Reinforcement-based Display Selection for Frugal Learning

Sébastien Deschamps, Hichem Sahbi

To cite this version:

Sébastien Deschamps, Hichem Sahbi. Reinforcement-based Display Selection for Frugal Learning. International Conference on Pattern Recognition (ICPR), Aug 2022, Montréal, Canada. hal-03848481

HAL Id: hal-03848481

https://hal.science/hal-03848481v1

Submitted on 10 Nov 2022

Reinforcement-based Display Selection for Frugal Learning

Sebastien Deschamps^{1,2} Hichem Sahbi^{1}
^{1} Sorbonne University, UPMC, CNRS, LIP6, France ^{2} Theresis Thales, France

***Abstract*—Most of the existing learning models, particularly deep neural networks, are reliant on large datasets whose hand-labeling is expensive and time demanding. A current trend is to make the learning of these models frugal and less dependent on large collections of labeled data. Among the existing solutions, deep active learning is currently witnessing a major interest and its purpose is to train deep networks using as few labeled samples as possible. However, the success of active learning is highly dependent on how critical are these samples when training models. In this paper, we devise a novel active learning approach for label-efficient training. The proposed method is iterative and aims at minimizing a constrained objective function that mixes diversity, representativity and uncertainty criteria. The proposed approach is probabilistic and unifies all these criteria in a single objective function whose solution models the probability of relevance of samples (i.e., how critical) when learning a decision function. We also introduce a novel weighting mechanism based on reinforcement learning, which adaptively balances these criteria at each training iteration, using a particular stateless Q-learning model. Extensive experiments conducted on staple image classification data, including Object-DOTA, show the effectiveness of our proposed model w.r.t. several baselines including random, uncertainty and flat as well as other work.**

I. Introduction

Visual recognition aims at translating the content of a given image into semantic categories [1], [11]. This task is currently witnessing a tremendous interest in pattern recognition and image processing through the use of deep learning models, and particularly convolutional neural networks (CNNs) [3]–[5] and more recently transformers [2]. Nonetheless, the success of these models is highly dependent on the availability of large collections of hand-labeled training data. In practice, labeling manually large datasets is very time and effort demanding, and the current trend is to *frugally* train models using transfer learning [6], domain adaptation [7], data augmentation [8], zero/few shot learning [17], self-supervision [9] and synthetic data/ground truth generation [10]. However, the relative success of these solutions relies upon a strong assumption that knowledge are enough in order to close the *accuracy gap* while actually labeled data are more important.

Another category of methods is active learning [23] which reserves the labeling effort only to critical data, i.e., on well selected and most qualitative subsets whose impact on the accuracy of the learned models is the most significant. This process is iterative and asks an oracle (annotator) to label a few samples deemed informative from a large pool of unlabeled data, prior to update a decision function that eventually maximizes generalization. Most of the active learning solutions are basically heuristics [14]–[16], [22]–[24] which select unlabeled data by considering relevance measures that capture how critical are these data when learning decision functions. These measures are usually based on diversity, representativity and uncertainty [12]. Diversity allows exploring different modes of data distribution while representativity seeks to select prototypical samples in each mode in order to avoid outliers. Uncertainty is instead used to locally refine the learned functions around ambiguous samples. A suitable tradeoff between these criteria makes it possible to balance exploration and exploitation, two widely known concepts in active learning [13], and this tradeoff is dependent on the distribution of the data and the task at hand.

In this paper, we introduce a novel active learning solution based on the minimization of a constrained objective function that mixes diversity, representativity and uncertainty criteria. In contrast to most of the aforementioned existing solutions (and those described in section II), which are basically heuristics, the proposed contribution is probabilistic and unifies all these criteria in a single objective function whose solution models the probability of relevance of samples (i.e., how critical are samples) when learning a decision function. We also introduce a novel weighting process based on reinforcement learning (RL), which adaptively tradeoffs these criteria at each iteration of active learning, and thereby avoids the combinatorial aspect of setting these criteria under the regime of frugal labeling. The proposed RL approach relies on a particular stateless Q-learning model. Extensive experiments conducted on challenging image classification datasets, including Object-DOTA, show the effectiveness of our proposed model w.r.t. several related works including flat, random and uncertainty display model selection.

