Towards Generic Image Classification: an Extensive Empirical Study

Raphaël Marée, Pierre Geurts, and Louis Wehenkel

Abstract—This paper considers the general problem of image classification without using any prior knowledge about image classes. We study variants of a method based on supervised learning whose common steps are the extraction of random subwindows described by raw pixel intensity values and the use of ensemble of extremely randomized trees to directly classify images or to learn image features. The influence of method parameters and variants is thoroughly evaluated so as to provide baselines and guidelines for future studies. Detailed results are provided on 80 publicly available datasets that depict very diverse types of images (more than 3800 image classes and over 1.5 million images).

Index Terms—Image classification, random subwindows, extremely randomized trees, feature learning, benchmarks, baselines.

1 Introduction

1.1 Context and motivation

The aim of supervised image classification is to automatically build computerized models able to predict accurately the class (among predefined ones) of new images, once trained from a set of labelled images. In the real world, this generic problem encompasses well-known tasks such as the automatic recognition of images of handwritten characters, faces, and road signs, to name but a few. In our preferred field of application, life science research, efficient classifiers have a huge potential ([1], [2], [3], [4], [5]) as they could ease the labor-intensive phenotype classification of biological samples (e.g. cells, tissues, organs, microbes, diatoms, plants, animal models, ...) thus facilitate obtaining more quantitative information in large-scale, high-throughput, screens.

However, since the early days of computer vision practice, when a researcher approaches a new image classification task, he or she often develops a dedicated algorithm to implement human prior knowledge as a sequence of specific operations, also known as a hand-crafted approach. Such an approach often involves the design and calculation of tailored filters and features capturing expected invariant image characteristics (e.g. [6] performed various operations to compute cell statistics such as nuclear/cytoplasmic area and nuclear density and texture to discriminate between normal and abnormal cells). Although several specific works have proved effective, the design choices are rarely straightforward hence such a strategy requires a lot of research and development efforts for each specific problem, and it might require major adjustments when parameters of the problem vary (e.g. sample preparation protocols, imaging modality, phenotypes to recognize, ...). In life science imaging, this engineering approach does not scale well as there are hundreds of thousands of biological entities that can be screened using many different sample preparation techniques and imaging modalities.

Over the last decade, the same kind of limitations in web-scale natural scene recognition and object categorization (where thousands of visual classes can be defined) motivated computer vision researchers to seek the development of more generic methods for the recognition of various classes of images that share some visual regularities, without relying on too strong assumptions about patterns to recognize and acquisition conditions. Towards that goal, a large variety of methods automatically extracting features and learning appearance statistics have been designed (see e.g. reviews of [7], [8], [9]), and multiple datasets have been published to train such models and assess their performances. Although high recognition performances are now achieved by diverse methods on a few datasets with more than 100 natural, coarse-grained, categories, machine performances are still inferior to human performances on challenging tasks [10], and the general image classification problem is rarely considered as a whole and thus remains largely unsolved. Indeed, the "image world" depicted by the few commonly used benchmarks, including the more recent and largest ones (e.g. [11]), remains limited as the space of possible images is much larger: potential applications hence the number of visual classes, real-world variations, and imaging acquisition procedures, are tremendous. Additionally, some dataset

• *All authors are with the Department of Electrical and Computer Engineering (Montefiore Institute), and with the GIGA-Research Center, University of Liège, Belgium.*
E-mail: raphael.maree@ulg.ac.be

issues have been raised [12], [13], [14], [15], [16], [17], [18] that somehow question the success of the best computer vision methods beyond these widely used benchmarks. Moreover, these methods are becoming more complex to implement and to parametrize as a decade of active research has produced numerous feature detectors and descriptors, filtering operations, combination schemes, classifier variants and architectures, and empirical studies have shown that their parameters influence significantly recognition performances (e.g. [19], [20], [21], [22]). Overall, it raises the problem of over-fitting to a limited number of specific datasets and the true, unbiased, potentials of such methods are only partially assessed. More dramatically, for real-world practice, although comprehensive studies were lately conducted on several datasets (e.g. [20]) from which some trends were observed (like the overall good performances of the bag-of-features framework for textures, natural scenes, and object recognition tasks), and although sophisticated toolboxes have emerged (e.g. [23], [24], [25], [22]), widely applicable, off-the-shelf, softwares and guidelines are hardly available.

In other words, difficulties of earlier practice requiring explicit programming of problem-dependent features have somehow been shifted to the issues of finding, for each new problem, the appropriate combination of algorithm steps and parameter values among many possibilities. Moreover due to their runtime complexity, some of these processing steps are not appropriate for real-time or large-scale applications where potentially hundreds of millions of objects have to be classified rapidly. Hence, scientific studies are often limited in scale, or still partially performed by hand (e.g. 50 millions of galaxies were manually labeled into morphological classes by almost 150000 humans within one year through the GalaxyZoo web-based project [26]), while others required very large computing infrastructures because they relied on dense feature computations (e.g. computers of the members of the Help Conquer Cancer project have contributed over 100 CPU-millenia for the automated classification of tens of millions of protein crystallization-trial images at a rate of 55 CPU-years per day [27]).

1.2 This work

Following and extending previous works [28], [29], [30], [31], we consider the generic problem of supervised image classification without any preconception about image classes, ie. it encompasses the recognition of numerous types of images under various image acquisition conditions. Indeed, with the design of a general-purpose yet simple and easily applicable image classifier in mind, [28], [29], [30], [31] proposed earlier an appearance-based, learning method, relying on dense random subwindow extraction in images, their description by raw pixel values, and the use of ensembles of extremely randomized trees to classify these subwindows hence images. Despite its conceptual simplicity and its rather low run-time complexity, it yielded interesting results on a few datasets. Subsequently, variants of the method were proposed in [32], [33], [34], [35] for object categorization, image segmentation, interest point detection, and content-based image retrieval.

The main objective of this paper is two-fold. First, to assess if results obtained by such an approach could be confirmed and extended on different application domains with various imaging conditions, we perform in this paper an extensive, systematic study of its performances on 80 publicly-available datasets (among which 25 bioimaging datasets). By conducting such a large-scale study, we seek to characterize the performances of the method and its recent variants, to study rigourously the influence of its parameters and classification schemes, to bring out the most influential design choices, and to draw general guidelines for future use so as to speed its application on new problems. Second, by summarizing publicly available databases and by providing our positive and negative results, we aim to foster research in generic methods, to encourage other researchers to evaluate and compare their methods on a wide range of imagery, and to draw attention to current dataset limitations. Additionally, to ease future research and comparison, a command-line software in Java will be freely provided on request, and supplementary materials with detailed results and links to download datasets are also available on a companion website: http://www.montefiore.ulg.ac.be/~maree/generic/.

2 Experimental setup

We will work with a large variety of datasets from many application domains. Our hypothesis is that by considering the image classification problem as a whole, it will possible to derive trends that are generally valuable, ie. applicable in several areas. For example, observations derived from experiments related to the recognition of traffic signs (captured with onboard cameras) or galaxies (captured during wide-field sky surveys) might be helpful for the recognition of cells (captured by microscopes) as these datasets are sharing some essential characteristics (they consist in different classes of shapes and they exhibit illumination and noise variations due to the acquisition process). Similarly, observations derived from material classification datasets might be of interest for biological tissue recognition (as their images have textured patterns).

2.1 Datasets

Our experimental setup comprises 80 image datasets that were previously published and are publicly and

freely available. They sum up roughly to 1.5 million images depicting approximately 3850 distinct classes. The choice of datasets was made a priori and independently of the results obtained with our method. The summary of their characteristics is given in Supplementary Table I, and an overview of image classes for all datasets is given in Supplementary Figures 1, 2, 3, and 4. Images were acquired worldwide, in controlled or uncontrolled conditions, using professional equipments in laboratory settings, individuals' digital camera in the real-world, various biomedical imaging equipements (fluorescence or brightfield microscopes, plain film radiography, etc.), robotic telescopes, synthetic aperture radars, etc. For a given dataset, image classes possibly exhibit subtle or prominent changes in their appearance due to various sources and levels of variations including possible changes in position, illumination, scale, and viewpoint, and/or presence of background clutter, occlusions, and noise. Moreover, either significant intra-class variations or high similarity between distinct classes could be present. Several of these datasets are synthetic and therefore variations are controlled (e.g. backgrounds are uniform) and well characterized, while many others contains real-world images so variations are mixed. Note that we only included in our experiments two widely used face datasets among tens of existing ones, given that face databases were recently summarized and evaluated thoroughly [36], [14], [37]. Also, we did not include the Pascal VOC challenge datasets [38] whose evaluation criteria (precision/recall curves for each object class) does not fit well into our evaluation framework (see below).

2.2 Protocols and Evaluation Criteria

Our evaluation protocols are summarized in Supplementary Table I. Our evaluation metric is the misclassification error rate evaluated on independant test images. If a precise dataset protocol was defined in the literature and was adopted in several papers, we also used it. However, for many datasets (e.g. those where the protocol was not rigourously described, or different between papers, or where the number of test images was rather small), we performed 10 runs where each run uses a certain number of images randomly drawn for the learning phase (e.g. 80% of the total number of images) and the remaining images for the testing phase (e.g. 20%). The misclassification error rate is then averaged over all the test sets which allows to have a reliable insight into the effects of method parameters. For some datasets with highly unbalanced classes the number of training/test images per run was explicitely fixed and balanced, or we use all available images and report mean class error rates (number of errors per class normalized by the number of test images in that class), similarly to other published works.

3 Method and Related Work

In this section, we present the two key components of the method and their relation with previous works. The method involves the extraction of random subwindows described by raw pixel values, presented in Section 3.1, and the use of ensemble of extremely randomized trees by different means, presented in Section 3.2. We will describe both steps and their parameters, and describe the way we will evaluate the influence of their parameters. Computational requirements are discussed in Supplementary Section 3.

3.1 Random Subwindows

The introduction of random subwindow sampling in [28] was motivated by the development of a generic method that does not make any assumption on the type of images to classify. That method tries to capture a rich representation of images by rapidly covering large parts of images and by describing subwindows by raw pixel values (in colors if available). This scheme contrasts with interest point detectors often used in object categorization and natural scene recognition ([8], [39]) that assume that the neighboorhood of corners, edges or contours capture interesting aspects of images that will help to classify them, an approach that might not be well-suited to capture the content of homogeneous regions without distinctive boundaries, or the content of low resolution images with weak signals. Moreover, the description of these random subwindows by raw pixel values is fast in comparaison with approaches that compute large sets of features (e.g. [40], [25]). It also differs from other methods based on low-dimensional invariant descriptors or filter banks outputs that might discard potential information contained in the original image patches (e.g. the color cue was often discarded by local invariant descriptors until recently [21], [22]). In contrast, the raw pixel descriptor is expected to capture fine-grained patterns and to allow discrimination between closely similar patches of a large number of classes, provided that one uses a robust classifier able to handle such high-dimensional input spaces. Different variants of random subwindow sampling were proposed: [28], [29] extracted square, fixed-size, subwindows and observed experimentally that very small subwindow sizes were optimal for some datasets (in particular for the texture dataset), while larger subwindows (close to the global image size) achieve better performance for others (in particular for the handwritten digit dataset). Influenced by these observations, [30] extracted square subwindows of random sizes chosen between 1×1 pixels and the minimum of the horizontal or vertical dimensions of each image. Subwindows were then resized to a fixed patch size whose pixels are used as input of the machine learning algorithm. The resizing step improved robustness to scale changes and it allows one

to use generic machine learning methods that work with fixed-size feature vectors. This procedure also introduces in the training set subwindows with slight pixel intensity variabilities through multiple over- or sub-sampling, a process that can help the algorithm to learn to be more robust to such changes that could occur naturally in unseen, test, images. To improve robustness to rotation, it was also proposed to apply a rotation (with a random angle) to subwindows so that the model can learn invariance from rotated subwindows. Later, [31] observed on one biological dataset that accuracy can be improved by constraining the size intervals of random subwindows compared to the full range of sizes. In these studies, not surprinsingly, the number of extracted subwindows has clearly an influence on accuracy (the more, the better).

3.1.1 This work: Assessment of random subwindow extraction parameters

While randomized patch extraction was then combined with various other steps (eg. [41], [32], [42], [43]), a large and systematic evaluation of randomization schemes is still lacking for image classification.

In this work, we first study systematically the influence of subwindow size intervals and the way random subwindows are encoded on all 80 datasets. Default tests are made using a total of $N_{ls} = 1$ million training subwindows (previous works [28], [30], [31] used only one hundred thousand subwindows) while a few others more intensive tests are performed with up to 50 millions subwindows. For a given dataset, the same number of subwindows are randomly drawn from each image, it equals N_{ls}/N_{img} where N_{img} is the number of training images. One can see subwindows as pixel context, support regions, or receptive fields of different sizes/scales whose intervals are systematically tested: we consider single pixels 1×1 as baseline, and 13 different configurations of square subwindows ranging from small image regions $[0\% - 10\%]$ to large ones $[90\% - 100\%]$, and including the default unconstrained size $[0\% - 100\%]$ used in [30]. Constraining sizes to e.g. $[25\% - 50\%]$, means that the size of each subwindow is randomly choosen between 25% and 50% of $min(width, height)$ in each image, then the position is randomly choosen in order to guarantee square subwindows are always fully contained within images. Note that in configurations with zero minimum ($[0\% - x\%]$), the minimum size is actually 1×1. The random subwindow procedure is illustrated in Figure 1. For all configurations (except baseline 1×1 where no resizing is performed), each subwindow is subsequently resized by bilinear interpolation to a patch of fixed size (8×8, 16×16 (default) or 32×32) and its pixel values encoded in HSV or graylevels are used as the subwindow descriptors. Whereas more elaborated or specific sampling schemes could be designed and might improve results on specific datasets (e.g.: localized sampling for datasets where positions of patterns of interest are known, rectangular subwindows for elongated objects, adaptive sampling [32], ...), we want here to investigate how far a basic, systematic, and generic random sampling could lead us in terms of accuracy on many datasets so as to provide baselines before developing more complex sampling schemes.

