



HAL
open science

Frugal Learning for Interactive Satellite Image Change Detection

Hichem Sahbi, Sébastien Deschamps, Andrei Stoian

► **To cite this version:**

Hichem Sahbi, Sébastien Deschamps, Andrei Stoian. Frugal Learning for Interactive Satellite Image Change Detection. IEEE International Geoscience and Remote Sensing Symposium, Jul 2021, Brussels, Belgium. pp.2811-2814, 10.1109/IGARSS47720.2021.9553385 . hal-03431343

HAL Id: hal-03431343

<https://hal.science/hal-03431343>

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FRUGAL LEARNING FOR INTERACTIVE SATELLITE IMAGE CHANGE DETECTION

Hichem Sahbi¹

Sebastien Deschamps^{1,2}

Andrei Stoian²

¹Sorbonne University, UPMC, CNRS, LIP6, France

²Therisis Thales, France

ABSTRACT

We introduce in this paper a novel active learning algorithm for satellite image change detection. The proposed solution is interactive and based on a question & answer model, which asks an oracle (annotator) the most informative questions about the relevance of sampled satellite image pairs, and according to the oracle’s responses, updates a decision function iteratively. We investigate a novel framework which models the probability that samples are relevant; this probability is obtained by minimizing an objective function capturing representativity, diversity and ambiguity. Only data with a high probability according to these criteria are selected and displayed to the oracle for further annotation. Extensive experiments on the task of satellite image change detection after natural hazards (namely tornadoes) show the relevance of the proposed method against the related work.

Index Terms— Frugal learning, interactive satellite image change detection.

1. INTRODUCTION

Satellite image change detection is the process of identifying occurrences of targeted changes in a scene, at a given instant t_1 , w.r.t. the same scene acquired at an earlier instant t_0 . One of the major applications of change detection is damage assessment after natural hazards (e.g., tornadoes, earth-quakes, etc.) in the purpose of prioritizing disaster response accordingly. This task consists in finding relevant changes (such as building destruction, damage of transportation routes and infrastructures, etc.) while discarding irrelevant ones (due to weather conditions, occlusions, sensor artefact and alignment errors, as well as effects due to oscillating objects, trees, etc.).

Early change detection solutions were initially based on simple comparisons of multi-temporal signals, via image differences and thresholding, using vegetation indices, principal component and change vector analyses (see [1–3, 5, 6, 9] and references therein). These methods also rely on a preliminary pre-processing step that attenuates the effects of irrelevant changes, by finding parameters of sensors for registration as well as correcting radiometric effects, occlusions and shadows (see for instance [10–15]). Other methods either ignore irrelevant changes or consider them as a part of appearance model design and are able to detect relevant changes while

being resilient to irrelevant ones [16–20]. In particular, machine learning methods are promising [7, 8], but their success is highly dependent on the availability of large collections of hand-labeled training data [4, 22, 33, 34]. Indeed, these approaches are limited by the huge scene variability and the lack of labeled training data in order to characterize rare relevant changes and abundant irrelevant ones. Besides, even when labeled data are available, these labels could be erroneous and oblivious to the annotator’s (user’s) subjectivity and intention.

Many existing solutions try to bypass the aforementioned limitations by making learning frugal and less dependent on large collections of labeled data; this includes few shot [21] and self-supervised learning [32]. However, these solutions are limited as their design is agnostic to the user’s subjectivity and intention. Weakly-supervised and active learning solutions [23, 25–30] are rather more appropriate where users annotate very few samples of the most relevant and irrelevant changes (according to a targeted intention), and a training model updates a decision criterion accordingly. In this paper, we follow this line and we introduce a novel interactive satellite image change detection algorithm which asks the user the most informative questions about the relevance of changes and according to the user’s responses updates a change detection criterion. The new proposed model is probabilistic and assigns for each unlabeled training sample a measure which captures how critical is that sample in order to update the decision criteria. We obtain this measure as the optimum of a constrained objective function mixing representativity, diversity, ambiguity and cardinality. In contrast to related active learning solutions [24], which are basically heuristics, the one proposed in this paper is probabilistic and unifies all the aforementioned terms in a single objective function whose solution models the probability of relevance of samples when learning decision functions. Extensive experiments conducted on the highly imbalanced task of satellite image change detection show the effectiveness of the proposed model.