II. Related work

Early active learning solutions are based on Bayesian inference [14], meta learning [15], [30]–[32] and more recent ones are dedicated to deep learning [16], [27], [28]. While some of these methods have shown a relative gain w.r.t. random sampling [25] and other baselines [26], they are basically heuristics and lack groundedness. In general, state-of-the-art active learning algorithms include pool-based and generative models. Pool-based methods use different acquisition strategies to sample informative examples among a pool of unlabeled data. This category includes diversity [42], [43] and uncertainty-based techniques [9], [46]–[48], [62] as well as their combination [27], [33], [49]. A representative work in diversity [18] casts the problem of active learning as a core-set selection [34], and proceeds by optimizing an euclidean distance between selected and non-selected data. The goal is to choose a subset of unlabeled data such that a model trained on it would perform similarly to a model trained on the whole dataset. However, core-set methods reach their limitation when distances between data become confound in high-dimensional spaces. Other methods, based on uncertainty [19], [20], [29], [62], attempt to select samples deemed ambiguous, with the assumption that the more uncertain a model is about its prediction, the more informative the sample is for that model. For instance, authors in [19] select the most uncertain (and hence informative) samples, with the least confidence scores, using minimal margin and entropy on top of softmax class probabilities [21]. However, in spite of being widely used, softmax may not reflect the actual uncertainty in the learned models [45]. Besides, using only uncertainty, particularly in batch-based CNNs, may result into redundant sampling which may lead to worse performances compared to random samples. Hence, several works attempt to combine uncertainty with diversity in order to overcome this limitation [27], [33], [49].

Another category of methods relies on generative adversarial networks (GANs) in order to synthesize informative training samples [35], [36]. A variant known as cGAN [38] conditions GAN on real images whereas ASAL [36] uses generated images to select/add similar real-world images to the training set together with their labels. In the latter, authors combine uncertainty, adversarial sample generation and matching to synthesize uncertain data. This is achieved, without an exhaustive search over pools of unlabeled data, with supposedly more resilience to sampling bias compared to other generative adversarial active learning approaches. In GAAL [39], authors annotate synthetic samples and use them for linear SVMs and deep convolutional (DC) GANs training. Nevertheless, their method performs worse than random sampling; this is due to the sampling bias and also the difficulty in annotating the generated (poor quality) samples. Overall, the gain of these GAN-based approaches has, thus far, not been consistently established w.r.t. other strategies including random sampling [25], maximal entropy and minimal distance baselines [29], [30], [50], and other approaches [51]–[53] as well as self-taught learning [40].

III. Proposed model

Let $\mathcal{X}$ denote the set of all possible images drawn from an existing but unknown probability distribution $P(X, Y)$. In this definition, the random variable X refers to an input image and Y to its unknown class label. Considering nc visual classes (a.k.a. labels or categories), and $\mathcal{U}$ as a large subset of $\mathcal{X}$ whose labels are initially unknown, our goal is to design classifiers $\{g_c\}_{c=1}^{nc}$ by interactively labeling a very *small* fraction of $\mathcal{U}$, and training the parameters of $\{g_c\}_c$. This interactive labeling and training is known as active learning.

Let $\mathcal{D}_t$ be a *display* (defined as a subset of $\mathcal{U}$) 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 performances or exhausting a labeling budget:

- Get the labels of $\mathcal{D}_t$ as $\mathcal{Y}_t \leftarrow \text{oracle}(\mathcal{D}_t)$. This oracle function may depend on an *only-user-known* ground-truth.
- Train $\{g_{c,t}\}_c$ using $\bigcup_{\tau=1}^{t}(\mathcal{D}_\tau, \mathcal{Y}_\tau)$, where the second subscript in $g_{c,t}$ refers to the decision function at iteration t. In the remainder of this paper, different learning models will be considered including deep convolutional networks.
- Select the next display $\mathcal{D} \subset \mathcal{U} - \bigcup_{\tau=1}^{t} \mathcal{D}_\tau$ that possibly increases the generalization performances of the subsequent classifiers $\{g_{c,t+1}\}_c$. As the labels of the display $\mathcal{D}$ are unknown and also expensive, one cannot combinatorially sample all the possible subsets $\mathcal{D}$, train the associated classifiers, and select the best display. Alternative 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. Nonetheless, one should be cautious in the way these sampling strategies are applied as many of them may lead to equivalent or worse performances compared to simple random sampling (see for e.g. [54] and references within).

In what follows, we introduce our main contribution: a novel display model which allows selecting the most representative samples to label by an oracle. The proposed approach relies both on a constrained objective function and a weight selection strategy based on reinforcement learning. This whole model turns out to be highly effective compared to different related display selection strategies including random, uncertainty as well as other related work as corroborated later in experiments.