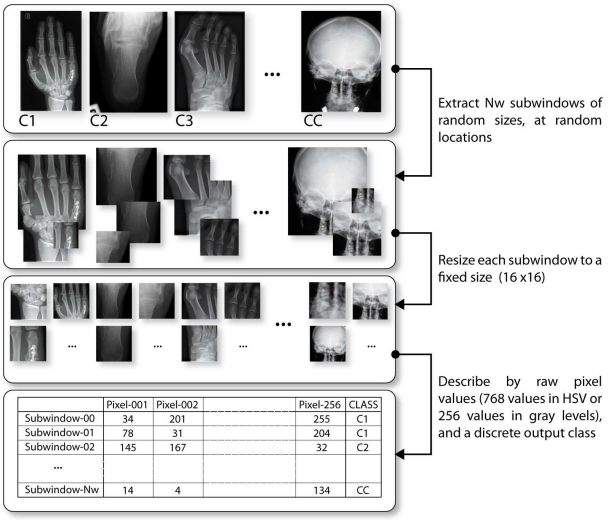

Fig. 1. Random subwindows extraction and description by pixel intensity values of resized patches, illustrated on a X-ray dataset with C classes.

3.2 Ensembles of Extremely Randomized Trees

Ensembles of randomized trees are increasingly used in machine learning and computer vision (see [44] for their recent developments in computer vision and medical imaging applications), and they have been shown to perform remarkably well on a broad range of domains, even competing with Boosted trees and SVMs in high-dimensional classification tasks [45].

The Extra-Trees algorithm was proposed in [46] where the reader can find a precise algorithm description. For image classification with random subwindows, the algorithm is initialized with a training sample of N_{ls} (randomly extracted) subwindows, each of which is described by a fixed number of input variables (raw pixel values of the resized subwindows) and a discrete output class (the class of the image from which the subwindow was extracted). Starting with the learning set of N_{ls} subwindows at the top-node, the Extra-Trees algorithm builds each decision tree of an ensemble according to the classical top-down decision tree induction procedure [47] that generates tests on input variables to progressively partitions the input space into hyperrectangular regions so as to yield regions where the output is constant. The two main differences between the Extra-Trees algorithm and other tree-based ensemble methods are that it splits nodes by using the best test (among k tests)

where both attributes and cut-points are choosen at random (rather than choosing the best cut-point for each attribute that optimizes a score measure like in Tree Bagging [48] or Random Forests [49]), and that it uses the whole learning sample (rather than a bootstrap replica in Tree Bagging and Random Forests) to grow the trees. In our case where subwindows are described by raw pixel values, a test associated to an internal node of a tree simply compares the value of a pixel (intensity of a grey level or of a certain color component) at a randomly choosen location within a subwindow to a cut-point value (note that tests involving several pixels will be evaluated in Section 4.2). The development of a node is stopped as soon as either all input variables or the output variable are constant in the local subset of the leaf (in which cases impurity can not be further reduced), or the number of subwindows in the leaf is smaller than a predefined threshold $n_{\min}$. Once an ensemble of T trees are built, they can be used in different ways to perform image classification, as described below.

3.2.1 Extremely Randomized Trees for Direct Image Classification (ET-DIC)

Previous works [28], [29], [30], [31] use trees as subwindow classifiers where the class of a subwindow is the class of its parent image, then to directly use them to predict the class of full images by aggregating predictions for individual subwindows given by all trees. This simple way of merging piecewise information yielded convincing results on various problems in previous, smaller, studies. In this scheme, when the developement of a node is stopped, it becomes a leaf where one computes and stores class frequencies according to the set of training subwindows that were propagated in this node. Each terminal node or leaf thus contains a vector of real values which dimensionality equals the number of classes.

Once the trees are built, the database of subwindows extracted from the training images are no longer used after training, and can be discarded. One only uses the ensemble model to classify subwindows of a test image as follows. The method similarly extracts a certain number, N_{ts}, of subwindows randomly within the test image, and then each test subwindow is propagated into each decision tree of the ensemble. Each decision tree outputs conditional class probability estimates for each subwindow. Each subwindow thus receives T class probability estimate vectors where T denotes the number of trees in the ensemble. All the subwindow predictions are then averaged and the class corresponding to the largest aggregated probability estimate is assigned to the image. This method variant is illustrated in Supplementary Figure 6.

3.2.2 Extremely Randomized Trees for Feature Learning (ET-FL)

Extremely randomized trees can also be used to generate visual features. In [33], an extension of ET-DIC was proposed for content-based image retrieval where a similarity measure reminiscent of tf-idf was based on terminal nodes of totally randomized tree ($k = 1$) . In [32], another extension was proposed for supervised object recognition, inspired by older works on textons and bags of visual words [50], [51], [52]. It used extremely randomized trees to build a global "bag-of-visual words" image representation (a "visual dictionary"), instead of the K-Means clustering algorithm traditionaly used in this setting. A linear SVM was then trained over this global image representation to perform the final image classification. That variant improved results over ET-DIC for 7 object classes [32].

In this classification scheme, instead of retaining probability estimates at terminal nodes and use trees to perform subwindow classification hence image classification like in ET-DIC, each terminal node (leaf) of a tree is thus considered as a "codebook" or "visual word". After propagating subwindows downto trees, each image is described by a single global feature vector which dimensionality equals the number of terminal nodes in the ensemble of trees, and where features can be encoded as binary or quantitative frequency values. Such a "bag-of-features" representation can then be fed into any classifier to build the final image classification model. To predict the class of a new image, its random subwindows are propagated into the ensemble of trees to build its global feature vector subsequently classified by a final classifier.

3.2.3 This work: Assessment of Extra-Trees modes and parameters

In this work, we will first stick strictly to the original Extra-Trees algorithm [46]. We will study systematically the influence of its parameters and the way trees are used to achieve image classification: Using them to directly classify subwindows hence images (ET-DIC), or using them to build features whose statistics are encoded into a global image representation (ET-FL) subsequently classified by a linear classifier.

For both variants, we study systematically on all 80 datasets the influence of the minimum node sample size n_{min} by picking a few of its possible values (from 1 to 1000 in ET-DIC and from 1 to 50000 in ET-FL), the number of random tests k (from 1 to the maximum number of input variables), and the number of trees T (from 1 to 20 in ET-DIC and from 1 to 40 in ET-FL although more extensive tests use up to $T = 1000$ trees). In ET-FL, we also study systematically the influence of the encoding of the global feature vector: We evaluate terminal binary encoding (where a feature equals to 1 if at least one of its subwindow was propagated to that terminal node, and 0 otherwise), but we also

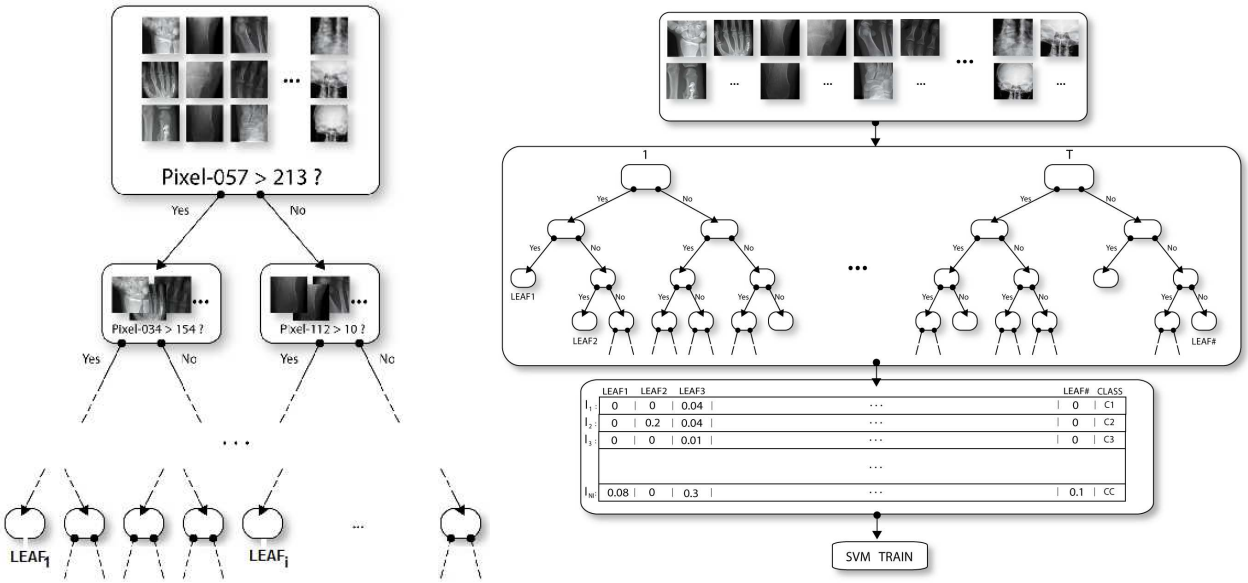

Fig. 2. Left: A single tree induced from a training set of random subwindows, using node tests with single pixel thresholding, for the ET-FL scheme. Right: An ensemble of T trees, the derived, quantitative frequency global representation for training images, and training of a final linear SVM classifier in ET-FL mode.

propose a quantitative frequency representation that rather computes the number of image subwindows that reach a given terminal node divided by the total number of subwindows extracted in the image (a bin value is included in $[0, 1]$, and the sum over all terminal nodes equals to 1 in a given tree for a given image), as illustrated by Figure 2. Another variant we evaluated is a hierarchical encoding where these binary values or frequencies are retained in all the tree nodes (internal and terminal nodes). Like [32], we use in ET-FL a linear SVM classifier to perform the final classification whose parameters were set to default values (see Supplementary Material for implementation details).

4 RESULTS

We first present a summary of our main findings in Section 4.1, then we will propose simple algorithmic extensions to improve results in Section 4.2. The present study was made possible due to the availability of a computer cluster (on a single processor this study would have required a few decades of computing times). The amount of intermediate data generated was also substantial (on the order of several tens of terabytes).

4.1 Summary of main results

Our main results are summarized in Figures 3 to 6 and summarized below (Detailed results are available in Supplementary Tables II to XII).

4.1.1 Random subwindows enable classification of many image types

Regarding the random subwindow extraction scheme, the most influential parameter is the size intervals of subwindows that allows the method to be adapted to very different types of problems. The optimal sizes are problem-dependant and could be very small or very large, and the default scheme using subwindows of any randomly choosen sizes (as in [30]) is a good compromise on average but significantly below results we can obtain by constraining the sizes on each dataset. Small subwindows allow to capture fine details and generally perform best for images with highly repeatable patterns ie. textured images (e.g. histological tissues, man-made materials, or assays with populations of cells), while larger subwindows yield better results for shape-like datasets (e.g. handwritten characters, red-blood cells, leaves, and traffic signs), as illustrated by Figure 4. For these latter type of datasets, extracting large subwindows augments the training set with (small) scale and translation variations, and allows models to directly capture global patterns. We observe the optimal range of sizes is often different between ET-DIC and ET-FL, with slightly smaller subwindows for the latter in general. 16×16 in color (when available) appears to be a good size for resized patch raw pixel descriptors. Concerning the number of extracted subwindows, a total of 1 million training subwindows performs well, but using a denser sampling can still improve results on several datasets (see Section 4.2).