2. PROPOSED METHOD

Let $\mathcal{I}_r = \{p_1, \dots, p_n\}$, $\mathcal{I}_t = \{q_1, \dots, q_n\}$ be two satellite images captured at two different instants t_0 and t_1 with $t_0 < t_1$, $p_i, q_i \in \mathbb{R}^d$ (here $d = 30 \times 30 \times 3$, see experiments). \mathcal{I}_r , \mathcal{I}_t , referred to as reference and test images respectively, are assumed registered, i.e., pixels in pairs $\{(p_i, q_i)\}_i$ correspond

to the same locations. Now, we define $\mathcal{I} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, with $\mathbf{x}_i = (p_i, q_i)$, and $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ the underlying unknown labels. Our goal is to design a change detection function $f : \mathcal{I} \rightarrow \{-1, +1\}$ which predicts the unknown labels in $\{\mathbf{y}_i\}_i$ with $\mathbf{y}_i = +1$ if the test patch $q_i \in \mathcal{I}_t$ corresponds to a ‘‘change’’ w.r.t. its reference patch $p_i \in \mathcal{I}_r$ and $\mathbf{y}_i = -1$ otherwise. As ‘‘changes’’ are scarce, it is very reasonable to assume that $|\{\mathbf{x}_i : \mathbf{y}_i = +1\}| \ll |\{\mathbf{x}_i : \mathbf{y}_i = -1\}|$. Learning f requires a subset of training data annotated by an oracle¹. As these annotations are usually highly expensive, the design of f should consider *as few annotations as possible* while minimizing the generalization error $P(f(X) \neq Y)$.

2.1. Interactive satellite image change detection

Our design principle is iterative and relies on a question & answer model, which suggests the most informative display² to an oracle, collects annotations and updates the decision function f accordingly. Let $\mathcal{D}_t \subset \mathcal{I}$ be a subset of images shown to an oracle at iteration t and let \mathcal{Y}_t be the unknown labels of \mathcal{D}_t ; in practice $|\mathcal{D}_t|$ is fixed depending on the targeted annotation *budget*. We build our learning function f iteratively by asking the oracle ‘‘questions’’ about labels in \mathcal{D}_t according to the following steps

Display zero. Select a display \mathcal{D}_0 including representatives of \mathcal{I} . In practice, \mathcal{D}_0 corresponds to centroids of a partition $\{h_1, \dots, h_K\}$ of \mathcal{I} obtained with K-means clustering. Run the following steps for $t = 0, \dots, T - 1$ (T may also depend on a predefined annotation budget):

-Oracle model. Label display \mathcal{D}_t with an oracle function (denoted $\mathcal{C}(\cdot)$) and assign $\mathcal{C}(\mathcal{D}_t)$ to \mathcal{Y}_t .

-Learning model. Train a decision function $f_t(\cdot)$ on data labeled, so far, $\cup_{k=0}^t (\mathcal{D}_k, \mathcal{Y}_k)$ and use it to predict labels on test data depending on $f_t(\cdot)$. In practice, we use support vector machines (SVMs) in order to build/update $f_t(\cdot)$.

-Display model. Select the next display $\mathcal{D}_{t+1} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$ (to show to the oracle) that minimizes $P(f_{t+1}(X) \neq Y)$. It is clear that a brute force strategy that (i) considers all the possible displays $\mathcal{D} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$, (ii) learns the underlying classifiers $f_{t+1}(\cdot)$ on $\mathcal{D} \cup_{k=0}^t \mathcal{D}_k$ and (iii) evaluates their generalization error is highly combinatorial and out of reach. Display selection heuristics, related to active learning, are usually used instead, but one should be cautious when using these heuristics since many of them can perform worse than the basic display strategy consisting in choosing data uniformly randomly (see [31] and references within). Our alternative display model proposed in this paper, combines diversity and ambiguity as well as representativity; diversity aims to select data in order to discover different modes of $f_{t+1}(\cdot)$ while ambiguity seeks to locally refine the decision boundary of $f_{t+1}(\cdot)$. Details about our display model, shown

¹i.e., user or expert.

²subset of images.

subsequently, constitute the main contribution of this work.

2.2. Proposed display model

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

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

here $\circ, '$ are respectively the hadamard product and the matrix transpose, $\|\cdot\|_1$ 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, (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 $(\mathbf{F}_{i1}, \mathbf{F}_{i2}) = (\hat{f}_t(\mathbf{x}_i), 1 - \hat{f}_t(\mathbf{x}_i))$ and $\hat{f}_t \in [0, 1]$ being a normalized version of f_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 $\sum_i \sum_c \mu_i \mathbf{F}_{ic} \log \mathbf{F}_{ic}$) 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 and helps obtaining a closed form solution (as also shown subsequently). The impact of these terms is controlled by $\alpha, \beta, \gamma \geq 0$.