A. Display selection model

We consider a probabilistic framework which defines for each sample $\mathbf{x}_i \in \mathcal{U}$ a membership value μ_i that measures how likely is "$\mathbf{x}_i$ belongs to subsequent display $\mathcal{D}_{t+1}$"; consequently, $\mathcal{D}_{t+1}$ will correspond to the unlabeled data in $\{\mathbf{x}_i\}_i \subset \mathcal{U}$ with the highest memberships $\{\mu_i\}_i$. Considering $\mu \in \mathbb{R}^n$ (with $n = |\mathcal{U}|$) as a vector of these memberships $\{\mu_i\}_i$, we propose to find μ as the optimum of the following constrained minimization problem

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

here $\odot$, $'$ 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

- $\mathbf{D} \in \mathbb{R}^{n \times K}$ and $\mathbf{D}_{ik} = d_{ik}^2$ is the euclidean distance between $\mathbf{x}_i$ and k^{th} cluster centroid of a partition of $\mathcal{U}$ ($\{h_1, \dots, h_K\}$) obtained with K-means clustering.
- $\mathbf{C} \in \mathbb{R}^{n \times K}$ is a binary indicator matrix with each entry $\mathbf{C}_{ik} = 1$ iff $\mathbf{x}_i$ belongs to the k^{th} cluster, and 0 otherwise.

[1] The oracle is defined as an expert annotator providing labels for any given subset of images.

- And $\mathbf{F} \in \mathbb{R}^{n \times nc}$ is a scoring matrix with $\mathbf{F}_{ic} = \hat{g}_{c,t}(\mathbf{x}_i)$ and $\{\hat{g}_{c,t}\}_c$ being a stochastic variant of the initial decision functions $\{g_{c,t}\}$, i.e., $\hat{g}_{c,t}(.) \in [0,1]$ and $\sum_c \hat{g}_{c,t}(.) = 1$. In the particular context of deep convolutional networks, these normalized classifiers correspond to softmax layer.

The first term $\mathbf{tr}\big(\text{diag}(\mu'[\mathbf{C} \odot \mathbf{D}])\big)$ in Eq. 1 (equal to $\sum_i \sum_k 1_{\{\mathbf{x}_i \in h_k\}} \mu_i d_{ik}^2$) measures the *representativity* of the selected samples in $\mathcal{D}$; it captures how close is each data $\mathbf{x}_i$ w.r.t. the centroid of its cluster, and this term reaches its smallest value when the centroids are sufficiently numerous and when they coincide with the selected samples. The second term $[\mathbf{C}'\mu]' \log[\mathbf{C}'\mu]$ in Eq. 1 (equivalent to $\sum_k [\sum_{i=1}^n 1_{\{\mathbf{x}_i \in h_k\}} \mu_i] \log[\sum_{i=1}^n 1_{\{\mathbf{x}_i \in h_k\}} \mu_i]$) captures the *diversity* of the selected samples, defined as the entropy of the probability distribution of the underlying clusters; this term is minimized when the selected samples belong to different clusters and vice-versa. The third criterion $\mathbf{tr}\big(\text{diag}(\mu'[\mathbf{F} \odot \log \mathbf{F}])\big)$ (equal to $\sum_i \sum_c^{nc} \mu_i \mathbf{F}_{ic} \log \mathbf{F}_{ic}$) captures the *ambiguity* in $\mathcal{D}$ measured by the entropy of $\{\hat{g}_{c,t}(.)\}_c$; this third term reaches it 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; without any a priori about the three other criteria, the fourth term favors a flat μ-distribution and acts as a regularizer.

B. Optimization

Proposition 1: The optimality conditions of (1) lead to the solution

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

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

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

here $\mathbf{1}_{nc}$, $\mathbf{1}_K$ denote two vectors of nc and K ones respectively.

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 $\{\hat{g}_{c,t+1}\}_c$ (see also algorithm 1). The setting of γ in Eq. 3 controls the sharpness of the μ-distribution; larger values result into flat distribution while smaller values to Dirac-like distribution. A reasonable setting of γ consists in dividing the numerator inside the exponential by its norm.

As shown in the remainder of this paper, the setting of the other 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. Besides, the setting of these hyper-parameters should be iteration-dependent as early, intermediate and late iterations t may require different display selection strategies. Moreover, since labeling is sparingly achieved and on-the-fly, no extra labeled validation sets could be made available *beforehand* in order to "optimally" set these hyper-parameters; and even when labeled validation sets are available, tuning these hyper-parameters through all the iterations t is highly combinatorial and intractable[2].