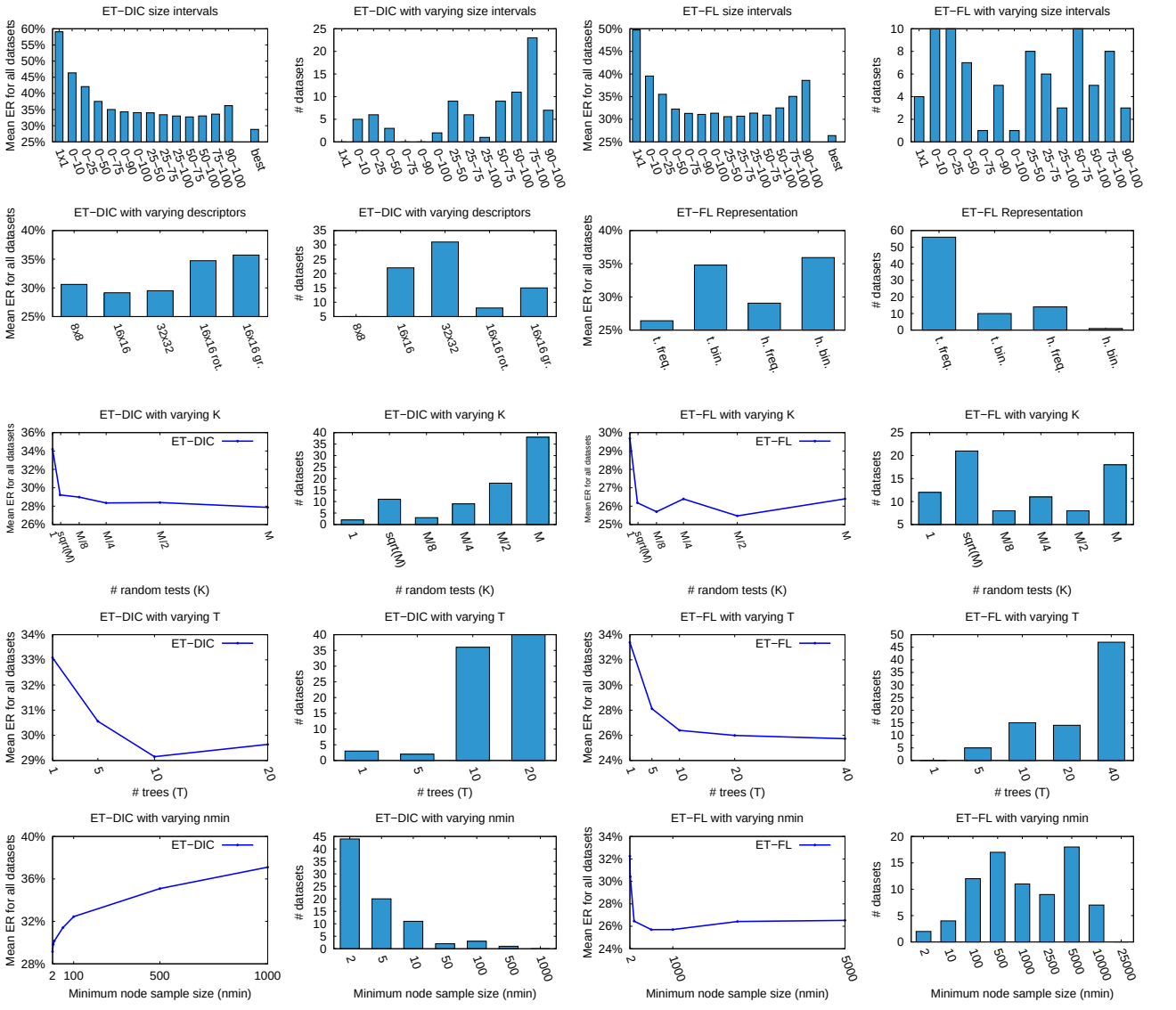

Fig. 3. Results averaged over all 80 datasets. ET-DIC (two first columns): Left: average of error rates for all datasets with subwindow size intervals (1st row), pixel descriptors (2nd), number of random tests (3rd), number of trees (4th), minimum node sample sizes (5th). Second column: Number of datasets for which the parameter values yield the best error rates. ET-FL (two last columns): Third column: Average of error rates for all datasets with subwindow size intervals (1st row), image representation (2nd), number of random tests (3rd), number of trees (4th), minimum node sample sizes (5th). Right: Number of datasets for which the parameter values yield the best error rates. See Supplementary Tables II to XII for detailed results.

4.1.2 *Observations for ET-DIC classification scheme*

With ET-DIC, the method mainly follows trends observed in previous, smaller, studies. Observing the average plots, an ensemble of trees yields better results than a single tree and 10 trees achieve good results and most often it seems not necessary to increase this number, given the additional memory requirements. However, on datasets where optimal subwindow sizes are large (close to full image sizes), increasing further the number of trees could be beneficial. The number of random tests allows to filter irrelevant attributes and using highest values often yields the best results especially for datasets where all pixels are not equally relevant. Using this scheme, trees should be fully developed ie. without pruning or limiting a priori their depth ($n_{min} = 2$).

4.1.3 *Observations for ET-FL classification scheme*

With ET-FL, increasing the number T of trees (hence the number of features) brings more improvement compared to what we observed with ET-DIC, although the improvement is not always important. Trees should be pruned ie. n_{min} value should be higher (in order to build features that are not too much specific), except for a few problems including

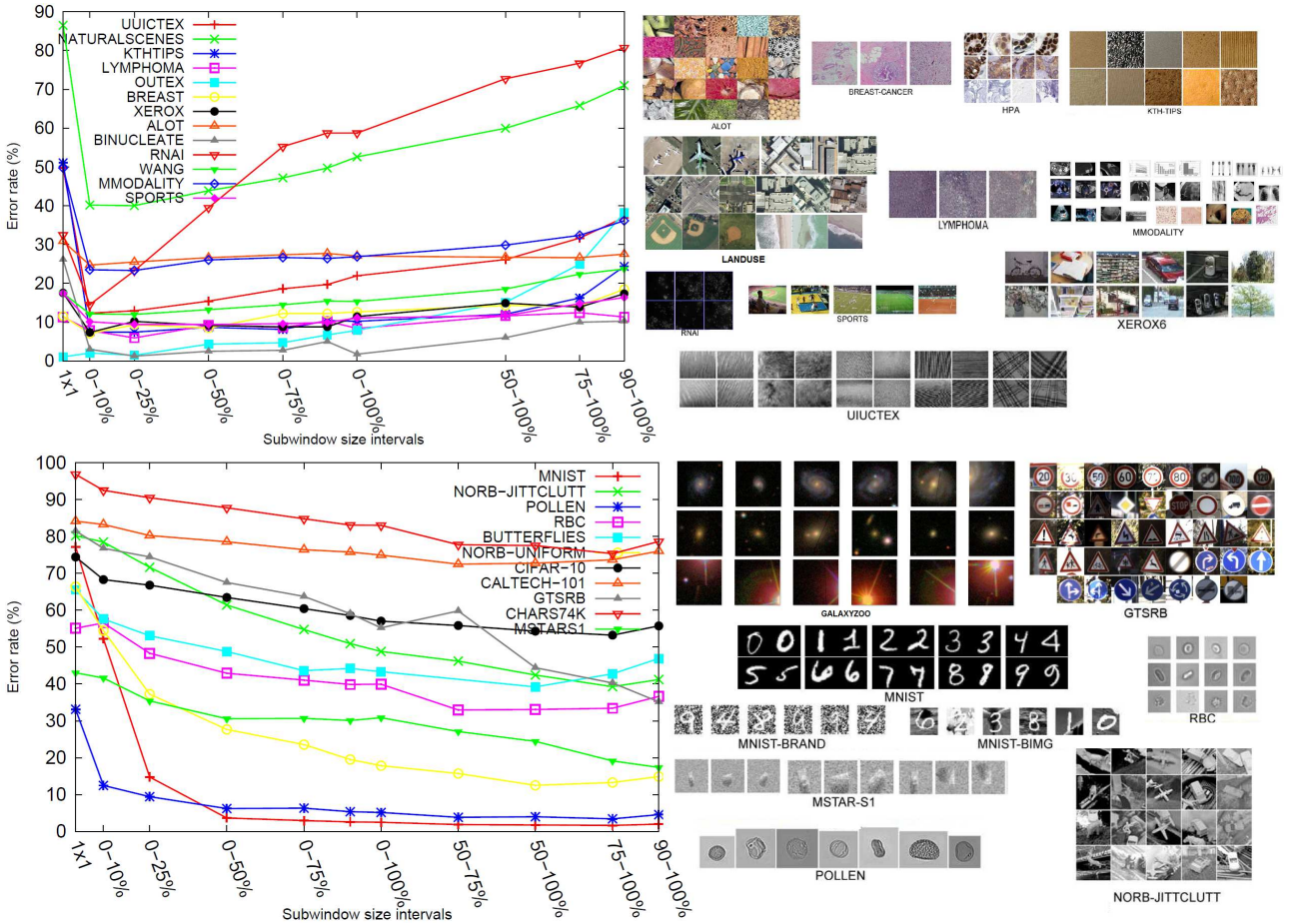

Fig. 4. Error rates using different size intervals with ET-FL. Top: Several datasets for which smalller subwindow sizes yield lowest error rates. Bottom: Several datasets for which larger subwindow sizes perform best.

object identification tasks in controlled conditions (for which specific features work best), but where ET-DIC is in fact performing better. On average, n_{min} = 500 (when using 1 million training subwindows) is the best choice, which yields different numbers of learned features depending on the dataset. Given our results, the recommended encoding to describe images is frequency at terminal nodes which significantly outperforms binary encoding, while hierarchical encoding decreases results when combined with a linear classifier. On average, the default value of the filtering parameter ($k = \sqrt{nbatts}$) achieves better results than unsupervised feature construction ($k = 1$), but increasing that parameter to higher values does not seem so important, although for several problems it is still beneficial to do so.

4.1.4 Comparison of ET-DIC and ET-FL

ET-DIC is slightly better for a quarter of the datasets, including particular object identification datasets in controlled conditions, but ET-FL yields better results on others (60 datasets among 80).

On average on all our datasets, the difference of error rate averages between the two methods is less than 4% once their parameters are optimized, more precisely the average of ET-FL best error rates is 23.10 ± 24.21 while ET-DIC yields an averaged best error rate of 26.78 ± 24.29 , see Figure 6.

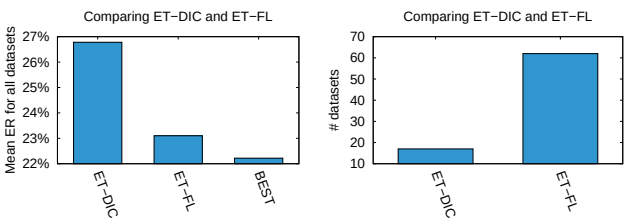

Fig. 6. Comparing ET-DIC and ET-FL. Left: Average of best error rates for all datasets. Right: Number of datasets for which each variant wins.

These results show that on a majority of datasets, the construction of a global image representation based on tree terminal node frequencies subsequently classified by a linear classifier (ET-FL) yields better results compared to the direct classification of individual subwindows (ET-DIC). Although individual subwindows can be strongly predictive with respect to the class of the image they come from (when ET-DIC is performing well), ET-FL allows to describe images by a higher-level representation than raw pixels. It learns image features (from small or large patterns), as each tree leaf contains subwindows that fulfills a serie of tests on pixel intensities in (small to large) subwindows. The final classification model that combines

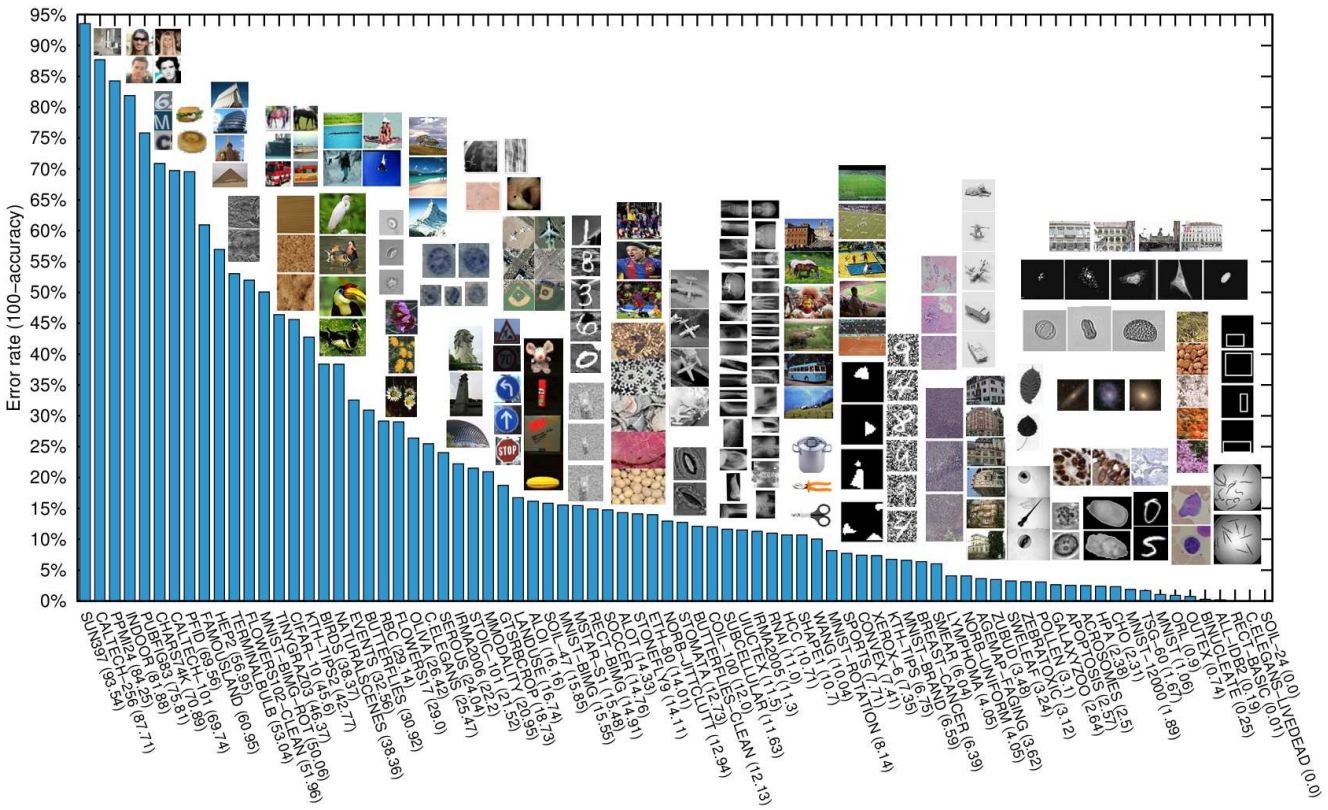

Fig. 5. Summary of the best error rate obtained for each dataset without optimizations, gathered from Supplementary Tables II to XII. Illustrative image examples from ~ 50 datasets are positionned (approximately) according to dataset recognition performances. Note that some results (e.g. on CIFAR-10, GTSRB, and IRMA-2005) are significantly improved using further optimizations, see Section 4.2 and Supplementary Table XIII.

such feature "responses" is more discriminative than the combination of individual predictions for every subwindows (ET-DIC).

