Label-frugal satellite image change detection with generative virtual exemplar learning

Hichem Sahbi

Sorbonne University, CNRS, LIP6, F-75005, Paris, France

Abstract

Change detection is a major task in remote sensing which consists in finding all the occurrences of changes in multi-temporal satellite or aerial images. The success of existing methods, and particularly deep learning ones, is tributary to the availability of hand-labeled training data that capture the acquisition conditions and the subjectivity of the user (oracle). In this paper, we devise a novel change detection algorithm, based on active learning. The main contribution of our work resides in a new model that measures how important is each unlabeled sample, and provides an oracle with only the most critical samples (also referred to as virtual exemplars) for further labeling. These exemplars are generated, using an invertible graph convnet, as the optimum of an adversarial loss that (i) measures representativity, diversity and ambiguity of the data, and thereby (ii) challenges (the most) the current change detection criteria, leading to a better re-estimate of these criteria in the subsequent iterations of active learning. Extensive experiments show the positive impact of our label-efficient learning model against comparative methods.

1 Introduction

Satellite image change detection is nowadays becoming a hotspot in remote sensing with applications ranging from anthropogenic activity monitoring to phenology mapping, and natural hazard damage assessment [2]–[4], [31]. However, this task is known to be challenging due to the pervasive changes in satellite images resulting from sensors and acquisition conditions (weather variations, radiometric changes, scene intrinsic content, etc). Early work relies on comparisons of multi-temporal series by analyzing how changes unfold over time using image differences and thresholding, spectral vegetation indices, principal component and change vector analysis [6], [7], [9]–[12]. Other methods either (i) rely on normalization as a preprocessing step that minimizes the effects of variations due to atmospheric conditions (e.g., clouds, haze,...) and sensor calibration [14], [15], [17], [19], [44], or (ii) model the expected appearance of relevant changes (objects or regions of interest over time) using statistical machine learning [5], [20]–[27], [40] while accounting for irrelevance changes.

The success of machine learning models, particularly deep neural networks [1], [149], depends on the availability of large collections of hand-labeled reference images that capture the diversity of relevant and irrelevant changes [13], [16]. These collections should also capture the user's targeted changes. However, hand-labeling large data collections is time-consuming and impractical, and even when relatively tractable, may suffer from domain-shift and may also struggle to capture the nuances of the user's subjectivity and intention. Several works explore solutions to make machine learning more efficient and less reliant on large collections of labeled data including few-shot and transfer learning [28], [29], as well as semi/self-supervised learning [41], [42]. Nonetheless, these methods often lack the ability to directly incorporate and understand

the user's intention. Solutions based on active learning [30], [33]–[39], [114], [143] are rather more appropriate and consist in efficiently probing the user about the relevance of *only the most informative* observed changes prior to train decision criteria that best suit the user's intention and scene acquisition conditions.

In this paper, we propose a new label-efficient satellite image change detection algorithm based on invertible graph convnets. The proposed model is interactive and queries the oracle about the labels of the *most critical* data (also dubbed as virtual exemplars) whose positive impact on the trained convnets is the most important. These virtual exemplars are designed as the most representative, diverse and ambiguous data that optimize an adversarial loss. This loss is optimized while learning an invertible graph convnet that achieves both classification and generation so as *to constrain the synthesized virtual exemplars to lie on a nonlinear manifold enclosing both "change" and "no-change" classes*. Note that the solution presented in this work, despite being adversarial, is conceptually different from the ones widely used in generative adversarial networks (GANs) [43]. Indeed, whereas GANs aim at generating *fake* data that mislead the trained discriminators, our formulation seeks instead to generate *critical* data — for annotation — which are the most impactful on the subsequent learned classifiers. Put differently, the proposed method allows to sparingly query the oracle *only* on the most representative, diverse and uncertain data which challenge the current discriminator, and ultimately lead to more accurate ones in the subsequent iterations of change detection. Extensive change detection experiments show the effectiveness of our invertible graph convnets and virtual exemplar learning models against comparative methods.

2 Proposed method

Given two satellite images $\mathcal{I}_r = \{p_1, \ldots, p_n\}$, $\mathcal{I}_t = \{q_1, \ldots, q_n\}$ as a collection of registered patch-pairs taken at two different instants t_0, t_1, with p_i and $q_i \in \mathbb{R}^d$. Our goal is to train a convnet f that predicts the unknown labels $\{\mathbf{y}_i\}_i$ in $\mathcal{I} = \{\mathbf{x}_i = (p_i, q_i)\}_i$ with $\mathbf{y}_i = 1$ iff $\mathbf{x}_i$ corresponds to a change pair, and $\mathbf{y}_i = 0$ otherwise. As training f requires patch-pairs (hand-labeled by an oracle), we seek to make f as *label-efficient* as possible and also accurate.

2.1 A glimpse on interactive change detection

Our change detection algorithm relies on a question & answer (Q&A) process that iteratively probes the oracle about the labels of the most *critical patch-pairs*, and then updates change detection criteria accordingly. The most critical patch-pairs, at iteration t, constitute a *display* $\mathcal{D}_t$ whose unknown labels are denoted as $\mathcal{Y}_t$. As shown subsequently, our algorithm trains change detection criteria $f_0, \ldots, f_{T-1}$ iteratively, starting from a random display $\mathcal{D}_0$, and according to the following steps

i) Gather $\mathcal{Y}_t$ from the oracle and train $f_t(.)$ on $\cup_{k=0}^{t}(\mathcal{D}_k, \mathcal{Y}_k)$ using graph convolution networks (GCNs) (see section 2.3).
ii) Select the subsequent display $\mathcal{D}_{t+1} \subset \mathcal{I} \backslash \cup_{k=0}^{t} \mathcal{D}_k$ to show to the oracle. A strategy considering all configurations of displays $\mathcal{D} \subset \mathcal{I} \backslash \cup_{k=0}^{t} \mathcal{D}_k$, prior to training the underlying criterion $f_{t+1}(.)$ on $\mathcal{D} \cup_{k=0}^{t} \mathcal{D}_k$, and keeping only the display $\mathcal{D}$ with the highest positive impact on $f_{t+1}(.)$'s accuracy, is *clearly intractable*. Besides, this requires labeling each of these configurations by the oracle. Our contribution introduced in this paper (see 2.2, 2.4) is based instead on active learning strategies which are rather more appropriate. However, one should be cautious as many of these strategies are found to be equivalent to (or worse than) basic random data selection strategies (as also

discussed in [114] and references therein).