Algorithm 1: Display selection mechanism

```
Input: Images in $\mathcal{U}$, display $\mathcal{D}_0 \subset \mathcal{U}$, budget $T$, $B$.
Output: $\cup_{t=0}^{T-1}(\mathcal{D}_t, \mathcal{Y}_t)$ and $\{g_t\}_t$.
for $t := 0$ to $T-1$ do
    $\mathcal{Y}_t \leftarrow$ oracle($\mathcal{D}_t$);
    $g_t \leftarrow \arg\min_g P(g(X) \neq Y)$ ;    // Learning model (built on top of $\cup_{k=0}^{t}(\mathcal{D}_k, \mathcal{Y}_k)$)
    $\hat{\mu}^{(0)} \leftarrow$ rand; $\mu^{(0)} \leftarrow \frac{\hat{\mu}^{(0)}}{\|\hat{\mu}^{(0)}\|_1}$; $\tau \leftarrow 0$
    while ($\|\mu^{(\tau+1)} - \mu^{(\tau)}\|_1 \geq \epsilon \wedge \tau < maxiter$) do
        Set $\mu^{(\tau+1)}$ using Eqs. (2) and (3) ;
         // Display model
        $\tau \leftarrow \tau + 1$
    end
    $\tilde{\mu} \leftarrow \mu^{(\tau)}$
    $\mathcal{D}_{t+1} \leftarrow \{\mathbf{x}_i \in \mathcal{U} \backslash \cup_{k=0}^{t} \mathcal{D}_k : \tilde{\mu}_i \in \mathcal{L}_B(\tilde{\mu})\}$ ;
     // $\mathcal{L}_B(\tilde{\mu})$ being the $B$ largest values of $\tilde{\mu}$
end
```

C. RL-based display selection

In what follows, we rewrite the classifiers $\{g_{c,t}\}_c$ trained at a given iteration t simply as g_t. 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 subsequent iteration $t+1$, and for any instance $\lambda_{t+1} \in \Lambda$ (written for short as λ), one may obtain a display (now rewritten as $\mathcal{D}_{t+1}^{\lambda}$) by solving Eq. 1, and the best configuration λ^* that yields an optimal display could be defined as

$$\lambda^* \leftarrow \arg\min_{\lambda \in \Lambda} \mathcal{R}_{\text{emp}}(g_{t+1}; \mathcal{V}_t), \tag{4}$$

here $\mathcal{V}_t \subset \cup_{\tau=1}^{t} \mathcal{D}_\tau$ is a holdout set taken from the previous oracle's annotations[3] and $\mathcal{R}_{\text{emp}}(g_{t+1}; \mathcal{V}_t)$ denotes the empirical risk of g_{t+1} on $\mathcal{V}_t$. As solving Eq. 4 requires generating and labeling multiple displays $\mathcal{D}_{t+1}^{\lambda}$ for different λ, and training the underlying classifiers, Eq. 4 makes finding the

[2] This tuning is intractable as the number of hyper-parameters scales linearly w.r.t. the max number of iterations T, and the number of possible grid search configurations scales polynomially as $\mathcal{O}(p^T)$ where p is the number of possible tested configurations for each hyper-parameter.

[3] Note that classifiers $\{g_t\}_t$ are trained on $\{\cup_{\tau=1}^{t} \mathcal{D}_\tau\}_t$ but these training sets are deprived from $\{\mathcal{V}_t\}_t$, and the latter are used only for validation.

best configuration λ^* clearly intractable. Moreover, in the frugal learning regime, one may not afford labeling multiple displays; besides, the holdout sets $\{\mathcal{V}_t\}_t$ are not sufficiently large in practice to make the setting of λ^* reliable which may lead to weak generalization. In order to bypass all these limitations, we consider in what follows an efficient and effective framework, based on RL, which allows training these hyper-parameters while considering not only the immediate reward (current classifier accuracy) but also future estimates of these rewards.

Hyper-parameter selection. We consider an RL algorithm based on Markov Decision Process (MDP) (see for instance [55]). The latter corresponds to a tuple $\langle S, A, R, q, \delta\rangle$ with S being a state set, A an action set, $R : S \times A \mapsto \mathbb{R}$ an immediate reward function, $q : S \times A \mapsto S$ a transition function and δ a discount factor. An RL agent interacts with an environment by running a sequence of actions from A with the goal of maximizing an expected discounted reward. The agent follows a stochastic policy, $\pi : S \mapsto A$, which computes the true state-action value as

$$Q(s,a) = E_\pi \left[\sum_{t=0}^{\infty} \delta^t r_t | S_0, = s, A_0 = a \right], \qquad (5)$$

where $r_t = R(s,a)$ is an immediate reward at iteration t 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., $\pi_*(s) \leftarrow \arg\max_a Q(s,a)$ with $Q(s, \pi_*(s))$ being the optimal action value. One of the most used methods to solve this type of RL problems is Q-learning [56], which directly estimates the optimal value function and obeys the fundamental identity, the Bellman equation

$$Q(s,a) \leftarrow (1-\gamma_{rl})Q(s,a) + \gamma_{rl}(r_t + \delta \max_{a'} Q(s',a')), \quad (6)$$