4.1.5 Overall results

Regarding overall performances, we achieve more than 80% recognition rate for 52 datasets among 80, and more than 90% recognition rate for 30 datasets (see Figure 5). However, our results are much lower for some other well-known challenging datasets or recently published ones that exhibit a lot of variabilities. In particular we achieve less than 50% recognition rate on 13 datasets, most of them containing images from the web that depict coarse-grained categories (natural scenes or various object/face classes with complex backgrounds and strong intensity and illumination changes). The difficulty induced by complex backgrounds was verified on variants of four datasets (three synthetic object categorization datasets, and a dataset of web images of butterflies) where results were significantly better when objects of interest where cropped on uniform backgrounds. The mean of the best error rate computed over all 80 datasets is 22.22%, while on the subset of 25 bioimaging datasets, the mean best error rate is 12.03%.

Overall, these results (and similarly those obtained by other methods) must be qualified because of limitations of some datasets (see Section 5.2). It has also to be noted that results presented in this paper do not constitute an upper limit of the method accuracy since combining simple optimizations and increasing parameter values (see Section 4.2) might result in higher recognition performances on specific datasets. Interestingly, the potential of our approach for a new, specific, problem could be evaluated in a short period of time as the evaluation of the influence of the main parameters of both classification schemes can be achieved in approximately one day as long as one has access to a small computing cluster (~ 50 cores).

4.2 Further optimizations

In practice, if the method does not achieve satisfactory results on a specific problem, it is possible to further optimize its parameters and implement slight algorithm variations to get better results. Although this alters somewhat the generality of the method, we believe these optimizations (some requiring only a few lines of codes) are simpler than designing a completely new, specific, approach. Although further work is needed to assess if some of these variants could be generalized and applied successfully on

a larger number of datasets, Supplementary Table XIII present promising results obtained with several (combinations of) simple optimizations for a dozen datasets. These optimizations are discussed below.

4.2.1 Extending parameter ranges

A first, straightforward way, to try to improve results, is to simultaneously enlarge ensembles of trees (ie. more than 40 used in previous sections) and the number of extracted subwindows for both learning (i.e. more than 1 million training subwindows used in previous sections) and prediction. Increasing these parameters and tuning the n_{min} value in ET-FL can lead to millions of features as input of the final linear classifier that turned out to be effective for some datasets. In addition, increasing the number of random tests evaluated in each tree node could also help.

4.2.2 Synthetic data

Instead of simply increasing the number of extracted subwindows, augmenting the original training sets with image or subwindow variations can help to improve results. For red-blood cell recognition, right and straight angle rotated and mirrored subwindows improved recognition rates while only increasing the number of extracted training subwindows to 10 millions but without transformation did not bring improvement. Other variations such as spatially-variant blurring, or adding noise, might also be investigated, depending on the application.

4.2.3 Normalizing random subwindow descriptors

Normalizing the pixel value distributions within each subwindow for each RGB channel independently (by substracting the mean and then dividing by the standard deviation) improves significantly results (compared to HSV encoding) on several datasets where strong illumination variations occur (including the traffic signs and characters recognition benchmarks).

4.2.4 Evaluating different node tests in Extra-Trees

More complex node tests can be implemented in decision tree nodes [44], [53], [54]. Instead of single pixel thresholding, we evaluated node tests that threshold the difference of a pixel and one of its 8 direct neighbours (choosen randomly). It improved results on various datasets, including for the recognition of traffic signs, X-Rays, and acrosomes. We hypothesize such improvement is due to the fact that tests on difference of pixel intensities allow to better capture contrast (edges) in images, enable invariance to linear intensity changes, and they also potentially reduce individual decision tree bias [55]. In future work, we are interested to evaluate if performances could be improved if the tree induction algorithm can choose tests among different test types at each node, e.g. linear combination of neighboring pixels, which somehow generalizes the following variant.

4.2.5 Applying filters to original images

On a web-scale object recognition dataset, we applied various filters (linear combination of neighboring pixels to detect oriented lines and contours) and spatial pooling operations (maximum and mean statistics over 3×3, 5×5, 7×7, 9×9, 11×11 neighborhoods) to original images in each RGB color channel independently, somewhat in the same way as first filter bank layers in convolutional network approaches [56]. Then we extracted random subwindows and describe them by raw pixel intensity values from all these filtered images, then build our tree models. On that dataset it improves results very significantly while increasing the number of training subwindows extracted in the original image space to 10 millions did not bring such improvement. It also improved results on two datasets of natural scenes and X-rays, although less dramatically.

4.2.6 Describing subwindows with statistical features

Several approaches combine many statistical features, and let the learning algorithm select the most relevant ones for each problem (such as in [40], [25]). Combining raw pixel descriptors with the default 328 features of [25]) computed within each subwindow yields significant error rate decrease on the red-blood cell dataset, but with a very important computational burden. Alternatively, one might rely on performances of subwindow size intervals to select a priori the type of features to include: if small subwindows yields the best results then texture features might help to improve results, while shape features or pooling of local features within larger spatial regions might perform best if large subwindows yield better results. For example, on dataset of cells in immunofluorescence, small to medium subwindows perform best in our evaluations in line with the best performing method on this dataset using a descriptor that encodes spatial relations among adjacent LBP texture features. However, in the future, we seek to enable decision trees to automatically generate such kind of features by combining previous optimizations in order to improve the versatility of the method.

4.2.7 Other sources of data

Whenever possible, one could combine visual data with other sources of information, such as textual data, or spectral data invisible to the human eye. In [57], we obtained on the dataset of biomedical imaging modalities a significant improvement by combining ET-FL with bags of textual terms extracted from image descriptions. In [58], we used hyperspectral data to better detect materials in outdoor scenes. In a preliminary study on classification of protein crystallization experiments, we obtained better classification results by describing subwindows with pixels from additional images obtained by a rotating-polarizer microscope (data not shown).

5 Discussion and Future Work

In this Section we compare the method with other works, discuss dataset issues, then suggest future work.

5.1 Comparison with other methods

Without a centralized repository of results, gathering state-of-the-art results from the wide computer vision literature for all the datasets included in our study could hardly be up to date. To the best of our knowledge, no other image classification method was evaluated on so many datasets. We will therefore only draw general trends from what we observed. Detailed comparisons for a subset of datasets are provided in Supplementary Material.

First, we compared on several datasets our approach with other approaches using Extremely Randomized Trees. On a few datasets with fixed image sizes, we first compared our approach to the direct application of Extra-Trees without subwindow extraction, ie. where each image is represented by a single input vector encoding all its pixel values. Our results (see Supplementary Tables XIV) were significantly better using our approaches based on subwindow extraction, in particular on datasets where small subwindows yield better results (e.g. on immunostaining patterns) but also on datasets when large subwindows performed best. Compared to [30] using unconstrained subwindow size intervals and ET-DIC on a few datasets, we observe that adjusting parameters (such as the subwindow size intervals, the number of subwindows, the number of random tests, and the classification scheme) can yield very important accuracy improvements. Compared to [32] that uses ET-FL with binary encoding at terminal nodes and used a fixed number of features (by post-pruning) on a few object classes, we observed that quantitative encoding and problem-dependent numbers of learned features (from a few thousands up to millions of features) have a significant influence on results.

Second, we observed the method often performs better than previously published baselines used in original publications presenting several datasets. This is particularly true for global approaches e.g. using classifiers (nearest neighbor classifier with euclidian distance, logistic regression, or SVMs) applied on down-sampled images (see Supplementary Table XV). It also sometimes performs better than first specific methods developed once new datasets were published, e.g. for a building recognition dataset, a sport categorization dataset, a leaf recognition task, a dataset about land uses from overhead imagery, and several bioimaging datasets (See Supplementary Material). On several datasets, our approach is also on par with, or better than, methods using application-specific features (e.g. on galaxy recognition, leaves, zebrafish phenotypes, ...), and better than many other methods (e.g. proposed during international challenges), while not reaching state-of-the-art performances on each and every problem (e.g. on cells in immunofluorescence).

On several other problems (especially datasets with images from the web depicting e.g. wild animals, faces of celebrities, or natural scenes or actions), our results using raw pixel values from original images are not satisfying. On most of these datasets, our approach without optimizations yields worse results than GIST [59], and it is also significantly inferior than more elaborated approaches, e.g. methods combining numerous image descriptors [40], or multi-stage architectures that combine various steps of normalization, filtering and spatial pooling ([60], [61], [62], [63]). On the web-scale object recognition dataset on which we evaluated optimizations using filtered images (see Section 4.2.5), our approach then becomes better than GIST [64] and also slightly better than other multi-stage approaches e.g. tiled convolutional neural networks [65] and factorized third-order Boltzmann Machines [64], but still significantly inferior to the best known method on this dataset [61]. On other problems (such as traffic sign recognition, and synthetic images of object categories), it seems not necessary to perform image filtering to be competitive with a variety of multi-stage approaches.

5.2 Dataset limitations and biases

In image classification, publicly available datasets are essential to enable continued progress, as they allow quantitative evaluation and comparison of algorithms. Recently, some dataset issues have been raised [12], [13], [14], [15], [16], [17], [18], including towards a few datasets we used in our study. These authors have shown that some hidden regularities can be exploited by learning algorithm to classify images with some success, e.g. background environments in face recognition benchmarks [14], [16], and illumination, focus or staining settings in biomedical imaging [18]. These biases specific to some training sets will prevent an algorithm to work on new images and are potentially guiding research in the wrong direction. Moreover, the realism of several benchmarks has to be questionned beside the large amounts of imaging data of high-throughput applications. For example, in diagnostic cytology a single patient sample might contain hundreds of thousands of cells while typical benchmarks contains only a few hundreds individual cells from a limited number of samples, and therefore variations induced by laboratory practice and by biological factors are often not well represented. Here, we suggest to implement a few quality control tests for assessing datasets and detecting biases before publishing and intensively working on new datasets.

Firstly, we propose to evaluate recognition performances with the extraction of 1×1 subwindows (ie.

individual pixels). While this scheme often yields bad recognition rates, it reveals that individual pixel intensities are in some cases (strongly) related to image classes as this setting yields with ET-DIC less than 30% error rate for several problems with between 3 and 201 classes, or even less than 10% error rate for two datasets, including one with more than 50 classes. Similarly, using individual pixels with ET-FL could be seen as a procedure to build an image intensity histogram that yields rather good results on a few datasets, in accordance to other reported results using color histograms. However, for five bioimaging datasets, such settings yield less than 9% error rate, an unexpected result given these datasets contain patterns of cells or tissues that a kind of global histogram approach should not be able to solve. We hypothesize some acquisition artefacts are inadvertently correlated to classes e.g. cell images from a given class were maybe acquired in a single sample (so the classifier might rely on sample acquisition characteristics rather than class-specific patterns), or acquired during imaging sessions with specific parameters, as already discussed by [18].

Secondly, similarly to [14] for face datasets, one can easily evaluate our method accuracy on regions not centered on the objects of interest (e.g. a 50×50 square from the top-left corner of each image). We performed such an experiment on a few datasets and reported results in Supplementary Table XVI. Obviously, a few datasets, including a recently published cell dataset, are biased as recognition rates for all or for a subset of classes are significantly better than majority/random voting while only using background data, as observed by inspection of confusion matrices.

Thirdly, it is possible to use ET-DIC to visualize directly individual subwindow predictions, hence to identify which image regions are mostly used to classify whole images. Such an approach can help to visually identify if patterns of interest or foreground regions are well detected or if unexpected image regions (artefacts or background) contribute to class recognition. For a dataset of natural scenes this allowed us to visualize that many subwindows in the background were discriminative for a subset of classes (see Supplementary Figures 8 and 9).

Anyway, in the future we believe it is important that the imaging acquisition protocols try to reduce the non-relevant differences between image classes (for bioimaging studies, see experimental considerations for effective pattern recognition in [66] and discussion in [18]), and we suggest that acquisition protocols should be better described once a dataset is released so that they can also be peer-reviewed.

5.3 Future algorithmic development

Future work should try to jointly increase the recognition results on the wide range of imagery studied herein while seeking fully autonomous parameter tuning. Interesting future directions towards generic image classification include the combination of both classification schemes, and the combination with ideas from other approaches [60], [61], [67] given our preliminary results with image filtering and complex tree node tests. Also, it would be interesting to evaluate algorithms for the selection of image-level features generated by ET-FL rather than using a simple linear SVM model (e.g. another layer of randomized trees, or the group lasso method). Extensions to deal with multi-label and hierarchical classification (e.g. to address the ImageNet dataset [11]) is also thinkable.

6 Conclusions

This paper addressed the generic problem of supervised image classification without any preconception about image classes. An extensive empirical study on 80 multi-class datasets (among which 25 bioimaging datasets) has been conducted to evaluate overall performances of variants of a method using random subwindows extraction, raw pixel intensity descriptors, and extremely randomized trees either to classify directly images or to learn features. Influence of its few parameters has been reported thoroughly. While the method does not reach state-of-the-art results on each and every problem, it is rather easy to evaluate and it achieves good performances for diverse image collections including images from real-word applications that exhibit various factors of variations. We therefore suggest it could be used as a first try on any new image classification problem. Additionally, a few optimizations were proposed which showed the method can be extended fairly simply to deal with more challenging images.