Considering the aforementioned goal, the main contribution of this paper includes a novel display selection strategy that finds in a *flexible way* virtual exemplars instead of using rigid unlabeled data (in contrast to [65]). The design principle of our method allows selecting the most diverse, representative, and uncertain data that challenge (the most) the current change detection criteria and leads to a better re-estimate of these criteria in the subsequent iterations of active learning. Besides, our contribution also includes a novel invertible GCN design that achieves both classification and virtual exemplar generation more effectively as shown through extensive experiments on the challenging task of interactive satellite image change detection.

2.2 Virtual exemplar design

As labeling images in $\mathcal{I}$ is highly prohibitive, one should focus only on the most critical samples. The latter define the subsequent display $\mathcal{D}_{t+1}$, also referred to as virtual exemplars $\{\mathbf{V}_k\}_{k=1}^K$, used to train $f_{t+1}(.)$ on $\mathcal{D}_{t+1} \cup \cdots \cup \mathcal{D}_0$. We consider for each training sample $\mathbf{x}_i \in \mathcal{I}$ a conditional probability distribution $\{\mu_{ik}\}_k$ that measures the membership of $\mathbf{x}_i$ to the K-virtual exemplars in $\{\mathbf{V}_k\}_{k=1}^K$. These memberships $\mu = \{\mu_{ik}\}_{ik}$ together with the underlying virtual exemplars are found by minimizing the following objective function

$$\begin{array}{ll} \min\limits_{\mathbf{V};\mu\in\Omega} & \mathbf{tr}\big(\mu\ d(\mathbf{V},\mathbf{X})^\top\big)\ +\ \alpha\ [\mathbf{1}_n^\top\mu]\log[\mathbf{1}_n^\top\mu]^\top \\ & +\ \beta\ \mathbf{tr}\big(f(\mathbf{V})^\top\ \log f(\mathbf{V})\big)\ +\ \gamma\ \mathbf{tr}(\mu^\top\log\mu), \end{array} \tag{1}$$

being $\Omega = \{\mu : \mu \geq 0; \mu\mathbf{1}_K = \mathbf{1}_n\}$ a convex set that constrains μ to be row-stochastic, and $\mathbf{1}_K$, $\mathbf{1}_n$ are two vectors of K and n ones respectively. In Eq. 1, $\mathbf{tr}(.)$ is the matrix trace operator, $\mu \in \mathbb{R}^{n\times K}$ is a learned matrix whose i-th row corresponds to the conditional probability of assigning $\mathbf{x}_i$ to each of the K learned virtual exemplars in $\mathbf{V} \in \mathbb{R}^{d\times K}$, and $\log$ is applied entry-wise. In the above equation, the matrix $d(\mathbf{V},\mathbf{X}) \in \mathbb{R}^{K\times n}$ captures the euclidean distances between the virtual exemplars in $\mathbf{V}$ and the input data in $\mathbf{X}$ whilst $f(\mathbf{V}) \in \mathbb{R}^{\#Classes\times K}$ corresponds to the softmax layer. The first term in Eq. 1 measures the representativity of the virtual exemplars by constraining them to be *close* to their assigned input data in $\mathcal{I}$ and thereby inheriting more accurate labels when the oracle annotates their closest input data. The second term in Eq. 1 models diversity by maximizing the spread of the probability distribution of samples across the learned virtual exemplars; this term reaches its minimum when virtual exemplars attract *diversely* input data in $\mathcal{I}$. The third term measures the *ambiguity* (or uncertainty) in $\mathbf{V}$ as the negative of the softmax entropy, and it reaches its smallest value when virtual exemplars in $\mathbf{V}$ are evenly scored w.r.t different classes. Finally, the fourth term acts as a regularizer which considers that without any a priori about the three other terms, the membership distribution $\{\mu_{ik}\}_k$ should be uniform for each sample $\mathbf{x}_i$. All these terms are combined using $\alpha, \beta, \gamma \geq 0$.

In order to solve the optimization problem in Eq. 1 for each change detection cycle t, we consider a bi-level optimization procedure that fixes the display $\mathbf{V}$ and finds μ, and vice versa. One may show that Eq. (1) admits the following solution

$$\begin{array}{lll} \mu^{(\tau+1)} & := & \mathbf{diag}\big(\hat{\mu}^{(\tau+1)}\mathbf{1}_K\big)^{-1}\ \hat{\mu}^{(\tau+1)} \\ \mathbf{V}^{(\tau+1)} & := & \hat{\mathbf{V}}^{(\tau+1)}\ \mathbf{diag}\big(\mathbf{1}_n^\top\mu^{(\tau)}\big)^{-1}, \end{array} \tag{2}$$

being $\hat{\mu}^{(\tau+1)}$, $\hat{\mathbf{V}}^{(\tau+1)}$ respectively

$$\begin{array}{l} \exp\big(-\frac{1}{\gamma}[d(\mathbf{X},\mathbf{V}^{(\tau)}) + \alpha(\mathbf{1}_n\mathbf{1}_K^\top + \mathbf{1}_n\log\mathbf{1}_n^\top\mu^{(\tau)})]\big), \\ \\ \mathbf{X}\ \mu^{(\tau)} + \beta\sum_c \nabla_v f_c(\mathbf{V}^{(\tau)}) \circ (\mathbf{1}_d\ [\log f_c(\mathbf{V}^{(\tau)})]^\top + \mathbf{1}_d\mathbf{1}_K^\top), \end{array} \tag{3}$$

here **diag**(.) maps a vector to a diagonal matrix and $\circ$ is the Hadamard matrix product. Due to space limitation, details of the proof are omitted and result from the optimality conditions of Eq. 1's gradient. Initially, $\mu^{(0)}$ and $\mathbf{V}^{(0)}$ are set to random values and the procedure converges to an optimal solution in few iterations, defining the most relevant virtual exemplars of $\mathcal{D}_{t+1}$ used to train the subsequent GCN f_{t+1} (see algorithm 1).