being γ_{rl} and δ the learning rate and the discount factor respectively set (in practice) to 0.1, 0.9 and $s' = q(s,a)$. We consider in our hyper-parameter optimization, a stateless version, so $Q(s,a)$, $R(s,a)$ are rewritten simply as $Q(a)$ and $R(a)$ respectively. One may turn the optimization of the hyper-parameters (α, β, η) either on continuous or discrete domain Λ. In the continuous case, Λ is equal to $\mathbb{R}^3 \backslash (0,0,0)$ and the underlying action set A corresponds to 27 possible joint incremental updates of α, β, η by three multiplicative factors taken (in practice) from $\{1, 0.95, (0.95)^{-1}\}^3$. In the discrete case, Λ equates $\{0,1\}^3 \backslash (0,0,0)$ so the underlying action set A corresponds instead to 7 possible binary configurations of α, β, η. At each iteration t, the reward $R(a)$ of a given action $a \in A$ will be evaluated once the action executed and the underlying subsequent display $\mathcal{D}_{t+1}$ and classifier g_{t+1} trained. Following Eq. 4, the reward $R(a)$ is measured using the accuracy $1 - \mathcal{R}_{\text{emp}}(g_{t+1}; \mathcal{V}_t)$ of the learned decision function g_{t+1} evaluated on the holdout set $\mathcal{V}_t \subset \cup_{\tau=1}^{t} \mathcal{D}_\tau$ whose cardinality does not exceed (in practice) 10% of the oracle's annotated displays. Note that this holdout set is used only for reward estimation (and hence hyper-parameter update) and not for classifier training. The detailed steps of our RL-based display selection are shown in algorithm 2.

Algorithm 2: RL-based Display selection mechanism

```
Input: Images in $\mathcal{U}$, display $\mathcal{D}_0 \subset \mathcal{U}$, budget $T$, $B$.
Output: $\cup_{t=0}^{T-1}(\mathcal{D}_t, \mathcal{Y}_t)$ and $\{g_t\}_t$.
$\forall$action, $Q$(action) $\leftarrow$ rand ;   // rand $\in [0,1]$
for $t := 0$ to $T-1$ do
    $\mathcal{V}_t \leftarrow$ RSubset($\mathcal{D}_t$) ;   // Random subset
    $\mathcal{Y}_t \leftarrow$ oracle($\mathcal{D}_t$) ;   // Oracle annotations
    $g_t \leftarrow \arg\min_g P(g(X) \neq Y)$ ;   // Learning model (built on top of $\cup_{k=0}^{t}\mathcal{D}_k \backslash \mathcal{V}_k$)
    if $t \geq 1$ then
        $r_t \leftarrow 1 - P(g_t(\mathcal{V}_t) \neq Y(\mathcal{V}_t))$ ;   // Reward
        Set $Q$(prev_action) using the stateless version of Eq. 6;
    end
    $v \leftarrow$ rand ;
    if $v \leq \exp(-t)$ then
        best_action $\leftarrow$ random_action in $A$ ;   // explore with prob $\exp(-t)$
    end
    else
        best_action $\leftarrow \arg\max_{\text{action}} Q$(action) ;   // otherwise take best Q-value
    end
    prev_action $\leftarrow$ best_action ;
    $(\alpha, \beta, \eta) \leftarrow$ update(best_action, $\alpha, \beta, \eta$) ;
    $\hat{\mu}^{(0)} \leftarrow$ rand ; $\mu^{(0)} \leftarrow \frac{\hat{\mu}^{(0)}}{\|\hat{\mu}^{(0)}\|_1}$ ; $\tau \leftarrow 0$
    while $(\|\mu^{(\tau+1)} - \mu^{(\tau)}\|_1 \geq \epsilon \wedge \tau < maxiter)$ do
        Set $\mu^{(\tau+1)}$ using Eqs. (2) and (3) ;   // Display model
        $\tau \leftarrow \tau + 1$
    end
    $\tilde{\mu} \leftarrow \mu^{(\tau)}$ ;
    $\mathcal{D}_{t+1} \leftarrow \{\mathbf{x}_i \in \mathcal{U} \backslash \cup_{k=0}^{t} \mathcal{D}_k : \tilde{\mu}_i \in \mathcal{L}_B(\tilde{\mu})\}$ ;   // $\mathcal{L}_B(\tilde{\mu})$ being the $B$ largest values of $\tilde{\mu}$
end
```