Overall, our work also emphasizes the need for additional research to design generic and adaptive methods, and evaluate them on very diverse types of imagery to better assess new method potentials and avoid overfitting. Towards that direction, we believe repositories and infrastructures should be promoted to ease the comparison of algorithms. We are thus currently working to integrate in a rich internet application [68] multi-threaded versions of these algorithms together with general-purpose annotation tools in order to ease and speedup the creation of realistic ground-truths and the evaluation of classifiers on large imaging datasets, with the hope to address readily the imaging data deluge occuring in life sciences.

In addition, we have made some observations regarding current dataset biases and we suggested to implement simple dataset quality control tests before publishing new benchmarks.

Author's contributions

R.M. designed and performed the study; R.M. wrote the first draft of the paper; and all authors designed

the methods, analyzed data, contributed to, and approved the final draft.

ACKNOWLEDGEMENTS

R.M. is supported by the CYTOMINE research grant of the Wallonia (DGO6, WIST3, 1017072), and by the GIGA interdisciplinary cluster of Genoproteomics of the University of Liège with financial support from the Wallonia and the European Regional Development fund. R.M. thanks the following persons for technical assistance or fruitful discussions (in alphabetical order): Vincent Botta, Alain Empain, Gilles Louppe, Axel Mathei, Giuseppe Saldi, Benjamin Stévens, and Sarah Ungaro.

REFERENCES

[1] R. F. Murphy, "An active role for machine learning in drug development," *Nature Chemical Biology*, vol. 7, pp. 327–330, 2011.

[2] G. Danuser, "Computer vision in cell biology," *Cell*, vol. 147, no. 5, pp. 973–978, 2011.

[3] N. de Souza, "Machines learn phenotypes," *Nature Methods*, vol. 9, no. 10, 2012.

[4] P. Liberali and L. Pelkmans, "Towards quantitative cell biology," *Nature Cell Biology*, vol. 14, no. 12, December 2012.

[5] F. Li, Z. Yin, G. Jin, H. Zhao, and S. T. C. Wong, "Chapter 17: Bioimage informatics for systems pharmacology," *PLoS Comput Biol*, vol. 9, no. 4, 04 2013.

[6] D. Zahniser, P. Oud, M. Raaijmakers, G. Vooys, and R. Van De Walle, "Biopepr: a system for the automatic prescreening of cervical smears," *Journal of Histochemistry and Cytochemistry*, vol. 27, no. 1, pp. 635–641, 1979.

[7] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, *Toward category-level object recognition*. Springer-Verlag, 2006, vol. 4170.

[8] A. Pinz, "Object categorization," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 4, pp. 255–353, 2006.

[9] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[10] F. Fleuret, T. Li, C. Dubout, E. K. Wampler, S. Yantis, and D. Geman, "Comparing machines and humans on a visual categorization test," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 108, no. 43, pp. 17 621–17 625, 2011.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *Proc. CVPR*, 2009.

[12] J. Ponce, T. L. Berg, M. Everingham, D. A. Forsyth, M. Hebert, S. Lazebnik, M. Marszalek, C. Schmid, B. C. Russell, A. Torralba, C. K. I. Williams, J. Zhang, , and A. Zisserman., *Toward Category-Level Object Recognition*. Springer-Verlag Lecture Notes in Computer Science, 2006, ch. Dataset Issues in Object Recognition.

[13] N. Hervé and N. Boujemaa, "Image annotation: which approach for realistic databases?" in *Proc. ACM International Conference on Image and Video Retrieval (CIVR)*, 2007, pp. 170–177.

[14] L. Shamir, "Evaluation of face datasets as tools for assessing the performance of face recognition method," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 225–230, 2008.

[15] C. D. Pinto N, Barhomi Y and D. JJ, "Comparing state-of-the-art visual features on invariant object recognition tasks," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 2011.

[16] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and Simile Classifiers for Face Verification," in *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009.

[17] A. Torralba and A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[18] L. Shamir, "Assessing the efficacy of low-level image content descriptors for computer-based fluorescence microscopy image analysis," *Journal of Microscopy*, vol. 243, no. 3, pp. 284–292, 2011.

[19] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on PAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[20] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, jun 2007.

[21] G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, pp. 48–62, 2009.

[22] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.

[23] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[24] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[25] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, and I. Goldberg, "Wnd-charm: Multi-purpose image classification using compound transforms," *Pattern Recognition Letters*, vol. 29, no. 11, pp. 1684–1693, 2008.

[26] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, and J. Vandenberg, "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey," *Monthly Notices of the Royal Astronomical Society*, vol. 389, pp. 1179–1189, Sep. 2008.

[27] T. Kotseruba, C. Cumbaa, and I. Jurisica, "High-throughput protein crystallization on the world community grid and the gpu," *Journal of Physics: Conference Series*, vol. 341, 2012.

[28] R. Marée, P. Geurts, G. Visimberga, J. Piater, and L. Wehenkel, "An empirical comparison of machine learning algorithms for generic image classification," in *Proc. 23rd SGAI AI,*, F. Coenen, A. Preece, and A. Macintosh, Eds. Springer, 2003, pp. 169–182.

[29] R. Marée, P. Geurts, J. Piater, and L. Wehenkel, "A generic approach for image classification based on decision tree ensembles and local sub-windows," in *Proceedings of the 6th Asian Conference on Computer Vision*, K.-S. Hong and Z. Zhang, Eds., vol. 2, 2004, pp. 860–865.

[30] ——, "Random subwindows for robust image classification," in *Proc. IEEE CVPR*, vol. 1. IEEE, 2005, pp. 34–40.

[31] R. Marée, P. Geurts, and L. Wehenkel, "Random subwindows and extremely randomized trees for image classification in cell biology," *BMC Cell Biology supplement on Workshop of Multiscale Biological Imaging, Data Mining and Informatics*, vol. 8, no. S1, July 2007.

[32] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Transactions on PAMI*, vol. 30, no. 9, pp. 1632–1646, 2008.

[33] R. Marée, P. Geurts, and L. Wehenkel, "Content-based image retrieval by indexing random subwindows with randomized trees," *IPSJ Transactions on Computer Vision and Applications*, vol. 1, no. 1, pp. 46–57, jan 2009.

[34] M. Dumont, R. Marée, L. Wehenkel, and P. Geurts, "Fast multi-class image annotation with random subwindows and multiple output randomized trees," in *Proc. VISAPP*, 2009.

[35] O. Stern, R. Marée, J. Aceto, N. Jeanray, M. Muller, L. Wehenkel, and P. Geurts, "Automatic localization of interest points in zebrafish images with tree-based methods," in *To appear Proc. 6th IAPR International Conference on Pattern Recognition in Bioinformatics*, ser. Lecture Notes in Bioinformatics. Springer-Verlag, 2011.

[36] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[37] N. Pinto, J. Dicarlo, and D. Cox, "Establishing good benchmarks and baselines for face recognition," in *ECCV 2008 Faces in 'Real-Life' Images Workshop*, October 2008.

[38] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes (VOC)

challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
[39] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.
[40] P. V. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *IEEE International Conference on Computer Vision (ICCV)*, 2009.
[41] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification." in *Proc. ECCV*, 2006, pp. 490–503.
[42] G. H. Le Lu, "Dynamic foreground/background extraction from images and videos using random patches," in *NIPS*, 2006.
[43] B. Yao, A. Khosla, and L. Fei-Fei, "Combining randomization and discrimination for fine-grained image categorization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, June 2011.
[44] A. Criminisi and J. Shotton, Eds., *Decision Forests for Computer Vision and Medical Image Analysis*, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2013.
[45] R. Caruana, N. Karampatziakis, and A. Yessenalina, "An empirical evaluation of supervised learning in high dimensions," in *Proc. International Conference on Machine Learning (ICML)*, 2008.
[46] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 36, no. 1, pp. 3–42, 2006.
[47] L. Breiman, J. Friedman, R. Olsen, and C. Stone, *Classification and Regression Trees*. Wadsworth International (California), 1984.
[48] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
[49] ——, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[50] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.
[51] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. ICCV*, vol. 2, Oct. 2003, pp. 1470–1477.
[52] C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in *ECCV International Workshop on Statistical Learning in Computer Vision*, 2004.
[53] D. Maturana, D. Mery, and A. Soto, "Face recognition with decision tree-based local binary patterns," in *Proceedings of the 10th Asian conference on Computer vision - Volume Part IV*, ser. ACCV'10. Springer-Verlag, 2011, pp. 618–629.
[54] F. Maes, P. Geurts, and L. Wehenkel, "Embedding monte carlo search of features in tree-based ensemble methods," in *European Conference on Machine Learning (ECML'12)*, Bristol, UK, September 2012.
[55] P. Geurts, "Contributions to decision tree induction: bias/variance tradeoff and time series classification," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, University of Liège, May 2002.
[56] Y. LeCun, K. Kavukvuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. International Symposium on Circuits and Systems (ISCAS'10)*, IEEE, Ed., 2010.
[57] R. Marée, O. Stern, and P. Geurts, "Biomedical imaging modality classification using bags of visual and textual terms with extremely randomized trees," in *Working Notes, ImageCLEF LAB Notebook, Multilingual and Multimodal Information Access Evaluation (CLEF)*, CELCT. Springer, Sep 2010.
[58] R. Marée, B. Stevens, P. Geurts, Y. Guern, and P. Mack, "A machine learning approach for material detection in hyperspectral images," in *To appear in Proc. 6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum (CVPR09)*. IEEE, Jun 2009.
[59] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
[60] N. Pinto, D. Doukhan, J. DiCarlo, and D. Cox, "A high-throughput screening approach to discovering good forms of biologically-inspired visual representation," *PLoS Computational Biology*, vol. 5, no. 11, 2009.
[61] D. C. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.
[62] A. Quattoni and A.Torralba, "Recognizing indoor scenes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
[63] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010)*, 2010.
[64] M. Ranzato, A. Krizhevsky, and G. E. Hinton, "Factored 3-way restricted boltzmann machines for modeling natural images," *Journal of Machine Learning Research - Proceedings Track*, vol. 9, pp. 621–628, 2010.
[65] Q. V. Le, J. Ngiam, Z. Chen, D. Chia, P. W. Koh, and A. Y. Ng, "Tiled convolutional neural networks," in *In NIPS*, 2010.
[66] L. Shamir, J. Delaney, N. Orlov, D. M. Eckley, and I. G. Goldberg, "Pattern recognition software and techniques for biological image analysis," *PLoS Computational Biology*, vol. 6, no. 11, 2010.
[67] Y. J. andd C. Huang and T. Darrell, "Beyond spatial pyramids: Receptive field learning for pooled image features," in *Proc. CVPR*, 2012.
[68] R. Marée, B. Stevens, L. Rollus, N. Rocks, X. Moles-Lopez, I. Salmon, D. Cataldo, and L. Wehenkel, "A rich internet application for remote visualization and collaborative annotation of digital slide images in histology and cytology," *BMC Diagnostic Pathology*, vol. 8 S1, 2013.

Raphaël Marée received the PhD degree in computer science in 2005 from the University of Liège, Belgium, where he is now coordinating the CYTOMINE project. His current research interests are in the broad area of machine learning and computer vision techniques with specific focus on their applications in life sciences.

Pierre Geurts graduated as an Electrical Engineer (in computer science) in 1998 and received the PhD degree in applied sciences, in 2002, both from the University of Liège, where he is Associate Professor of Computer Science. His research interests include machine learning and its applications in various fields, such as bioinformatics, computer vision, and computer networks.

Louis Wehenkel graduated as an Electrical Engineer (electronics) in 1986 and received the PhD degree in 1990, both at the University of Liège, where he is full Professor of Electrical Engineering and Computer Science. His research interests lie in the fields of stochastic methods for systems and modeling, machine learning and data mining, with applications in electric power systems planning, operation and control, image analysis and bioinformatics.