Algorithm 1: Virtual exemplar learning

Input: $\mathcal{I}$, $\mathcal{D}_0 \subset \mathcal{I}$, budget T.
Output: $\cup_{t=0}^{T-1}(\mathcal{D}_t, \mathcal{Y}_t)$ and $\{f_t\}_t$.

for $t := 0$ **to** $T-1$ **do**
- $\mathcal{Y}_t \leftarrow$ oracle$(\mathcal{D}_t)$;
- $f_t \leftarrow \arg\min_f$ CrossEntropyLoss$(f, \cup_{k=0}^{t}(\mathcal{D}_k, \mathcal{Y}_k))$;
- $\tau \leftarrow 0$; $\hat{\mu}^{(0)} \leftarrow$ random; $\hat{\mathbf{V}}^{(0)} \leftarrow$ random;
- Set $\mu^{(0)}$ and $\mathbf{V}^{(0)}$ using Eqs. (2) and (3);
- **while** ($\|\mu^{(\tau+1)} - \mu^{(\tau)}\|_1 + \|\mathbf{V}^{(\tau+1)} - \mathbf{V}^{(\tau)}\|_1 \geq \epsilon$ $\wedge\, \tau <$ *maxiter*) **do**
  - Set $\mu^{(\tau+1)}$ and $\mathbf{V}^{(\tau+1)}$ using Eqs. (2) and (3);
  - $\tau \leftarrow \tau + 1$;
- $\tilde{\mu} \leftarrow \mu^{(\tau)}$; $\tilde{\mathbf{V}} \leftarrow \mathbf{V}^{(\tau)}$;
- $\mathcal{D}_{t+1} \leftarrow \left\{\mathbf{x}_i \in \mathcal{I} \setminus \cup_{k=0}^{t} \mathcal{D}_k : \mathbf{x}_i \leftarrow \arg\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{V}_k\|_2^2\right\}_{k=1}^{K}$.

2.3 Graph convnets

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as a set of nodes $\mathcal{V}$ and edges $\mathcal{E}$ associated to a given patch-pair $\mathbf{x} \in \mathcal{I}$. The graph $\mathcal{G}$ is endowed with a signal[1] and associated with a learnable adjacency matrix $\mathbf{A}$. Graph convnets (GCNs) seek to learn a set of C filters $\mathcal{F}$ that define convolution on n nodes of $\mathcal{G}$ (with $n = |\mathcal{V}|$) as $(\mathcal{G} \star \mathcal{F})_{\mathcal{V}} = g_2\big(g_1(\mathbf{A}\ \mathbf{U}^\top)\mathbf{W}\big)$, here $\mathbf{U} \in \mathbb{R}^{s \times n}$ is the graph signal, $\mathbf{W} \in \mathbb{R}^{s \times C}$ is the matrix of convolutional parameters corresponding to the C filters and g_1, g_2 are nonlinear activations applied entry-wise. In $(\mathcal{G} \star \mathcal{F})_{\mathcal{V}}$, the input signal $\mathbf{U}$ is first mapped using $\mathbf{A}$ and this defines for each node u, the set of neighbor aggregates. Beside learning the convolution parameters in $\mathbf{W}$, entries of $\mathbf{A}$ are also learned in order to capture an "optimal" graph topology when achieving aggregation so as $(\mathcal{G} \star \mathcal{F})_{\mathcal{V}}$ models both attention and convolution layers; the first layer aggregates signals in $\mathcal{N}(\mathcal{V})$ (sets of node neighbors of $\mathcal{V}$) by multiplying $\mathbf{U}$ with $\mathbf{A}$ while the second layer achieves convolution by multiplying the resulting aggregates with the C filters in $\mathbf{W}$. Stacking one or multiple attention+convolution layers $(\mathcal{G} \star \mathcal{F})_{\mathcal{V}}$ together with fully connected and softmax layers defines the whole architecture of our GCNs. In the remainder of this paper, we go one step further, and make these GCNs bijective, i.e., invertible. As shown subsequently, this property is valuable, as it allows us *to learn the virtual exemplars in a mapping (latent) space while capturing the topology of their manifold in the input (ambient) space.* This turns out to be more effective when learning display models as shown later in experiments.

2.4 Invertible graph convnet design

Subsequently, we subsume the aforementioned GCN architecture as a multi-layered neutral network $f_{t,\theta}$ with t being the active learning cycle, $\theta = \{\mathbf{W}_1, \ldots, \mathbf{W}_L\}$, L its depth, $\mathbf{W}_\ell \in \mathbb{R}^{d_{\ell-1} \times d_\ell}$ its ℓ^{th}-layer weight tensor, and d_ℓ the output dimension of the ℓ^{th}-layer. In what follows, we omit both θ and t in the definition of $f_{t,\theta}$ and we rewrite the output of a given layer ℓ of this GCN as $\phi^\ell = g_\ell(\mathbf{W}_\ell^\top \phi^{\ell-1})$, $\ell \in \{2, \ldots, L\}$, being g_ℓ an activation function; without a loss of generality, we also omit the bias in the definition of ϕ^ℓ. Provided that the matrices in θ are invertible, and the

1. In change detection, this signal is the difference between patch-pair pixel values.

activation functions $\{g_\ell\}_\ell$ bijective[2], one may guarantee that the trained GCN is also bijective and henceby invertible. Besides, the dimensionality of all the layers remains constant[3]. This bijection property is valuable, as it reduces the complexity of solving Eq. 1 by finding virtual exemplars in the latent space instead of the ambient space. On another hand, by the invertibility of the GCN, this guarantees that the inverted latent exemplars belong to the nonlinear manifold/distribution enclosing data in the ambient space.