IV. Experiments

We study the impact of our proposed display selection model on two remote sensing tasks: satellite image change detection [37], [41], [44] using the Jefferson dataset [57], and remote sensing object classification using Object-DOTA [58]. In the first task, i.e., change detection, the goal is to find occurrences of targeted changes in satellite image pairs taken at different instants. The Jefferson dataset, used in change detection, consists of 2,200 non-overlapping patch pairs (of 30 × 30 RGB pixels each). These pairs correspond to registered (bi-temporal) GeoEye-1 satellite images of 2,400 x 1,652 pixels with a spatial resolution of 1.65m/pixel, taken from the area of Jefferson (Alabama) in 2010 and in 2011. These images show multiple changes due to tornadoes in Jefferson (building destruction, etc.) as well as no-changes (including irrelevant ones as clouds). The underlying ground-truth consists of 2,161 negative pairs (no/irrelevant changes) and only 39 positive pairs (relevant changes), so more than 98% of this area corresponds to no-changes and this makes the task of finding relevant changes even more challenging. In our experiments, half of the dataset is used to train the display and the learning models while the remaining half for evaluation.

The second dataset — Object-DOTA as a variant of DOTA [58] — is larger and used for image classification. Object-DOTA contains 127,759 remote sensing snapshots, belonging to 15 categories (including harbors, ships, etc.). These snapshots were taken from 2,806 large remote sensing images, both in gray-scale and RGB, of dimensions ranging from 800^2 to $20{,}000^2$ pixels. DOTA images were originally collected from Google-Earth as well as GF-2 and JL-1 satellites. The number of images per category ranges from 98 to 37,028, so categories are also highly imbalanced. Training and test sets include 98,906 and 28,853 data respectively. As classes in both Jefferson and Object-DOTA are highly imbalanced, we measure the classification performances using the equal error rate (EER); the latter is a balanced generalization error that evenly weights errors through different classes. Smaller EER (or equivalently larger *accuracy* defined as 1-EER) implies better performances.

A. Backbones and pretraining

Images in Jefferson and Object-DOTA are encoded using two pretrained backbones; the GCN (graph convolutional network) in [60] for the former, and the ViT [59] for the latter. The GCN consists of multiple blocks of aggregation and inner product layers followed by pooling and fully connected layers. Note that these networks are pretrained differently; indeed, the used GCN is pretrained on Jefferson, but using a self-supervised pretext loss similar to the one in [61] while the ViT is pretrained on a different set (namely ImageNet [1]) using a supervised loss. In both cases, no ground-truth is used on Jefferson and Object-DOTA for pretraining.

B. Model Analysis and Ablation

In order to study the impact of different terms of our objective function, we consider them individually, pairwise and all jointly taken. In this study, the last term of Eq. 1 is always kept as it acts as a regularizer and allows obtaining the closed form in Eq. 2. The impact of each of these terms and their combination is shown in Tables I and II. From these results, we observe the highest impact of representativity+diversity especially at the earliest iterations of frugal learning, whilst the impact of ambiguity term raises later in order to locally refine the decision functions (i.e., once the modes of data distribution become well explored). These performances are shown for different sampling percentages at each iteration t.

According to these results, none of the settings (rows) in Table I and Table II obtains the best performance through all the iterations. Considering these observed ablation performances, a better setting of the α, β and η should be iteration-dependent using RL (as described in section III-C), and as also corroborated through performances shown Tables I, II and also the dynamics of the learned hyper-parameters through iterations t (shown in Fig. 2). Indeed, it turns out that this adaptive setting outperforms the other combinations (including "all", also referred to as "flat"), especially at the late iterations (highest sampling percentages) for RL-discrete (RL-D), and the low/mid sampling percentages for RL-continuous (RL-C) on Jefferson, and the late iterations on Object-DOTA. Nevertheless, the average performance of RL-C is better than RL-D.

TABLE I

THIS TABLE SHOWS AN ABLATION STUDY OF OUR DISPLAY MODEL ON JEFFERSON. 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). RL-D AND RL-C REFER TO OUR PROPOSED REINFORCEMENT BASED DISCRETE AND CONTINUOUS MODELS RESPECTIVELY. **In this table, smaller EERs imply better performances.**

Iter Samp%	2 2.90	3 4.36	4 5.81	5 7.27	6 8.72	7 10.18	8 11.63	9 13.09	10 14.54
rep	26.21	12.72	10.48	9.88	9.70	8.52	8.85	8.61	8.82
div	31.24	23.45	30.41	44.81	24.12	13.22	17.02	6.88.	7.98
amb	46.68	38.73	29.91	14.74	20.11	8.33	7.41	7.37	5.53
rep + div	26.21	33.35	25.10	21.55	11.71	2.84	1.65	1.59	1.43
rep + amb	26.21	12.62	10.81	9.82	9.70	8.53	9.23	8.60	8.82
div + amb	41.69	28.82	23.08	23.41	23.42	19.82	13.10	8.16	6.97
all (flat)	26.21	33.35	25.52	23.70	14.59	2.74	1.54	1.67	1.48
RL-D	31.75	**10.36**	14.83	13.36	14.70	**1.06**	**1.06**	**1.10**	**1.01**
RL-C	**9.91**	21.29	**9.95**	**6.54**	**4.65**	2.63	2.44	2.95	1.80