Towards Generic Image Classification: an Extensive Empirical Study: Supplementary data

Raphaël Marée, Pierre Geurts, and Louis Wehenkel

I. Introduction

This document includes some supplementary data (results, explanations, tables, figures, and images) which we could not include in the main paper due to space constraints. Topics covered in this supplement are:

- Listing and short description of all 80 datasets and their evaluation protocols: Table I, Figures 1 to 4
- Computational requirements and implementation details of the methods
- Results for both method variants and their parameters
  - ET-DIC subwindow size intervals: Table II
  - ET-DIC subwindow descriptors: Table III
  - ET-DIC number or random tests: Table IV
  - ET-DIC number of trees: Table V
  - ET-DIC minimum node sample size: Table VI
  - ET-FL subwindow size intervals: Table VII
  - ET-FL minimum node sample size: Tables VIII and IX
  - ET-FL number or random tests: Table X
  - ET-FL encoding scheme: Table XI
  - ET-FL number of trees: Table XII
- Results with simple optimizations, Table XIII
- Comparison with other methods
  - Comparison with extremely randomized trees without random subwindows: Table XIV
  - Comparison with baselines: Table XV
  - Comparison with various other approaches
- Dataset bias: Table XVI, Figures 8 to 11

II. Datasets

A. Description of all datasets and evaluation protocols

Datasets	Images	Img sizes	Protocols	Classes	Short description	Ref.
ACROSOMES	1851	$\pm 100 \times 100$	10 RS $\times$ 1296/555	2	intact/damaged acrosomes of boar spermatozoa	[1]
AGEMAP-FAGING	850	1388×1040	10 RS 423 $\times$ 106	4	mouse liver tissues at different development stages	[2]
ALL-IDB2	260	256×256	10 RS $\times$ 208/52	2	normal and lymphoblast cells	[3]
ALOI	72000	192×144	I 1000/71000	1000	uniform background, viewpoint changes	[4]
ALOT	25000	384×286	10 RS $\times$ 10000/15000	250	textures with varying illuminations	[5]
APOPTOSIS	700	67×67	10 RS $\times$ 630/70	2	DIC images of (non-)apoptotic cells	[6]
BINUCLEATE	40	640×512	10 RS $\times$ 20/20	2	DAPI images of binucleate and regular cells	[2]
BIRDS	600	various	10 RS $\times$ 300/300	6	background, illumination, orientation	[7]
BREAST-CANCER	361	760×570	10 RS $\times$ 289/72	3	biopsies of breast cancer (H&E straining)	[8]
BUILDINGSAB	249	$\pm 800 \times 534$	10 RS $\times$ 68/181	68	grayscale buildings	[9]
BUTTERFLIES	619	various	10 RS $\times$ 182/437	7	www butterfly pictures	[10]
BUTTERFLIES-CLEAN	619	various	10 RS $\times$ 182/437	7	same but butterflies are cropped	[10]
CALTECH-101	9145	various	10 RS $\times$ 3030/5168	101	objects and clutter	[11]
CALTECH-256	29780	various	10 RS $\times$ 7680/6400	256	objects and clutter	[12]
C.ELEGANS	237	1600×1200	10 RS $\times$ 120/117	4	C.elegans muscles at different ages	[2]
C.ELEGANS-LIVEDEAD	97	$\pm 400 \times 400$	10 RS $\times$ 78/19	2	C.elegans live/dead assays	[13]
CHARS74KEIMG	7705	$\pm 50 \times 50$	10 RS $\times$ 930/930	62	characters in natural images	[14]
CHO	327	512×382	10 RS $\times$ 100/65	5	subcellular localizations	[15]
CIFAR-10	60000	32×32	I 50000/10000	10	tiny object/scene images	[16]
COIL-100	7200	128×128	I 100/7100	100	uniform background, viewpoint changes	[17]
CONVEX	58000	28×28	I 8000/50000	2	single white convex regions on black background	[18]
ETH-80	3280	256×256	20 RS 2952/328	8	man-made objects	[19]
EVENTS	1579	various	10 RS $\times$ 560/480	8	sport scenes	[20]
FAMOUSLAND	1000	$\pm 300 \times 200$	10 RS $\times$ 800/200	100	www pictures of famous places	[21]
FLOWERS17	1700	various	10 RS $\times$ 1360/340	17	flowers, pose, light variations	[22]
FLOWERS102	8189	$> 500 \times 500$	10 RS $\times$ 2040/6149	102	larger set of flowers	[23]
GALAXYZOO	75746	200×200	10 RS $\times$ 49991/25755	4	galaxies from imaging sky survey	[24]
GTSRBCROP	51839	$\pm 75 \times 75$	I 39209/12630	43	traffic signs	[25]
HCC	363	$\pm 3096 \times 4140$	20 RS $\times$ 266/14	14	Human carcinoma cell lines	[26]
HEP2	721	$\pm 100 \times 100$	10 RS 361/360	6	cells in indirect immunofluorescence	[27]
HPA	1057	$\pm 64 \times 64$	10 RS $\times$ 952/105	4	Immunostaining patterns	[28]
INDOOR	15620	$> 200 \times 200$	10 RS $\times$ 5360/1340	67	indoor scenes	[29]
IRMA 2005	10000	$< 512 \times 512$	I 9000/1000	57	human body radiographs	[30]
IRMA 2006	11000	$< 512 \times 512$	I 10000/1000	116	human body radiographs (fine-grained)	[30]
KTH-TIPS	810	200×200	10 RS $\times$ 400/410	10	scale, illumination, pose changes	[31]
KTH-TIPS2	4752	200×200	20 RS 3564/1188	11	texture categorization	[32]
LANDUSE	2100	256×256	10 RS 1680/420	21	land use from overhead imagery	[33]
LYMPHOMA	374	1388×1040	10 RS $\times$ 337/37	3	biopsies of lymphoma (H&E staining)	[2]
MMODALITY	5010	$\pm 1200 \times 1200$	I 2390/2620	8	biomedical imaging modalities	[34]
MNIST	70000	28×28	I 60000/10000	10	handwritten digits centered on uniform background	[35]
MNIST-12000	62000	28×28	I 1000/50000	10	subset of MNIST	[18]
MNIST-ROTATION	62000	28×28	I 12000/50000	10	digits with random rotation	[18]
MNIST-BIMG	62000	28×28	I 12000/50000	10	digits with background image	[18]
MNIST-BRAND	62000	28×28	I 12000/50000	10	digits with random background noise	[18]
MNIST-BIMG-ROT	62000	28×28	I 12000/50000	10	digits with random background noise and rotation	[18]
MSTAR-S1	3617	128×128	I 1860/1757	3	synthetic aperture radar images	[36]
NATURALSCENES	4485	$\pm 300 \times 200$	10 RS $\times$ 1500/2985	15	graylevel natural scenes (OLIVA superset)	[37]
NORB-UNIFORM	58600	96×96	I 24300/24300	5	normalized object sizes and uniform background	[38]
NORB-JITTCLUTT	349920	108×108	I 291600/58320	6	jittered objects and cluttered background	[38]
OLIVA	2688	256×256	10 RS $\times$ 1200/1488	8	color images of natural scenes	[39]
ORL	400	92×112	10 RS $\times$ 200/200	40	graylevel, centered, faces	[40]
OUTEX	864	128×128	10 RS $\times$ 432/432	54	color textures	[41]
PFID	1098	$\pm 250 \times 250$	10 RS $\times$ 732/366	61	segmented fast food items	[42]
POLLEN	6039	25×25	10 RS $\times$ 5490/549	7	pollen grains	[2]
PPMI24	4800	258×258	I 2400/2400	24	people holding or playing instruments	[43]
PUBFIG83	13838	100×100	10 RS 7470/830	83	unconstrained faces of celebrities	[44]
RBC	5062	128×128	10 RS $\times$ 1500/1125	3	red-blood cells	[45]
RECT-BASIC	51200	28×28	I 1200/50000	2	tall or wide rectangles at variable position	[18]
RECT-BIMG	62000	28×28	I 12000/50000	2	same with backgrounds	[18]
RNAI	200	1024×1024	10 RS $\times$ 160/40	10	cells following RNA interference	[2]
SEROUS	3652	$\pm 25 \times 25$	10 RS $\times$ 3286/366	11	serous cells	[46]
SHAPE1	724	500×500	10 RS $\times$ 200/100	10	geometric object classes	[47]
SMEAR	917	$< 500 \times 500$	10 RS $\times$ 825/92	2	Cells from PAP smears	[48]
SOCCER	280	$\pm 300 \times 300$	10 RS $\times$ 175/105	7	soccer teams	[49]
SPORTS	2449	$\pm 500 \times 375$	10 RS $\times$ 1837/612	5	www pictures of sports	[50]
SOIL-24	264	$\pm 360 \times 288$	I 24/240	24	man-made objects, uniform background	[51]
SOIL-47	987	$\pm 360 \times 288$	I 47/940	47	man-made objects, uniform background	[51]
STOIC-101	3847	$\pm 320 \times 240$	I 3187/660	101	color buildings in Singapour	[52]
STOMATA	114	$\pm 320 \times 320$	10 RS $\times$ 103/11	2	opened/closed stomata in *Arabidopsis thaliana*	[6]
STONEFLY9	3845	$> 1000 \times 1000$	I 2697/1148	9	stonefly species	[53]
SUBCELLULAR	948	$\pm 512 \times 382$	10 RS $\times$ 862/86	10	subcellular localizations	[54]
SUN397	108754	$> 200 \times 200$	10 RS $\times$ 19850/19850	397	scenes from abbey to zoo	[55]
SWELEAF	1875	$> 1500 \times 1500$	10 RS $\times$ 1125/750	15	swedish leaves, uniform background	[56]
TERMINALBULB	970	300×300	10 RS $\times$ 280/690	7	DIC of pharynx terminal bulb	[2]
TINYGRAZ03	1148	32×32	10 RS $\times$ 1033/115	20	tiny object/scene images	[57]
TSG-60	180	320×240	10 RS $\times$ 120/60	60	color buildings	[58]
UIUCTEX	1000	640×480	10 RS $\times$ 500/500	25	grayscale textures	[59]
WANG	1000	384×480	10 RS $\times$ 500/500	10	from Corel collection	[60]
XEROX-6	1082	various	10 RS $\times$ 984/98	6	users' pictures of objects	[61]
ZEBRATOXIC	1249	696×520	I 1153/96	3	developmental status of zebrafish embryos	[62]
ZUBUD	1120	640×480	I 1005/115	201	color buildings in Zurich	[63]
TOTAL	1517317	-	-	3854	-	

TABLE I

SUMMARY OF DATASET CHARACTERISTICS AND EVALUATION PROTOCOLS (I MEANS ONE RUN WITH INDEPENDANT LEARNING AND TEST SETS OF GIVEN SIZES, X RS MEANS X RUNS WITH RANDOM SAMPLING OF INDEPENDANT LEARNING AND TEST SETS OF GIVEN SIZES).

B. Illustration of biomedical datasets

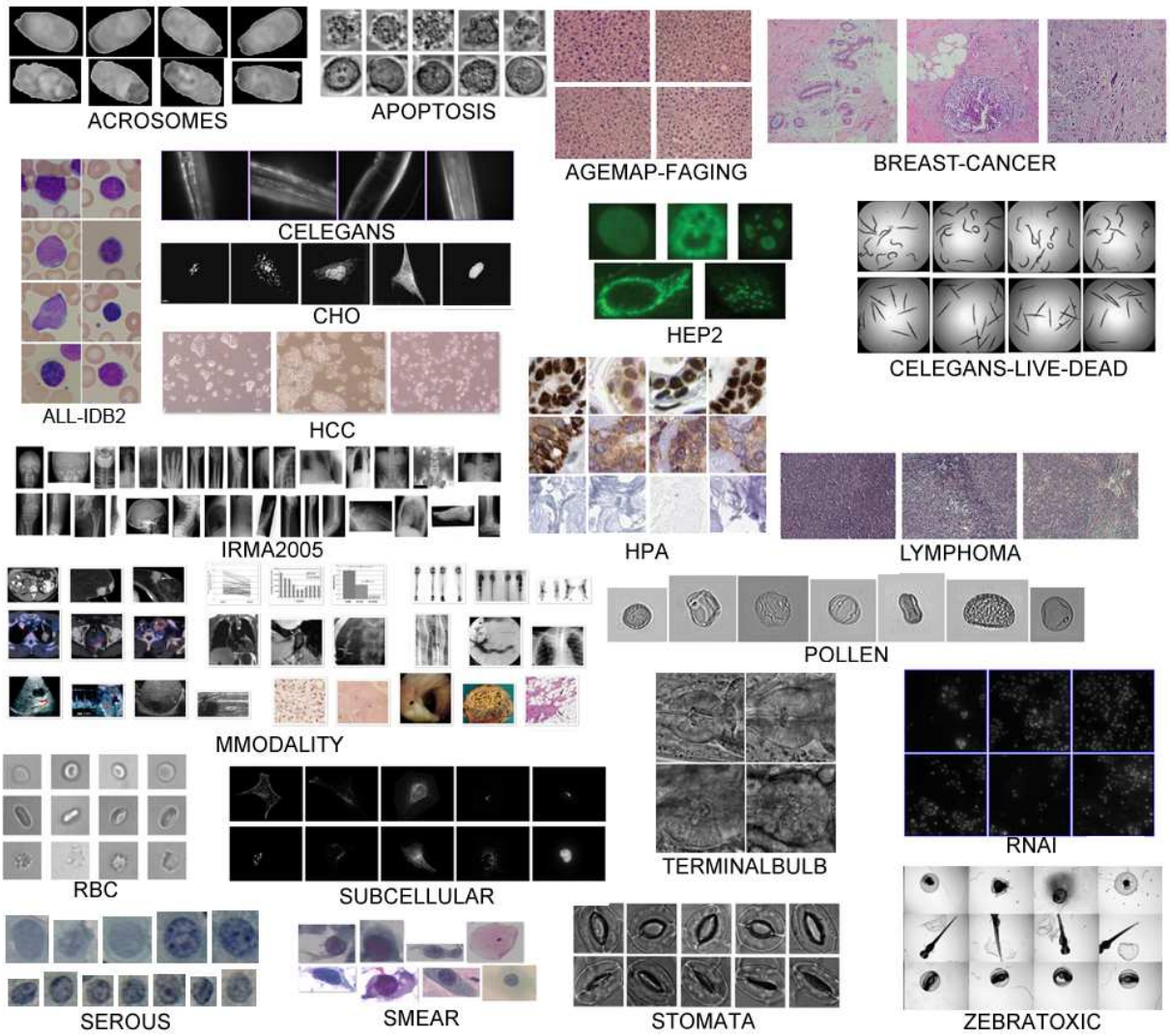

Fig. 1. Illustration of bioimaging classification datasets tackled in our experiments, including various cells, tissues, embryos, preparation protocols and imaging modalities.