We formulate the minimization problem by adding an equality constraint and bounds which ensure a normalization of the relevance values and allow us to consider $\{\mu_i\}_i$ as a probability distribution on \mathcal{I} .

2.3. Optimization

Proposition 1 *Let $\mathbf{1}_{nc}, \mathbf{1}_K$ denote vectors of nc and K ones respectively. The optimality conditions of (1) lead to the solution*

$$\mu^{(\tau+1)} := \frac{\hat{\mu}^{(\tau+1)}}{\|\hat{\mu}^{(\tau+1)}\|_1}, \quad (2)$$

with $\hat{\mu}^{(\tau+1)}$ being

$$\exp\left(-\frac{1}{\gamma}[(\mathbf{D}\circ\mathbf{C})\mathbf{1}_K+\alpha\mathbf{C}(\log[\mathbf{C}'\mu^{(\tau)}+\mathbf{1}_K])+\beta(\mathbf{F}\circ\log\mathbf{F})\mathbf{1}_{nc}]\right). \quad (3)$$

In view of space, details of the proof are omitted and result from the gradient optimality conditions of Eq. (1).

Considering the above proposition, the optimal solution is obtained iteratively as a fixed point of Eqs (2) and (3) with $\hat{\mu}^{(0)}$ initially set to random values. Note that convergence is observed in practice in few iterations, and the underlying fixed point, denoted as $\tilde{\mu}$, corresponds to the most *relevant* samples in the display \mathcal{D}_{t+1} (according to criterion 1) used to train the subsequent classifier f_{t+1} (see also algorithm 1).

Algorithm 1: Display selection mechanism

Input: Images in \mathcal{I} , display $\mathcal{D}_0 \subset \mathcal{I}$, budget T , B .

Output: $\cup_{t=0}^{T-1}(\mathcal{D}_t, \mathcal{Y}_t)$ and $\{f_t\}_t$.

for $t := 0$ **to** $T - 1$ **do**

$\mathcal{Y}_t \leftarrow \mathcal{C}(\mathcal{D}_t)$; // Oracle model
 $f_t \leftarrow \arg \min_f P(f(X) \neq Y)$; // Learning model (built on top of $\cup_{k=0}^t(\mathcal{D}_k, \mathcal{Y}_k)$)

$\hat{\mu}^{(0)} \leftarrow \text{random}$; $\mu^{(0)} \leftarrow \frac{\hat{\mu}^{(0)}}{\|\hat{\mu}^{(0)}\|_1}$; $\tau \leftarrow 0$

while $(\|\mu^{(\tau+1)} - \mu^{(\tau)}\|_1 \geq \epsilon \wedge \tau < \text{maxiter})$ **do**

Set $\mu^{(\tau+1)}$ using Eqs. (2) and (3);

// Display model

$\tau \leftarrow \tau + 1$

$\tilde{\mu} \leftarrow \mu^{(\tau)}$

$\mathcal{D}_{t+1} \leftarrow \{\mathbf{x}_i \in \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k : \tilde{\mu}_i \in \mathcal{L}_B(\tilde{\mu})\}$;

// $\mathcal{L}_B(\tilde{\mu})$ being the B largest values of $\tilde{\mu}$

3. EXPERIMENTS

We evaluate the performances of our interactive change detection algorithm using a dataset of 2,200 non-overlapping patch pairs (of 30×30 pixels in RGB) taken from two registered (reference and test) GeoEye-1 satellite images of $2,400 \times 1,652$ pixels with a spatial resolution of 1.65m/pixel. These images correspond to the same area of Jefferson (Alabama) taken respectively in 2010 and in 2011 with many changes due to tornadoes (building destruction, etc.) and no-changes (including irrelevant ones as clouds). The underlying ground truth contains 2,161 negative patch pairs (no-changes and irrelevant ones) and only 39 positive patch pairs (relevant changes), so $< 2\%$ of these patches correspond to relevant changes; half of this set is used to build the display and the learning models and the other half for evaluation. Performances are reported using equal error rate (EER) on the eval set of \mathcal{I} . EER is the balanced generalization error that equally weights errors in "change" and "no-change" classes. Smaller EER implies better performance.