By considering the matrix of virtual exemplars, denoted as $\mathbf{Z}$ in the latent space, and by the invertibility of f (i.e., $f(\mathbf{V}) = f(f^{-1}(\mathbf{Z})) = \mathbf{Z}$), Eq. 1 can be rewritten as

$$\begin{aligned}\min_{\mathbf{Z};\mu\in\Omega} \quad & \mathbf{tr}\big(\mu\, d(f^{-1}(\mathbf{Z}),\mathbf{X})^\top\big) \;+\; \alpha\, [\mathbf{1}_n^\top \mu] \log[\mathbf{1}_n^\top \mu]^\top \\ & +\; \beta\, \mathbf{tr}\big(\mathbf{Z}^\top \log \mathbf{Z}\big) \;+\; \gamma\, \mathbf{tr}(\mu^\top \log \mu).\end{aligned} \tag{4}$$

This objective function can be solved w.r.t. $\mathbf{Z}$ (similarly to Eqs. 1 and 2) using the fixed-point iteration process, while recovering $\mathbf{V} = f^{-1}(\mathbf{Z})$ thanks to the invertible GCN. Hence, this surrogate problem allows exploring nonlinear manifolds in the ambient space while being tractable in the latent space.

2.5 Regularization

The success of the generative properties of the invertible GCN f^{-1} is reliant on the stability of f. In other words, when f is M-Lipschitzian (with $M \approx 1$), the network f^{-1} will also be M-Lipschitzian (with $M \approx 1$) [152], so any slight update of virtual exemplars in the latent space (with the fixed-point iteration) will also result into a slight update in these exemplars in the ambient space when applying f^{-1}. This eventually leads to stable virtual exemplar generation in the ambient space, i.e., they follow the actual distribution of data manifold. As the Lipschitz constant of f is $M = \prod_\ell \|\mathbf{W}_\ell\|_2 . |g'_\ell|$, the sufficient conditions that guarantee that both f and f^{-1} are M-Lipschitzian (with $M \approx 1$) corresponds to (i) $\|\mathbf{W}_\ell\|_2 \approx 1$, and (ii) $|g'_\ell| \approx 1$ for all ℓ. Hence, by design, conditions (i)+(ii) could be satisfied by choosing the slope of the activation functions to be close to one (in practice to 0.99 and 0.95 respectively for the positive and negative parts of the leaky ReLU), and also by constraining the norm of all the weight matrices to be *orthonormal* which also guarantees their invertibility. This is obtained by adding a regularization term, to the cross-entropy (CE) loss, when training GCNs, as

$$\min_{\{\mathbf{W}_\ell\}_\ell} \mathrm{CE}(f;\{\mathbf{W}_\ell\}_\ell) + \lambda \sum_\ell \big\|\mathbf{W}_\ell^\top \mathbf{W}_\ell - \mathbf{I}\big\|_F, \tag{5}$$

here $\mathbf{I}$ stands for identity, $\|.\|_F$ denotes the Frobenius norm and $\lambda > 0$ (with $\lambda = \frac{1}{d}$ in practice); in particular, when $\mathbf{W}_\ell^\top \mathbf{W}_\ell - \mathbf{I} = 0$, then $\mathbf{W}_\ell^{-1} = \mathbf{W}_\ell^\top$ and $\|\mathbf{W}_\ell\|_2 = \|\mathbf{W}_\ell^{-1}\|_2 = 1$. With the aforementioned formulation, the learned GCNs are guaranteed to be discriminative, stable and invertible.

3 Experiments

We evaluate change detection performances using the Jefferson dataset which consists of $2,200$ non-overlapping and registered patch-pairs with each patch including 30×30 pixels in RGB. These registered pairs correspond to a large area of Jefferson (Alabama) captured by two (bi-temporal) GeoEye-1 satellite images of $2,400 \times 1,652$ pixels taken at two different instants (in

2. In practice, $\{g_\ell\}_\ell$ are chosen as leaky-ReLU activations which are bijective.

3. Excepting the softmax layer whose dimensionality depends on the number of classes. Nonetheless a simple trick consists in adding fictitious softmax outputs to match any targeted dimensionality.

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) again defined as $(\sum_k^t |\mathcal{D}_k|/(|\mathcal{I}|/2)) \times 100$. The AUC (Area Under Curve) corresponds to the average of EERs across iterations.

rep	div	amb	2	3	4	5	6	7	8	9	10	AUC.
✗	✗	✓	27.29	11.15	7.97	8.18	7.31	7.97	7.94	7.50	7.90	10.35
✗	✓	✗	18.72	11.24	7.97	8.18	7.29	7.59	7.88	7.50	7.90	9.36
✓	✗	✗	35.98	16.86	6.52	4.98	2.67	2.03	1.80	1.45	1.30	8.17
✓	✗	✓	40.40	23.86	9.56	7.65	5.75	5.47	6.12	4.40	5.72	12.10
✗	✓	✓	27.29	11.15	7.97	8.18	7.31	7.97	7.94	7.50	7.90	10.35
✓	✓	✗	29.84	17.63	6.21	4.40	2.70	1.98	1.92	1.65	1.52	7.53
✓	✓	✓	27.61	11.76	5.74	2.95	2.39	1.89	1.61	1.55	1.34	**6.31**
Samp%			2.90	4.36	5.81	7.27	8.72	10.18	11.63	13.09	14.54	-

2010 and in 2011) with a spatial resolution of 1.65m/pixel. These images capture a few relevant changes (road network damages, building destruction, etc., due to tornadoes that hit this area in 2011) together with no-changes (irrelevant ones such occlusions due to clouds, and radiometric variations). In total, the Jefferson dataset includes 39 positive pairs (relevant changes) and 2,161 negative pairs, so less than 2% of the observe area correspond to relevant changes, and this makes relevant changes rare and difficult to find. In all experiments, half of the dataset is used to train our display and GCN models whilst the other half is used for evaluation. As the two (positive and negative) classes are highly imbalanced, we score the performance of change detection, on the eval set, using the equal error rate (EER). In the observed results, smaller EERs imply better performances.