C. Illustration of other datasets

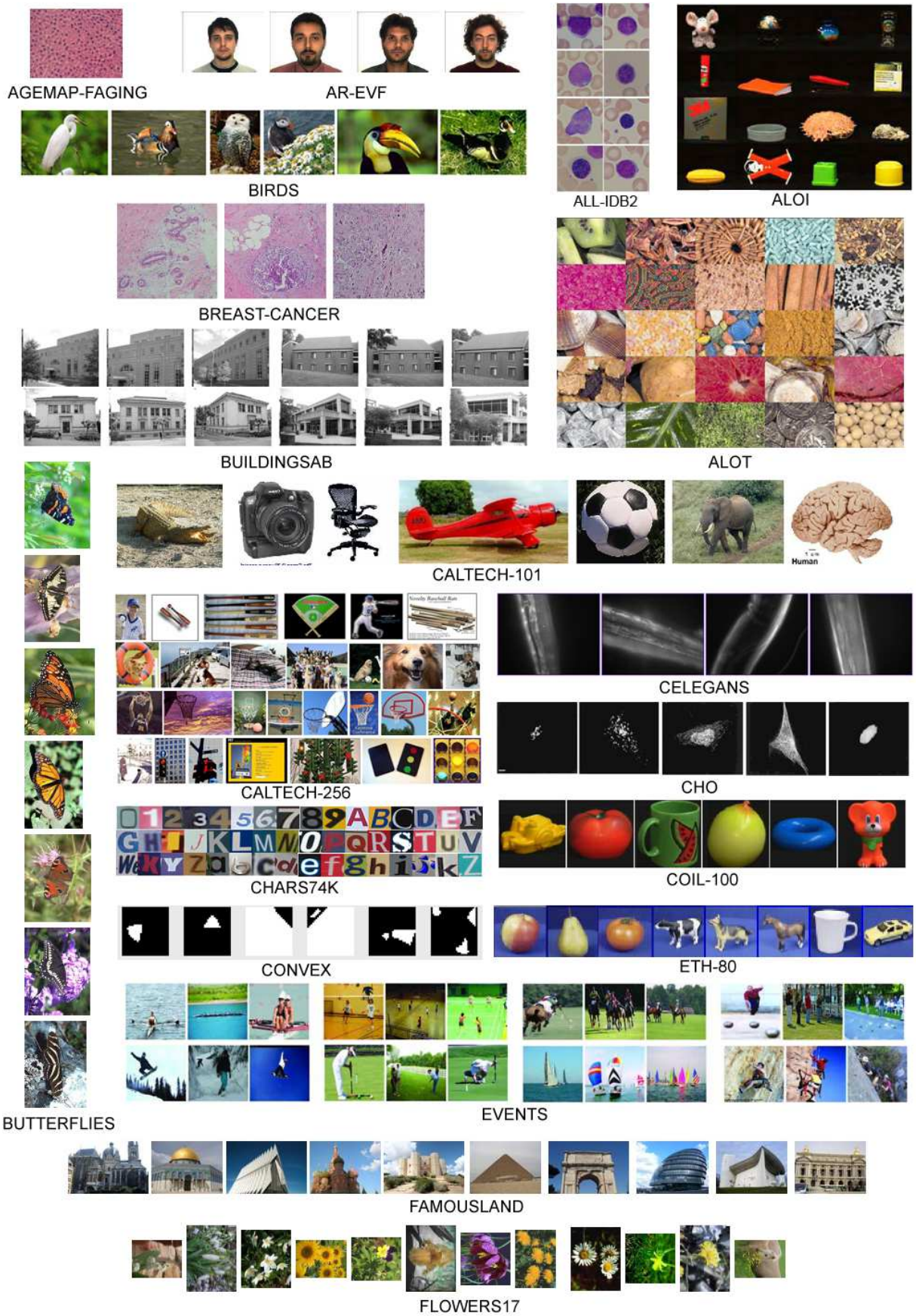

Fig. 2. An overview of several image classes used in our experimental setup. Images taken from publications or websites of the datasets from AGEMAP-FAGING to FLOWERS17 (see Table I for a summary of dataset characteristics).

Fig. 3. An overview of several image classes used in our experimental setup. Images taken from publications or websites of the datasets from FLOWERS102-CLEAN to RNAI (see Table I for a summary of dataset characteristics).

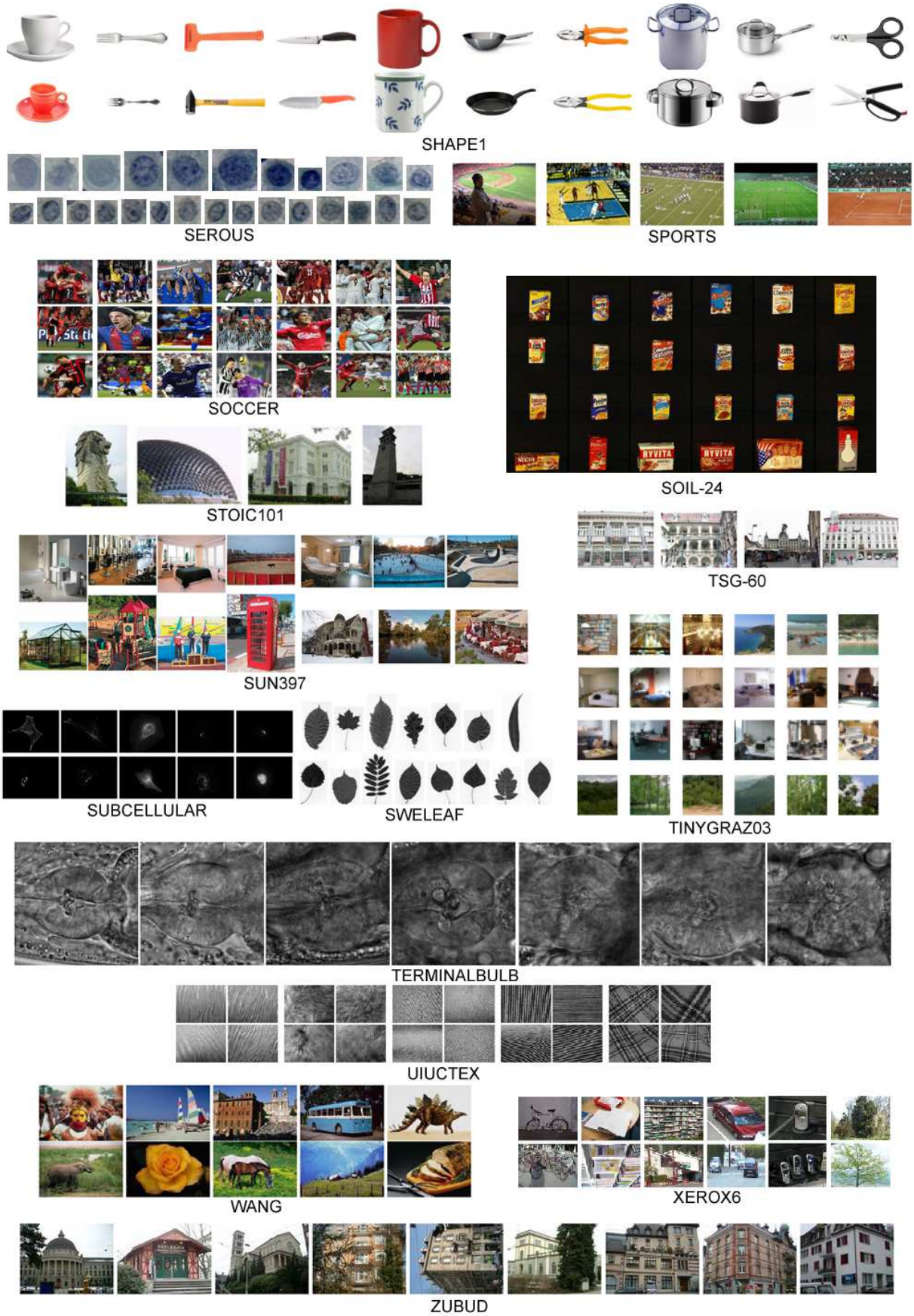

Fig. 4. An overview of several image classes used in our experimental setup. Images taken from publications or websites of the datasets from SEROUS to ZUBUD (see Table I for a summary of dataset characteristics).

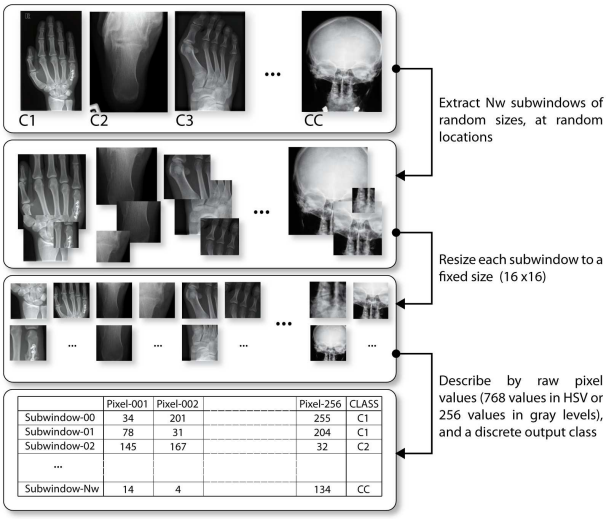

Fig. 5. Random subwindows extraction and description by raw pixel values.

III. METHODS

A. Computational requirements

The computational requirements could be subdivided in the method steps.

1) Random Subwindows: As location and size of random subwindow are choosen randomly, the detection process is very fast compared to common interest point detection methods that involve several image operations (normalization, convolution, resampling, ...). Then, resizing subwindows to a fixed-size essentially depends on the image size (number of pixels) in original subwindows, which could be time-consuming when large subwindows are extracted from large images. In practice, we observed nearest neighbor interpolation yields comparable recognition results (although on a few datasets recognition results were lower) while being much faster than bilinear or bicubic interpolation.

2) Extra-Trees: The time complexity of the Extra-Trees training algorithm is on the order of $kTN_{ls}logN_{ls}$. The size of the Extra-Trees may grow substantially with very large training sets of subwindows, but the propagation step is on the order of $N_{test}Td$, where N_{test} denotes the number of subwindows extracted from an image, and d the average tree depth (which is on the order of $logN_{ls}$ as we observed in our experiments). Thus, the approach scales very well and, moreover, it is highly parallelizable.

Regarding space complexity, during learning, each tree of the ensemble is learned individually and could be saved after its construction. Thus the minimum memory requirement is the memory space necessary for one tree, and the memory to store the training set of subwindows. A binary decision tree has a theoretical maximum number of nodes equals to $2N_{ls} - 1$ nodes when fully developed, but this number is generally smaller in practice. In internal nodes, an attribute index and a real-valued threshold are stored, in addition to two pointers to successor nodes. In each terminal node, for ET-DIC we store a vector of probability estimates which dimension equals to the number of classes. This could be highly memory demanding when there are a lot of classes and when we use a large number of training subwindows and trees are deeply constructed. For ET-FL, there is no information stored at terminal nodes, global image representation being calculated on-the-fly, so the tree ensemble model is significantly lighter.

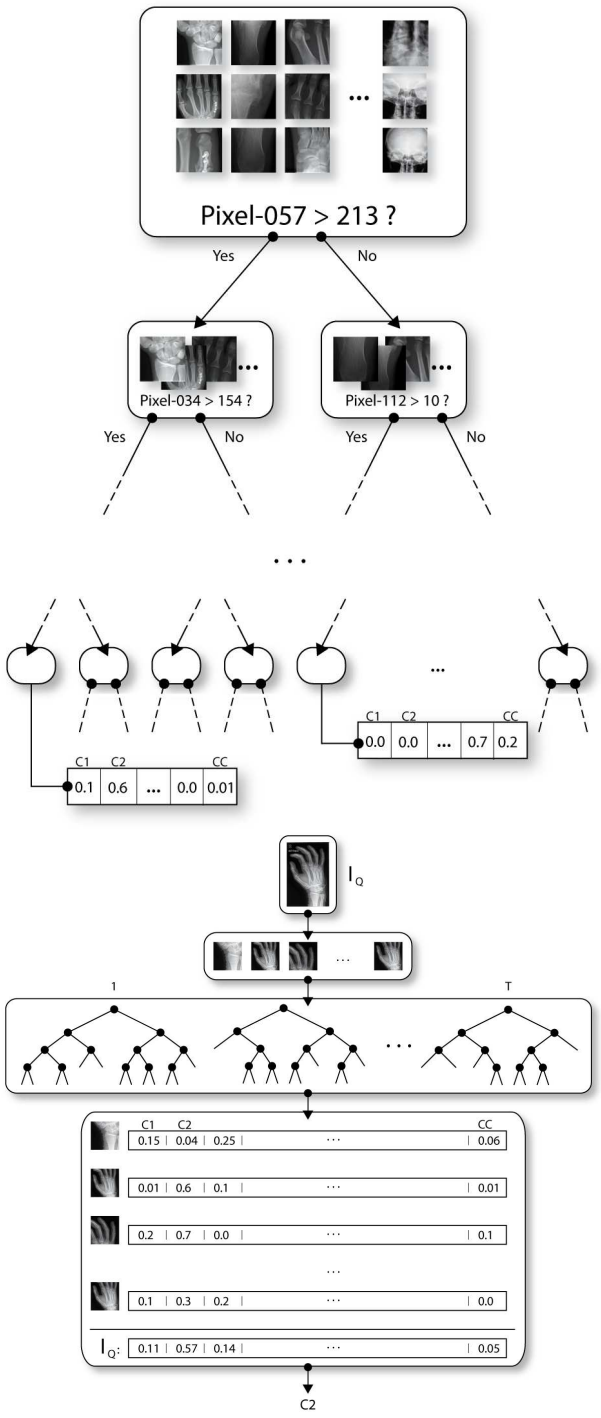

Fig. 6. ET-DIC. Training one tree (left), prediction with an ensemble of trees (right).