Iter Samp%	1	2	3	4	5	6	7	8	9	10
rep	48.05	26.21	12.72	10.48	9.88	9.70	8.52	8.85	8.61	8.82
div	48.05	31.24	23.45	30.41	44.81	24.12	13.22	17.02	6.88	7.98
amb	48.05	46.68	38.73	29.91	14.74	20.11	8.33	7.41	7.37	5.53
rep+div	48.05	26.21	33.35	25.10	21.55	11.71	2.84	1.65	1.59	1.43
rep+amb	48.05	26.21	12.62	10.81	9.82	9.70	8.53	9.23	8.60	8.82
div+amb	48.05	41.69	28.82	23.08	23.41	23.42	19.82	13.10	8.16	6.97
all	48.05	26.21	33.35	25.52	23.70	14.59	2.74	1.54	1.67	1.48

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 $t = 0, \dots, T - 1$ (Iter) and the underlying sampling rates (Samp) again defined as $(\sum_{k=0}^{t-1} |\mathcal{D}_k| / (|\mathcal{I}|/2)) \times 100$.

3.1. Ablation study

Initially, we study the impact of each term of our objective function when taken separately and jointly. Note that cardinality term is always kept as it acts as a regularizer that also defines the closed form solution shown earlier. Table 1 shows the impact of each of these terms individually, pairwise and all jointly taken. We observe that representativity+diversity are the most important criteria at the early stages of the iterative change detection process, while the impact of the ambiguity term comes later during the last iterations (when the modes of data are all explored) in order to locally refine change decision functions. These observations are shown through EER, and for different sampling rates defined at each iteration t as $(\sum_{k=0}^{t-1} |\mathcal{D}_k| / (|\mathcal{I}|/2)) \times 100$ with $|\mathcal{D}_k|$ set to 16 in practice and again $|\mathcal{I}| = 2, 200$.

3.2. Comparison

In order to further investigate the relevance of our method, we compare our display model against three different related strategies, namely *maxmin*, *uncertainty* as well as *random sampling*. Maxmin consists in sampling a display \mathcal{D}_{t+1} greedily at each iteration of the interactive process; each data in $\mathbf{x}_i \in \mathcal{D}_{t+1} \subset \mathcal{I} \setminus \cup_{k=0}^t \mathcal{D}_k$ is chosen in order to *maximize its minimum distance w.r.t.* $\cup_{k=0}^t \mathcal{D}_k$, thereby the display \mathcal{D}_{t+1} will correspond to the most distinct samples. Uncertainty sampling consists in choosing the display whose unlabeled samples are the most ambiguous (i.e., whose SVM scores are the closest to zero). Random consists in picking randomly samples from the pool of unlabeled training data. We also compare our method with fully-supervised learning where a monolithic classifier is trained using 100% of the ground truth of training data.

Figure 1 shows EER w.r.t different iterations (and the underlying sampling rates in table 1). We observe the positive impact of our display model compared to the other strategies. As the task is highly imbalanced, all these comparative strategies are unable to spot the rare class (changes) sufficiently; maxmin and random make it possible to capture the diversity of the data without being able to minimize the EER at the latest iterations. Uncertainty refines locally the decision functions but suffers from the lack of diversity. Whereas all these strategies (for $t \leq 1$) have high EERs, interactive

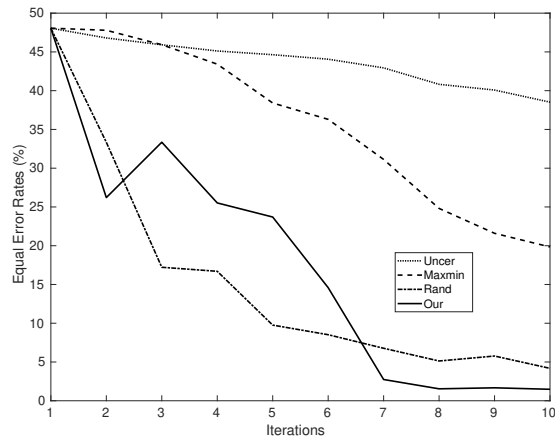


Fig. 1. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table 1 (Samp). Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER of 0.94%. See again section 3.2 for more details.

change detection (when combined with our display model) rapidly reduces the EER and overtakes all the other strategies, at the end of the iterative process. This comes from the rapid adaptation of decision functions $\{f_t(\cdot)\}_t$ to the oracle when frugal learning from the most relevant samples.