3.1 Ablation Study

In this section, we study the impact of different terms of our objective function individually, pairwise and all jointly taken. For all the settings, the last term of Eq. 1 is kept as it allows obtaining the closed form in Eq. 2. Table 1 shows the impact of each of these terms and their combination; from these results, we observe an important impact of representativity and diversity particularly at the earliest change detection iterations. The impact of ambiguity is more noticeable in the later iterations as it further refines change detection criteria when all the modes of "change" and "no-change" distribution are explored. The EERs are shown for increasing sampling percentages defined, at each iteration t, as $(\sum_k^t |\mathcal{D}_k|/(|\mathcal{I}|/2)) \times 100$ with again $|\mathcal{I}| = 2,200$ and $|\mathcal{D}_k|$ set to 16.

3.2 Comparison

We further compare the strength of our virtual exemplar (display) model against other display selection strategies such as *random, maxmin and uncertainty*. Random sampling consists in taking a random subset from the pool of unlabeled training data while uncertainty aims at picking the display whose classifier scores are the most ambiguous (i.e., similar through classes) on the same pool. Maxmin consists in greedily selecting data in $\mathcal{D}_{t+1}$ where each data $\mathbf{x}_i \in \mathcal{D}_{t+1} \subset \mathcal{I} \backslash \cup_{k=0}^t \mathcal{D}_k$ is taken by *maximizing its minimum distance w.r.t.* $\cup_{k=0}^t \mathcal{D}_k$, resulting into $\mathcal{D}_{t+1}$ with the most distinct samples. We also compare our sampling method against the closely related method [65] which consists in defining relevance measures on the whole unlabeled set and picking the display with the highest relevance. As an upper bound, we also show performances using the fully supervised setting which trains a single classifier on the full training set whose labels are taken from the

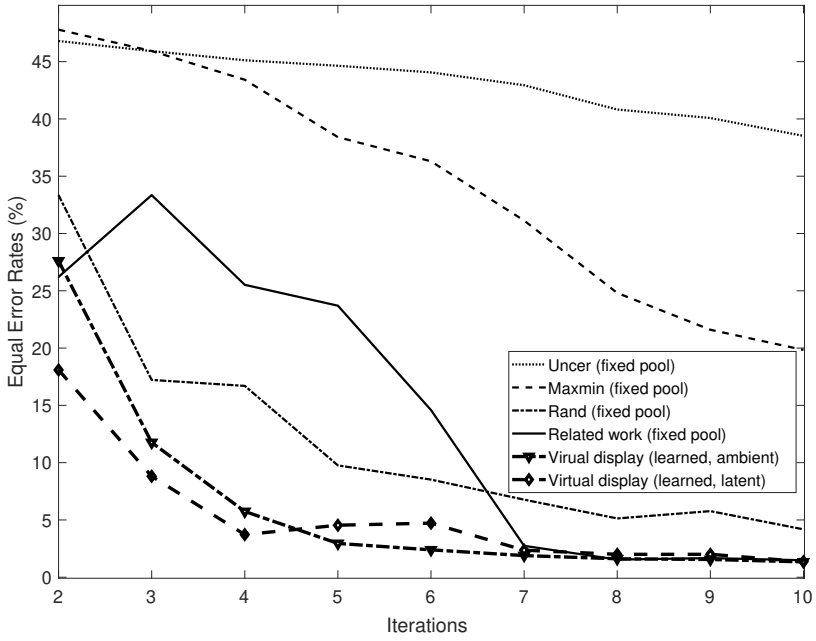

Fig. 1. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table 1 (Samp). Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER of 0.94%. Related work stands for the method in [65]; see again section 3.2 for more details.

ground-truth. Figure 1 shows the positive impact of the proposed display model against the aforementioned sampling strategies across increasing iterations (and sampling rates). From these observations, most of the comparative methods (excepting [65]) are unable to find the change class accurately; at the early iterations, both random and maxmin capture the diversity without being capable of sufficiently refining change detection results at the subsequent iterations, whereas uncertainly allows us to refine change detection results but without enough diversity. The method in [65] has the advantage of gathering the advantage of all these comparative strategies, but suffers at some extent from the rigidity of the selected display (which is taken from a fixed pool). In contrast to all these methods, our virtual exemplar model allows us to learn flexible displays particularly when these displays are constrained in the latent space with our invertible GCNs, particularly at frugal labeling regimes.

4 Conclusion

In this work, we introduce a new interactive change detection technique built on top of active learning. Our flexible display models constrained by our invertible convnets allow training highly label-efficient change detection criteria. The adversarial design of our display model enables effective selection of the most informative (representative, diverse and ambiguous) data that challenge (the most) the current change detection criteria and further improve the subsequent ones. Extensive experiments conducted on the challenging task of satellite image change detection show that the proposed virtual display model outperforms other sampling strategies.

REFERENCES

[1] P. Vo and H. Sahbi. "Transductive kernel map learning and its application to image annotation." BMVC. 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] Wen, Y. et al. GCD-DDPM: A generative change detection model based on difference-feature guided DDPM. IEEE Transactions on Geoscience and Remote Sensing, 2024

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

[6] 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.

[7] 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.