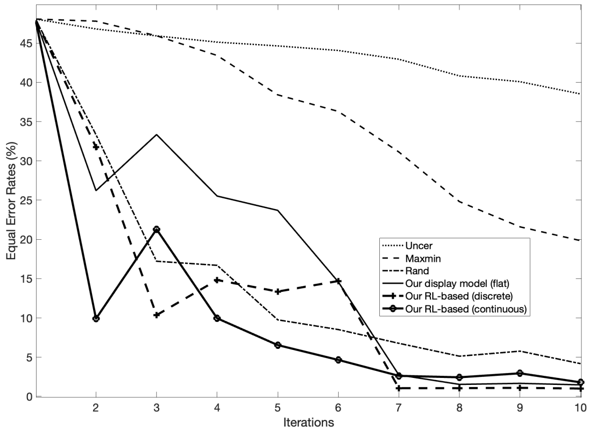

Fig. 1. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table I (Samp) on Jefferson. Here Uncer and Rand stand for uncertainty and random sampling respectively. Note that fully-supervised learning achieves an EER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

C. Extra Analysis and Comparison

Figure. 1 shows extra comparisons of our RL-based display model against different related display sampling strategies including *random, MaxMin and uncertainty*. Random selects data from $\mathcal{U}\backslash \cup_{k=0}^{t} \mathcal{D}_k$ whereas MaxMin (similar to [18])

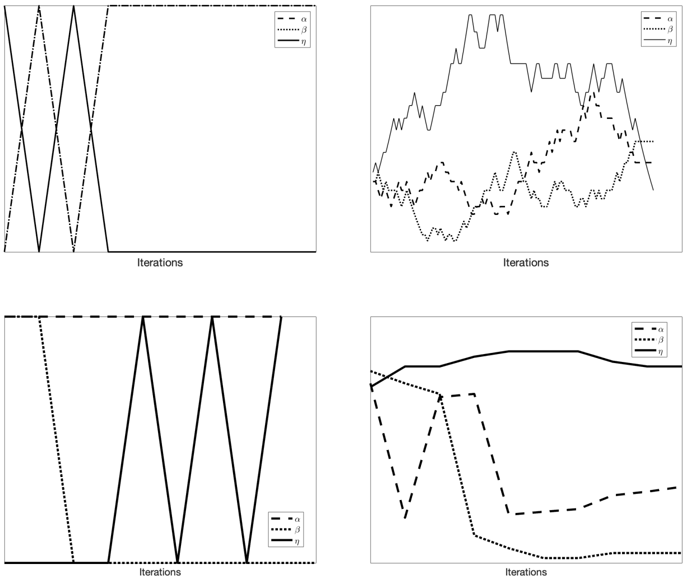

Fig. 2. This figure shows the dynamic of the learned hyper-parameters w.r.t different iterations of active learning on Jefferson (top) and Object-DOTA (bottom). These figures correspond to the discrete RL (left) and the continuous one (right). As observed, the variation in the continuous setting is gradual while in the discrete case the variation is more abrupt.

TABLE II

Same caption as Table. I, excepting experiments were achieved on Object-DOTA. Note that performances are reported via the accuracy (1-EER) with a higher frugal regime (i.e., lower percentages of labeled data compared to Jefferson data). **In this table, higher accuracies imply better performances.**

Iter Samp%	2 0.2	3 0.3	4 0.4	5 0.5	6 0.6	7 0.7	8 0.8	9 0.9	10 1.0
rep	25.77	27.91	29.81	30.69	31.96	33.51	34.08	34.53	35.05
div	26.16	34.04	37.97	41.33	41.85	44.44	45.57	48.12	48.72
amb	38.44	48.86	54.96	58.21	59.51	61.03	61.14	62.61	62.66
rep + div	49.57	51.38	53.60	54.44	54.74	55.49	55.47	55.87	56.25
rep + amb	42.04	49.43	53.49	57.23	59.73	61.88	62.78	63.42	64.16
div + amb	41.76	48.54	53.11	56.21	57.32	58.12	59.05	60.07	61.10
all (flat)	47.18	56.80	59.73	61.03	63.70	63.85	64.34	64.74	65.23
RL-D	35.42	41.19	43.44	46.28	51.29	53.43	54.09	53.47	56.59
RL-C	46.29	55.36	57.82	60.72	63.08	**64.43**	**64.88**	**66.20**	**66.61**