4. CONCLUSION

We introduce in this paper a novel interactive satellite image change detection algorithm. The proposed method is based on a query & answer model which suggests the most informative patch pairs to an oracle and according to the responses of the latter updates a decision criterion that captures the oracle’s intention. Our proposed display model is probabilistic and obtained by minimizing a constrained objective function mixing (i) diversity that explores different modes of data distribution, (ii) representativity which focuses on the most informative samples in these modes and (iii) ambiguity that returns samples with the most ambiguous classifications. We also consider a regularization term that smooths the learned probabilities and allows obtaining closed form solutions. Experiments conducted on the highly imbalanced task of satellite image change detection show the effectiveness of our proposed frugal learning approach.

5. REFERENCES

[1] D. Brunner et al., Earthquake damage assessment of buildings using vhr optical and sar imagery, *IEEE TGRS*, 48(5), 2403–2420, 2010.
 [2] H. Gokon et al., A method for detecting buildings destroyed by the 2011 tohoku earthquake and tsunami using multitemporal terrasars-x data, *GRSL*, vol. 12, no. 6, pp. 1277–1281, 2015.
 [3] 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.

[4] H. Sahbi, Coarse-to-fine deep kernel networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCVW)*, 2017.
 [5] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, Image change detection algorithms: A systematic survey, *IEEE TIP*, 14(3), 294–307, 2005.
 [6] S. Liu et al., Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images, *TGRS*, 53(8), 4363–4378, 2015.
 [7] H. Sahbi, Interactive satellite image change detection with context-aware canonical correlation analysis. *IEEE GRS Letters* 14(5), 607–611, 2017.
 [8] H. Sahbi, Relevance feedback for satellite image change detection. *IEEE ICASSP*, 1503–1507, 2013.
 [9] G. Chen et al., Object-based change detection, *IJRS*, vol. 33, no. 14, pp. 4434–4457, 2012.
 [10] J. Zhu et al., Reducing mis-registration and shadow effects on change detection in wetlands, *PERS*, 77(4), 325–334, 2011.
 [11] N. Bourdis, D. Marraud, and H. Sahbi, Camera pose estimation using visual servoing for aerial video change detection, in *IGARSS*, 2012.
 [12] A. Fournier et al., 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 et al., Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements, *TGRS*, 45(6), 1529–1558, 2007.
 [15] N. Bourdis, D. Marraud, and H. Sahbi, Constrained optical flow for aerial image change detection, in *IGARSS*, 2011, pp. 4176–4179.
 [16] Pollard, Comprehensive 3d change detection using volumetric appearance modeling, Phd, Brown University, 2009.
 [17] A. Nielsen, The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data, *IEEE TIP*, 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 seg., *IJRS*, 29(2), 399–423, 2008.
 [21] O. Vinyals et al. Matching networks for one shot learning. *arXiv preprint arXiv:1606.04080*, 2016.
 [22] M. Jiu and H. Sahbi. Deep kernel map networks for image annotation. *IEEE ICASSP*, 2016.
 [23] B. Settles. Active learning. *Synthesis lectures on artificial intelligence and machine learning*, 1-114, 2012.
 [24] H. Tianxu et al. An active learning approach with uncertainty, representativeness, and diversity. *The Scientific World Journal*, 2014.
 [25] Joshi et al. Multi-class active learning for image classification. *IEEE CVPR*, 2009.
 [26] S. Dasgupta. Analysis of a greedy active learning strategy. In *NIPS*, 337-344, 2005.
 [27] Houlshby et al., Bayesian active learning for classification and preference learning. *ArXiv:1112.5745*, 2011.
 [28] Campbell & Broderick, Automated scalable Bayesian inference via Hilbert coresets. *JMLR*, 551–588, 2019.
 [29] Gal et al., Deep bayesian active learning with image data. *ICML*, 2017.
 [30] Pang et al., Meta-learning transferable active learning policies by deep reinforcement learning. In *arXiv:1806.04798*, 2018.
 [31] F.R. Bach. Active learning for misspecified generalized linear models. *Advances in Neural Information Processing Systems (NIPS)*, 19, 2006.
 [32] A. Kolesnikov, X. Zhai, L. Beyer. Revisiting Self-Supervised Visual Representation Learning. In *IEEE/CVF CVPR*, pp. 1920–1929, 2019.
 [33] H. Sahbi. A particular Gaussian mixture model for clustering and its application to image retrieval. *Soft Computing*, 667-676, 2008.
 [34] M. Jiu and H. Sahbi. Semi supervised deep kernel design for image annotation. *IEEE ICASSP*, 2015.