[8] M. Jiu and H. Sahbi. Nonlinear deep kernel learning for image annotation. IEEE Transactions on Image Processing 26 (4), 1820-1832.

[9] 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, 53(8), pp. 4363–4378, 2015.

[10] 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.

[11] LI, Liangliang, MA, Hongbing, ZHANG, Xueyu, et al. Synthetic aperture radar image change detection based on principal component analysis and two-level clustering. Remote Sensing, vol. 16, no 11, p. 1861, 2024.

[12] LISTIANI, Indira Aprilia, ZANETTI, Massimo, et BOVOLO, Francesca. Time Series Directional Change Vector Analysis. In IEEE IGARSS 2024-2024, p. 8683-8686, 2024.

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

[14] 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.

[15] 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

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

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

[18] F. Yuan, G-S. Xia, H. Sahbi, V. Prinet. Mid-level features and spatio-temporal context for activity recognition. Pattern Recognition 45 (12), 4182-4191

[19] 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.

[20] Cheng, Guangliang, et al. "Change detection methods for remote sensing in the last decade: A comprehensive review." Remote Sensing 16.13: 2355., 2024.

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

[22] 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.

[23] 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

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

[25] L. Wang, H. Sahbi. Bags-of-daglets for action recognition. In IEEE ICIP 2014.

[26] J. Im, J. Jensen, and J. Tullis, Object-based change detection using correlation image analysis and image seg, IJRS, 29(2), 399–423, 2008.

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

[28] Qiu, Chunping, et al. "Few-shot remote sensing image scene classification: Recent advances, new baselines, and future trends." ISPRS Journal of Photogrammetry and Remote Sensing 209 (2024): 368-382.

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

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

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

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

[33] Tianxu et al., An Active Learning Approach with Uncertainty, Representativeness, and Diversity,

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

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

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

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

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

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

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

[41] Zhao, Maofan, et al. "Beyond Pixel-Level Annotation: Exploring Self-Supervised Learning for Change Detection With Image-Level Supervision." IEEE Transactions on Geoscience and Remote Sensing (2024).

[42] A. Kolesnikov, X. Zhai, L. Beyer. Revisiting Self-Supervised Visual Representation Learning. IEEE/CVF CVPR, 2019, pp. 1920-1929

[43] Creswell, Antonia, et al. "Generative adversarial networks: An overview." IEEE Signal Processing Magazine 35.1 (2018): 53-65.

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

[45] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009

[46] H. Sahbi and N. Boujemaa. "Robust matching by dynamic space warping for accurate face recognition." Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205). Vol. 1. IEEE, 2001.

[47] Sabrina Tollari, Philippe Mulhem, Marin Ferecatu, Hervé Glotin, Marcin Detyniecki, Patrick Gallinari, H. Sahbi, and Zhong-Qiu Zhao. "A comparative study of diversity methods for hybrid text and image retrieval approaches." In Workshop of the Cross-Language Evaluation Forum for European Languages, pp. 585-592. Springer, Berlin, Heidelberg, 2008.

[48] H. Sahbi. A particular Gaussian mixture model for clustering and its application to image retrieval. Soft Computing 12 (7), 667-676

[49] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need. arXiv:1706.03762. 2017.

[50] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.

[51] A. Mazari and H. Sahbi. "MLGCN: Multi-Laplacian graph convolutional networks for human action recognition." The British Machine Vision Conference (BMVC). 2019.

[52] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[53] M. Ferecatu and H. Sahbi. "TELECOM ParisTech at ImageClefphoto 2008: Bi-Modal Text and Image Retrieval with Diversity Enhancement." CLEF (Working Notes). 2008.

[54] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-first AAAI conference on artificial intelligence. 2017.

[55] Clemens-Alexander Brust, Christoph Kading, and Joachim Denzler. Active learning for deep object detection. arXiv preprint arXiv:1809.09875, 2018.

[56] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. Neurocomputing, 312:135–153, 2018.

[57] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data aug- mentation for deep learning. Journal of Big Data, 6(1):1–48, 2019.

[58] Keze Wang, Liang Lin, Xiaopeng Yan, Ziliang Chen, Dongyu Zhang, and Lei Zhang. Cost-effective object detection: Active sample mining with switchable selection criteria. CoRR, abs/1807.00147, 2018.

[59] Vladimir Haltakov, Christian Unger, and Slobodan Ilic. Framework for gen- eration of synthetic ground truth data for driver assistance applications. In German conference on pattern recognition, pages 323–332. Springer, 2013.

[60] H. Sahbi, Jean-Yves Audibert, Jaonary Rabarisoa, and Renaud Keriven. "Context-dependent kernel design for object matching and recognition." In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8. IEEE, 2008.

[61] Begum Demir, Claudio Persello, and Lorenzo Bruzzone. Batch-mode active-learning methods for the interactive classification of remote sens- ing images. IEEE Transactions on Geoscience and Remote Sensing, 49(3):1014–1031, 2010.

[62] H. Sahbi, Jean-Yves Audibert, and Renaud Keriven. "Graph-cut transducers for relevance feedback in content based image retrieval." 2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007.

[63] Krause, Andreas, and Carlos Guestrin. "Nonmyopic active learning of gaussian processes: an exploration-exploitation approach." Proceedings of the 24th international conference on Machine learning. 2007.

[64] M. Jiu and H. Sahbi. "Deep representation design from deep kernel networks." Pattern Recognition 88 (2019): 447-457.

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

[66] Robert Pinsler, Jonathan Gordon, Eric T. Nalisnick, and Jose Miguel Hernandez-Lobato. Bayesian batch active learning as sparse subset approximation. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alche Buc, Emily B. Fox, and Roman Garnett, editors, NeurIPS, pages 6356–6367, 2019.

[67] S. Thiemert, H. Sahbi, and M. Steinebach. "Applying interest operators in semi-fragile video watermarking." Security, Steganography, and Watermarking of Multimedia Contents VII. Vol. 5681. SPIE, 2005.