greedily selects a sample $\mathbf{x}_i$ in $\mathcal{D}_{t+1}$ from the pool $\mathcal{U} \backslash \cup_{k=0}^{t} \mathcal{D}_k$ by maximizing its minimum distance w.r.t $\cup_{k=0}^{t} \mathcal{D}_k$. We also compare our method w.r.t. uncertainty [62] which consists in selecting samples in the display whose scores are the most ambiguous (i.e., the closest to zero). 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 performances in figure 1 (and also figure 3) show the positive impact of the proposed RL-based display, both on the discrete and the continuous models, against the related sampling strategies for different amounts of annotated data. Excepting the flat model (also used in [57]), most of these comparative methods are powerless to classify data sufficiently well. Indeed, 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 decision functions but suffers from the lack of diversity). The flat display strategy [57] gathers the advantages of random, MaxMin and uncertainty, but suffers from the rigidity of the weights of representativity, diversity and ambiguity criteria which are fixed instead of being learned (i.e., iteration-dependent). In contrast, our proposed RL-based design adapts the choice of these criteria as iterations evolve; it's worth noticing that RL-C is effective including at the early iterations, and this makes it more suitable for high frugal regimes. In sum, the proposed RL-based display makes classification reaching lower EERs (and equivalently high accuracy) and overtakes all the other strategies at the end of the iterative learning process.

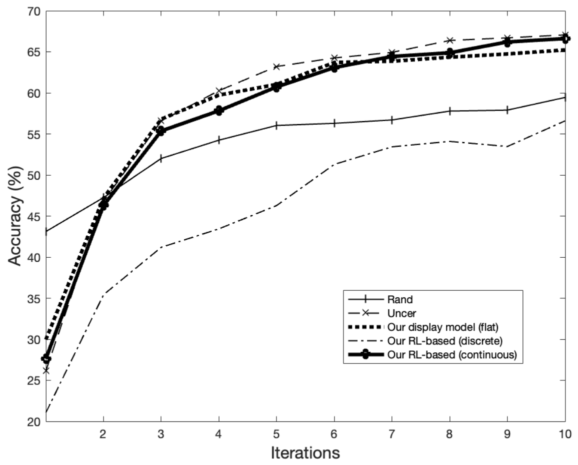

Fig. 3. This figure shows a comparison of different sampling strategies w.r.t. different iterations (Iter) and the underlying sampling rates in table II (Samp) on Object-DOTA. Note that fully-supervised learning achieves an accuracy of 74.92%. See again section IV-C for more details.

V. Conclusion

We introduce in this paper a novel display learning model based on the optimization of an objective function mixing representativity, diversity and ambiguity. The proposed approach is probabilistic and assigns membership measures to unlabeled samples, and selects the display as samples with the highest memberships. The proposed approach also relies on an RL-based mechanism which selects the best (discrete or continuous) combination of representativity, diversity and ambiguity through active learning iterations, thereby leading to better performances. Extensive experiments conducted on the task of remote sensing image classification and change detection show the outperformance of the proposed method against different settings as well as related work.

As a future work, we are currently investigating the use of self-supervised learning methods in order to further enhance the generalization capacity of our learning models as well as the use of generative networks for other display model design.

References

[1] 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

[2] 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.

[3] 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.

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

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

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

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

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

[9] 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.

[10] 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.

[11] Hichem 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

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

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

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

[22] Sanjoy Dasgupta. Analysis of a greedy active learning strategy. In NIPS, pages 337–344, 2004.

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

[24] Hichem Sahbi. "Interactive satellite image change detection with context-aware canonical correlation analysis." IEEE Geoscience and Remote Sensing Letters 14.5 (2017): 607-611.

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

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

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

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

[33] 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.

[34] Trevor Campbell and Tamara Broderick. Automated scalable bayesian inference via hilbert coresets. J. Mach. Learn. Res., 20:15:1–15:38, 2019.

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

[36] 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.

[37] Nicolas Bourdis, Denis Marraud, and Hichem Sahbi. "Camera pose estimation using visual servoing for aerial video change detection." 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2012.

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

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

[40] 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.

[41] Nicolas Bourdis, Denis Marraud, and Hichem Sahbi. "Spatio-temporal interaction for aerial video change detection." 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2012.

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

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

[44] Nicolas Bourdis, Denis Marraud, and Hichem Sahbi. "Constrained optical flow for aerial image change detection." 2011 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2011.

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

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

[47] 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.

[48] 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.

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

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

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

[52] 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.

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

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

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

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

[57] Hichem Sahbi, Sebastien Deschamps, and Andrei Stoian. ”Frugal Learning for Interactive Satellite Image Change Detection.” 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, 2021.

[58] Xia, Gui-Song, et al. ”DOTA: A large-scale dataset for object detection in aerial images.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

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

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

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

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