Y. Deng, and G. Qi, PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data, IJRS, vol. 29, no. 16, pp. 4823–4838, 2008.

[5] H. Sahbi. "Coarse-to-fine deep kernel networks." IEEE ICCV-W, 2017.

[6] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, Image change detection algorithms: A systematic survey, IEEE Trans. on Im Proc, vol. 14, no. 3, pp. 294–307, 2005.

[7] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images, TGRS, vol. 53, no. 8, pp. 4363–4378, 2015.

[8] G. Chen, G. J. Hay, L. M. Carvalho, and M. A. Wulder, Object-based change detection, IJRS, vol. 33, no. 14, pp. 4434–4457, 2012.

[9] N. Bourdis, D. Marraud and H. Sahbi. "Constrained optical flow for aerial image change detection." in IEEE IGARSS, 2011.

[10] J. Zhu, Q. Guo, D. Li, and T. C. Harmon, Reducing mis-registration and shadow effects on change detection in wetlands, Photogrammetric Engineering & Remote Sensing, vol. 77, no. 4, pp. 325–334, 2011.

[11] H. Sahbi. "Interactive satellite image change detection with context-aware canonical correlation analysis." IEEE GRSL, (14)5, 2017.

[12] A. Fournier, P. Weiss, L. Blanc-Fraud, and G. Aubert, A contrast equalization procedure for change detection algorithms: applications to remotely sensed images of urban areas, In ICPR, 2008

[13] Carlotto, Detecting change in images with parallax, In Society of Photo-Optical Instrumentation Engineers, 2007

[14] S. Leprince, S. Barbot, F. Ayoub, and J.-P. Avouac, Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements,TGRS, vol. 45, no. 6, pp. 1529–1558, 2007.

[15] H. Sahbi. "Relevance feedback for satellite image change detection." IEEE ICASSP, 2013.

[16] Pollard, Comprehensive 3d change detection using volumetric appearance modeling, Phd, Brown University, 2009.

[17] A. A. Nielsen, The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data, IEEE Transactions on Image processing, vol. 16, no. 2, pp. 463–478, 2007.

[18] C. Wu, B. Du, and L. Zhang, Slow feature analysis for change detection in multispectral imagery, TGRS, vol. 52, no. 5, pp. 2858–2874, 2014

[19] N. Bourdis, D. Marraud, and H. Sahbi, Spatio-temporal interaction for aerial video change detection, in IGARSS, 2012, pp. 2253–2256

[20] J. Im, J. Jensen, and J. Tullis, Object-based change detection using correlation image analysis and image segmentation, International Journal of Remote Sensing, vol. 29, no. 2, pp. 399–423, 2008.

[21] Vinyals et al., Matching networks for one shot learning. 2016.

[22] Dasgupta, Sanjoy. "Analysis of a greedy active learning strategy." Advances in neural information processing systems 17 (2004).

[23] Q. Oliveau and H. Sahbi. "Learning attribute representations for remote sensing ship category classification." IEEE JSTARS 10.6 (2017): 2830-2840.

[24] Burr, Settles. "Active learning." Synthesis Lectures on Artificial Intelligence and Machine Learning 6.1 (2012).

[25] He, Tianxu, et al. "An active learning approach with uncertainty, representativeness, and diversity." The Scientific World Journal 2014 (2014).

[26] Joshi et al., Multi-class active learning for image classification. 2009.

[27] Settles & Craven. An analysis of active learning strategies for sequence labeling tasks. 2008.

[28] Houlsby et al., Bayesian active learning for classification and preference learning. 2011.

[29] Campbell & Broderick, Automated scalable Bayesian inference via Hilbert coresets. 2019.

[30] P. Vo and H. Sahbi. "Transductive kernel map learning and its application to image annotation." BMVC. 2012.

[31] Gal et al., Deep bayesian active learning with image data. 2017

[32] Pang et al., Meta-Learning Transferable Active Learning Policies by Deep Reinforcement Learning

[33] A. Kolesnikov, X. Zhai, L. Beyer. Revisiting Self-Supervised Visual Representation Learning Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1920-1929

[34] H. Sahbi, S. Deschamps, A. Stoian. Frugal Learning for Interactive Satellite Image Change Detection. IEEE IGARSS, 2021.

[35] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.

[36] Jin, C., Allen-Zhu, Z., Bubeck, S., & Jordan, M. I. (2018). Is Q-learning provably efficient?. arXiv preprint arXiv:1807.03765.

[37] M. Jiu and H. Sahbi. "Laplacian deep kernel learning for image annotation." IEEE ICASSP, 2016.