[68] Ricardo BC Prudencio and Teresa B Ludermir. Selective generation of training examples in active meta-learning. International Journal of Hybrid Intelligent Systems, 5(2):59–70, 2008.

[69] Hiranmayi Ranganathan, Hemanth Venkateswara, Shayok Chakraborty, and Sethuraman Panchanathan. Deep active learning for image classification. In 2017 IEEE International Conference on Image Processing (ICIP), pages 3934–3938. IEEE, 2017.

[70] Snell, Jake, Kevin Swersky, and Richard S. Zemel. "Prototypical networks for few-shot learning." arXiv preprint arXiv:1703.05175 (2017).

[71] H. Sahbi. "Imageclef annotation with explicit context-aware kernel maps." International Journal of Multimedia Information Retrieval 4.2 (2015): 113-128.

[72] Sener, Ozan, and Silvio Savarese. "Active learning for convolutional neural networks: A core-set approach." arXiv preprint arXiv:1708.00489 (2017).

[73] David D Lewis and William A Gale. A sequential algorithm for training text classifiers. In SIGIR'94, pages 3–12. Springer, 1994.

[74] H. Sahbi. "Learning laplacians in chebyshev graph convolutional networks." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[75] Xin Li and Yuhong Guo. Adaptive active learning for image classification. In CVPR, pages 859–866. IEEE Computer Society, 2013.

[76] Ajay J. Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In CVPR, pages 2372–2379. IEEE Computer Society, 2009.

[77] L. Wang and H. Sahbi. "Bags-of-daglets for action recognition." 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014.

[78] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009

[79] Frank Olken. Random sampling from databases. PhD thesis, University of California, Berkeley, 1993.

[80] H. Sahbi. "Lightweight Connectivity In Graph Convolutional Networks For Skeleton-Based Recognition." 2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021.

[81] Maria E Ramirez-Loaiza, Manali Sharma, Geet Kumar, and Mustafa Bilgic. Active learning: an empirical study of common baselines. Data mining and knowledge discovery, 31(2):287–313, 2017.

[82] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning, 2019.

[83] H. Sahbi. "CNRS-TELECOM ParisTech at ImageCLEF 2013 Scalable Concept Image Annotation Task: Winning Annotations with Context Dependent SVMs." CLEF (Working Notes). 2013.

[84] Stefan Depeweg, Jose-Miguel Hernandez-Lobato, Finale Doshi-Velez, and Steffen Udluft. Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. In International Conference on Machine Learning, pages 1184–1193. PMLR, 2018.

[85] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. Journal of machine learning research, 2(Nov):45–66, 2001.

[86] Sahbi, H., and N. Boujemaa. "Robust face recognition using dynamic space warping." International Workshop on Biometric Authentication. Springer, Berlin, Heidelberg, 2002.

[87] Ashish Kapoor, Kristen Grauman, Raquel Urtasun, and Trevor Darrell. Active learning with gaussian processes for object categorization. In 2007 IEEE 11th International Conference on Computer Vision, pages 1–8. IEEE, 2007.

[88] Sachin Ravi and Hugo Larochelle. Meta-learning for batch mode active learning. 2018. In URL https://openreview.net/forum, 2018.

[89] H. Sahbi, D. Geman. A Hierarchy of Support Vector Machines for Pattern Detection. Journal of Machine Learning Research 7 (10).

[90] Kunkun Pang, Mingzhi Dong, Yang Wu, and Timothy Hospedales. Meta-learning transferable active learning policies by deep reinforcement learn- ing. arXiv preprint arXiv:1806.04798, 2018.

[91] Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G. Hauptmann. Multi-class active learning by uncertainty sampling with diversity maximization. Int. J. Comput. Vis., 113(2):113–127, 2015.

[92] S. Thiemert, H. Sahbi, and M. Steinebach. "Using entropy for image and video authentication watermarks." Security, Steganography, and Watermarking of Multimedia Contents VIII. Vol. 6072. SPIE, 2006.

[93] Christoph Mayer and Radu Timofte. Adversarial sampling for active learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3071–3079, 2020.

[94] Dwarikanath Mahapatra, Behzad Bozorgtabar, Jean-Philippe Thiran, and Mauricio Reyes. Efficient active learning for image classification and seg- mentation using a sample selection and conditional generative adversarial network, 2019.

[95] H. Sahbi and N. Boujemaa. "Coarse-to-fine support vector classifiers for face detection." Object recognition supported by user interaction for service robots. Vol. 3. IEEE, 2002.

[96] Jia-Jie Zhu and Jose Bento. Generative adversarial active learning. CoRR, abs/1702.07956, 2017.

[97] Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence, 2020.

[98] Yong Cheng Wu. Active learning based on diversity maximization. In Ap- plied Mechanics and Materials, volume 347, pages 2548–2552. Trans Tech Publ, 2013.

[99] L. Wang, H Sahbi. Directed acyclic graph kernels for action recognition. Proceedings of the IEEE International Conference on Computer Vision, 3168-3175.

[100] Sharat Agarwal, Himanshu Arora, Saket Anand, and Chetan Arora. Con- textual diversity for active learning. In European Conference on Computer Vision, pages 137–153. Springer, 2020.

[101] Yarin Gal. Uncertainty in Deep Learning. PhD thesis, University of Cambridge, 2016.

[102] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2016.

[103] H. Sahbi. Coarse-to-fine support vector machines for hierarchical face detection. Diss. PhD thesis, Versailles University, 2003.

[104] Donggeun Yoo and In So Kweon. Learning loss for active learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 93–102, 2019.

[105] Patrick Hemmer, Niklas Kuhl, and Jakob Schoffer. Deal: Deep evidential active learning for image classification. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 865– 870, 2020.

[106] Jordan T Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gra- dient lower bounds. arXiv preprint arXiv:1906.03671, 2019.

[107] N. Boujemaa, F. Fleuret, V. Gouet, and H. Sahbi. "Visual content extraction for automatic semantic annotation of video news." In the proceedings of the SPIE Conference, San Jose, CA, vol. 6. 2004.

[108] Yoram Baram, Ran El Yaniv, and Kobi Luz. Online choice of active learning algorithms. Journal of Machine Learning Research, 5(Mar):255–291, 2004.

[109] Ksenia Konyushkova, Raphael Sznitman, and Pascal Fua. Learning active learning from data. arXiv preprint arXiv:1703.03365, 2017.

[110] H. Sahbi. "Misalignment resilient cca for interactive satellite image change detection." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.

[111] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. Active learning by querying informative and representative examples. In John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Cu- lotta, editors, NIPS, pages 892–900. Curran Associates, Inc., 2010.

[112] T. Napoléon and H. Sahbi. "From 2D silhouettes to 3D object retrieval: contributions and benchmarking." EURASIP Journal on Image and Video Processing 2010 (2010): 1-17.

[113] Naoki Abe. Query learning strategies using boosting and bagging. Proc. of ICML98, pages 1–9, 1998.

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

[115] X. Li and H. Sahbi. "Superpixel-based object class segmentation using conditional random fields." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.

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

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

[118] H. Sahbi. Kernel PCA for similarity invariant shape recognition. Neurocomputing 70 (16-18), 3034-3045.

[119] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[120] Carl Doersch, Abhinav Gupta, Alexei A. Efros. Unsupervised Visual Representation Learning by Context Prediction, arXiv:1505.05192, 2015.

[121] M. Jiu and H. Sahbi. "Semi supervised deep kernel design for image annotation." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015.

[122] Culotta, Aron, and Andrew McCallum. "Reducing labeling effort for structured prediction tasks." AAAI. Vol. 5. 2005.

[123] M. Jiu and H. Sahbi. "Deep kernel map networks for image annotation." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016.

[124] Atwood, J., Towsley, D.: Diffusion-convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1993–2001 (2016)

[125] Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)

[126] Chen, J., Zhu, J., Song, L.: Stochastic training of graph convolutional networks with variance reduction. arXiv preprint arXiv:1710.10568 (2017)

[127] Chen, J., Ma, T., Xiao, C.: Fastgcn: fast learning with graph convolutional networks via importance sampling. arXiv preprint arXiv:1801.10247 (2018)

[128] H. Sahbi and N. Boujemaa. "From coarse to fine skin and face detection." Proceedings of the eighth ACM international conference on Multimedia. 2000.

[129] Dai, H., Kozareva, Z., Dai, B., Smola, A., Song, L.: Learning steady-states of iterative algorithms over graphs. In: International Conference on Machine Learning. pp. 1114–1122 (2018)

[130] Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in neural information processing systems. pp. 3844–3852 (2016)

[131] H. Sahbi and F. Fleuret. Kernel methods and scale invariance using the triangular kernel. Diss. INRIA, 2004.

[132] Gao, H., Wang, Z., Ji, S.: Large-scale learnable graph convolutional networks. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1416–1424. ACM (2018)

[133] Gori, M., Monfardini, G., Scarselli, F.: A new model for learning in graph domains. In: Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. vol. 2, pp. 729–734. IEEE (2005)

[134] Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems. pp. 1024–1034 (2017)

[135] Henaff, M., Bruna, J., LeCun, Y.: Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163 (2015)

[136] H. Sahbi and F. Fleuret. Scale-invariance of support vector machines based on the triangular kernel. Diss. INRIA, 2002.

[137] Huang, W., Zhang, T., Rong, Y., Huang, J.: Adaptive sampling towards fast graph representation learning. In: Advances in Neural Information Processing Systems. pp. 4558–4567 (2018)

[138] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)

[139] H. Sahbi, J-Y. Audibert, R. Keriven. Context-dependent kernels for object classification. IEEE transactions on pattern analysis and machine intelligence 33 (4), 699-708.

[140] Levie, R., Monti, F., Bresson, X., Bronstein, M.M.: Cayleynets: Graph convolutional neural networks with complex rational spectral filters. IEEE Transactions on Signal Processing **67**(1), 97–109 (2018)

[141] Li, R., Wang, S., Zhu, F., Huang, J.: Adaptive graph convolutional neural networks. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)

[142] L. Wang and H. Sahbi. "Nonlinear cross-view sample enrichment for action recognition." European Conference on Computer Vision. Springer, Cham, 2014.

[143] H. Sahbi, P. Etyngier, J-Y. Audibert, R. Keriven. Manifold learning using robust graph laplacian for interactive image search. In CVPR 2008.

[144] Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R.: Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493 (2015)

[145] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. arXiv preprint arXiv:1901.00596 (2019)

[146] Zhang, J., Shi, X., Xie, J., Ma, H., King, I., Yeung, D.Y.: Gaan: Gated attention networks for learning on large and spatiotemporal graphs. arXiv preprint arXiv:1803.07294 (2018)

[147] M. Ferecatu and H. Sahbi. "Multi-view object matching and tracking using canonical correlation analysis." 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, 2009.

[148] T. Ma, J. Chen, and C. Xiao, "Constrained generation of semantically valid graphs via regularizing variational autoencoders," in Proc. of NeurIPS, 2018, pp. 7110–7121.

[149] H. Sahbi. "Learning Connectivity with Graph Convolutional Networks." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.

[150] S. Pan, R. Hu, G. Long, J. Jiang, L. Yao, and C. Zhang, "Adversarially regularized graph autoencoder for graph embedding." in Proc. of IJCAI 2018, pp. 2609–2615.

[151] H. Sahbi, L .Ballan, G. Serra, A. Del-Bimbo. Context-dependent logo matching and recognition. IEEE Transactions on Image Processing 22 (3), 1018-1031.

[152] J. Heinonen, *Lectures on Lipschitz Analysis*, Springer, 2005.