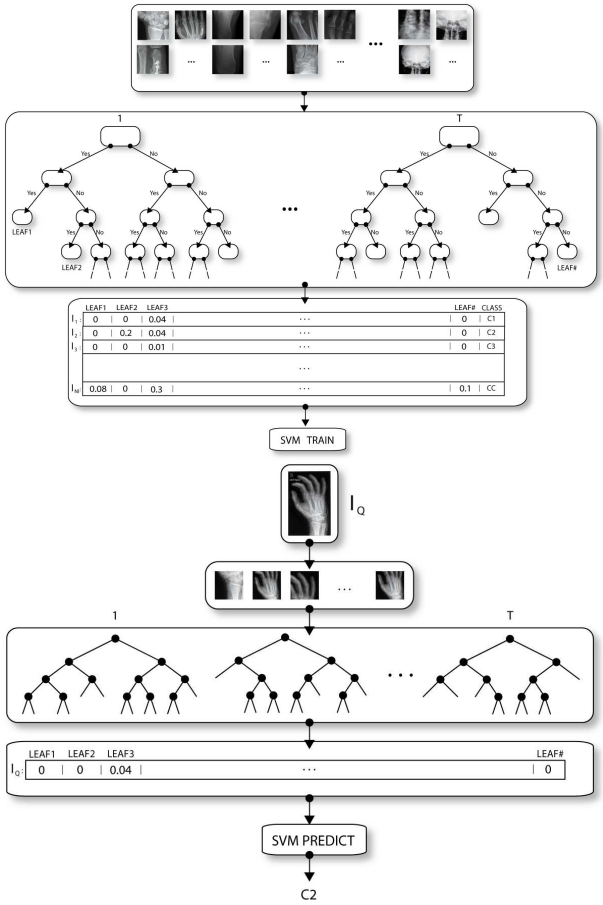

Fig. 7. ET-FL: training an ensemble of trees from random subwindows, building global feature vectors and training SVM (left); propagation of random subwindows of a test image in the ensemble of trees, building of its global feature vector, prediction of its class with SVM classifier (right).

3) Final classifier in ET-FL: For the ET-FL variant, the final classifier has to be built which space and time complexity depend on the specific implementation. With Libsvm, time complexity depends on the number of iterations which may be higher than linear to the number of training images [64]. With Liblinear implementation, time complexity scales linearly with the number of training images.

4) Implementation details: All our experiments were done using our Java implementation of random subwindow extraction and Extra-Trees. For ET-FL, we integrated the Java implementation of the SVM linear classifier of LIBSVM $v2.85$ [64] that uses a "one-against-one" approach for multi-class problems, with default parameters (C-SVC with cost $C = 1$, $\epsilon = 0.1$). For larger problems (e.g. where the number of instances or classes was very large, and for $n_{min} < 1000$ and $T > 10$) we used the faster LIBLINEAR [65] tool $v1.34$ also using default parameter values (L2-loss SVM (dual) solver, $C = 1$, $\epsilon = 0.1$). As suggested by authors of these packages, each attribute was normalized using the svm-scale tool before training the SVM model and the normalization parameters obtained on the training set are used to scale each test instance.

5) Availability: To ease future research and comparison, we offer our single-threaded, command-line, Java implementation of the method on request. We are also integrating in a rich internet application [66] multi-threaded versions of these algorithms in Python based on the scikit-learn toolbox [67]. See http://www.montefiore.ulg.ac.be/~maree/ for updates.

IV. RESULTS

A. Tables of method variants and influences of parameter values

We hereby report results for each parameter studied individually for both variants ET-DIC and ET-FL. Otherwise stated, our experiments are run using a fixed number of subwindows for training (a total of approximately 1 million training subwindows), those are resized to 16×16 pixels encoded by raw pixel values in greylevels or in the HSV colorspace if images are colorized, $T = 10$ trees are built with the filtering parameter equals to the rounded square value of the number of attributes (ie. 16 for greylevel images and 28 for color images), $n_{min} = 2$ corresponding to fully developped trees in ET-DIC and $n_{min} = 5000$ for ET-FL, terminal node frequency encoding for ET-FL, and a fixed number of test subwindows. In order to study the influence of each parameter individually, in each following table we vary a single parameter while the others remain constant. We first find the best subwindow size intervals for each method (ET-DIC in Table II and ET-FL in Table VII), then other parameter influences are evaluated based on the best size interval for each dataset and each method (ET-DIC in Tables III to VI, and ET-FL in Tables VIII to XII).

EVALUATION OF THE SIZE INTERVALS OF SUBWINDOWS ON ERROR RATES WITH ET-DIC. OTHER PARAMETERS ARE CONSTANT: $n_{min} = 1$, $T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS.

TABLE II

Datasets	1×1	$0 - 10\%$	$0 - 25\%$	$0 - 50\%$	$0 - 75\%$	$0 - 90\%$	$0 - 100\%$	$25 - 50\%$	$25 - 75\%$	$25 - 100\%$	$50 - 75\%$	$50 - 100\%$	$75 - 100\%$	$90 - 100\%$
ACROSOMES	37.87 ± 1.00	19.48 ± 3.12	9.75 ± 1.22	6.27 ± 1.27	5.26 ± 1.13	6.88 ± 1.46	6.29 ± 1.06	4.02 ± 0.31	4.99 ± 0.86	5.86 ± 0.94	5.60 ± 0.83	5.96 ± 0.91	5.59 ± 0.69	5.78 ± 0.97
AGEMAP-FAGING	54.48±3.01	55.99±1.94	55.43±4.08	56.10±5.19	53.62±3.19	54.10±2.98	53.71±3.74	51.90±4.92	51.71±3.90	49.62±4.41	49.14±2.56	50.29±3.27	47.90±4.78	47.23±3.02
ALL-IDB2	8.85±3.57	5.96±2.50	3.08±2.88	1.73±2.35	1.92±1.72	2.31±1.15	1.92±1.49	1.73±1.60	0.96±1.29	1.35±1.23	1.92±1.92	2.12±1.60	1.15±1.28	2.50±2.28
ALOI	65.59	50.22	41.81	30.47	24.08	23.01	22.63	23.03	22.57	22.63	22.87	23.33	25.00	25.76
ALOT	66.12±0.20	44.70±0.38	43.19±0.48	42.74±0.30	42.11±0.37	41.40±0.38	40.76±0.38	38.8±0.37	39.08±0.36	38.47±0.29	37.48±0.32	36.67±0.40	34.08±0.40	31.64±0.36
APOPTOSIS	42.86±0.0	31.71 ± 1.40	19.71 ± 2.91	12.29 ± 3.90	12.43 ± 3.56	14.57 ± 1.40	12.43 ± 3.89	13.14 ± 2.98	11.86 ± 3.94	12.43 ± 3.84	12.14 ± 2.87	11.71 ± 2.84	8.14 ± 2.86	11.29 ± 3.03
BINUCLEATE	48.72±0.00	4.36±5.26	4.36±5.85	7.44±4.36	6.15±6.51	8.97±4.33	4.62±3.59	6.41±6.20	9.74±5.94	6.92±7.07	12.56±6.12	7.95±6.33	11.54±7.80	12.56±5.67
BIRDS	66.07±1.93	62.8±2.69	58.5±1.29	56.77±1.65	51.87±2.26	51.30±2.15	50.07±2.50	50.67±1.53	47.60±2.37	48.60±2.05	44.97±3.43	46.53±2.29	46.60±3.34	49.10±2.62
BREAST-CANCER	20.69±2.59	15.00±5.23	14.72±3.06	18.06±5.38	14.44±3.41	16.11±2.93	17.78±3.15	17.78±3.39	13.33±4.49	14.72±4.13	15.28±5.16	14.58±1.89	16.94±3.15	17.78±3.72
BUILDINGSAB	82.82±2.15	44.42±3.04	42.82±2.07	44.59±1.75	44.59±3.42	45.91±2.10	47.29±3.54	45.03±2.23	47.29±2.17	48.01±2.15	48.39±2.71	50.77±2.22	58.67±2.02	66.29±2.45
BUTTERFLIES	69.31±2.54	62.88±1.68	58.44±2.51	53.34±1.47	51.10±2.11	51.92±2.72	52.81±3.06	50.64±2.09	50.85±2.08	51.60±3.89	50.43±3.03	49.86±2.39	54.30±2.37	56.84±2.35
BUTTERFLIES-CLEAN	42.8±2.46	29.4±2.36	24.22±3.39	21.78±2.49	20.91±1.96	21.22±2.25	20.58±1.61	17.80±1.41	18.69±1.43	18.56±1.49	20.29±1.42	22.49±1.26	24.64±1.78	30.02±1.55
CALTECH-101	88.24	85.85	82.38	79.8	76.89	76.11	74.66	77.11	74.80	72.92	73.43	71.82	70.61	73.21
CALTECH-256	96.47±0.21	94.92±0.08	93.82±0.33	92.24±0.29	91.01±0.22	90.12±0.37	89.84±0.37	90.57±0.34	89.64±0.26	88.83±0.36	88.82±0.40	88.30±0.24	88.26±0.26	89.39±0.19
C.ELEGANS	46.75±3.11	41.54±4.68	38.63±4.97	44.53±6.12	46.75±4.56	47.52±3.16	48.46±3.27	40.68±3.67	46.07±3.41	50.17±3.22	46.32±4.29	48.55±3.95	48.46±3.20	50.34±2.40
C.ELEGANS-LIVE-DEAD	47.89±1.58	10.53 ± 4.08	7.37 ± 6.32	4.74 ± 2.83	3.16 ± 3.49	5.79 ± 3.68	5.26 ± 4.08	6.84 ± 5.79	7.37 ± 4.82	3.16 ± 2.58	3.68 ± 3.37	7.89 ± 6.34	4.74 ± 4.37	6.84 ± 5.29
CHARS74K	97.29±0.36	93.55±0.88	91.58±0.84	87.86±0.99	83.31±1.00	81.86±1.03	79.83±1.06	86.67±1.00	81.87±1.43	77.78±1.57	77.69±1.37	74.20±0.94	71.16±1.66	73.99±1.67
CHO	27.38±3.63	2.92±1.75	3.38±2.15	2.62±1.83	4.00±2.19	3.54±2.49	5.54±3.45	3.85±2.94	5.23±1.71	5.85±2.73	8.46±3.10	8.77±3.23	14.77±4.25	21.38±4.98
CIFAR-10	77.70	69.07	67.59	64.21	61.16	58.71	57.45	59.91	57.25	54.02	52.89	50.84	48.91	49.61
COIL-100 nv1	27.75	15.11	12.66	12.15	12.73	13.11	13.72	14.20	14.79	16.21	19.70	19.87	25.25	31.96
CONVEX	50.60	35.72	22.86	15.87	12.58	10.07	8.80	14.36	11.86	8.49	10.83	8.04	9.34	13.00
ETH-80	41.04±10.35	34.67±8.17	28.72±6.35	28.51±8.22	19.15±8.00	21.31±7.48	22.13±6.26	22.26±9.92	22.07±5.65	21.43±21.80	21.80±5.83	17.74±7.39	16.83±5.17	17.44±6.05
EVENTS	49.92±2.04	43.71±2.20	43.25±1.98	42.29±1.89	37.48±1.63	37.33±2.18	37.38±2.47	38.71±2.40	37.19±2.16	36.60±1.64	37.23±1.73	36.04±1.28	37.19±2.93	38.13±2.41
FAMOUSLAND	88.25±2.03	84.45±2.37	81.75±2.37	77.7±2.75	74.85±1.98	73.75±1.86	74.3±2.93	73.55±2.62	71.9±1.95	70.25±1.74	66.85±3.55	70.45±1.52	72.05±1.39	74.15±1.96
FLOWERS17	52.53±1.69	43.82±2.75	41.35±2.53	35.94±0.93	37.32±2.15	35.59±1.93	33.97±2.00	35.06±1.74	33.50±2.51	33.71±1.98	33.88±3.18	33.24±1.16	34.79±2.43	38.44±1.83
FLOWERS102-CLEAN	78.23	70.97	67.54	62.84	60.41	59.04	58.47	59.01	57.76	58.08	57.19	57.38	60.17	64.99
GALAXYZOO	29.93±0.12	26.39±0.14	21.19±0.19	14.76±0.14	12.69±0.23	11.88±0.21	11.36±0.20	11.42±0.15	10.94±0.15	10.09±0.15	9.68±0.27	8.78±0.20	6.43±0.16	4.22±0.17
GTSRBCROP	83.09	77.48	73.65	64.38	56.63	51.61	48.04	52.53	48.96	43.34	37.80	37.45	31.55	29.01
HCC	45.71 ± 13.05	38.21 ± 8.23	37.50 ± 10.80	39.64 ± 10.22	38.57 ± 10.93	40.00 ± 10.20	38.93 ± 12.67	42.14 ± 9.01	36.43 ± 11.93	37.50 ± 12.55	42.86 ± 11.74	40.00 ± 9.15	33.21 ± 14.86	37.14 ± 10.50
HEP2-ICPR2012	66.89	66.21	63.62	64.17	64.03	63.62	64.44	63.90	62.67	63.90	64.17	64.85	64.85	65.12
HPA	17.43±1.91	14.67 ± 1.96	10.95 ± 2.60	8.38 ± 3.12	9.71 ± 2.33	8.76 ± 2.16	8.86 ± 1.95	6.19 ± 1.30	9.52 ± 2.79	7.52 ± 1.73	7.71 ± 2.46	9.24 ± 2.09	12.19 ± 2.20	14.57 ± 2.04
INDOOR	91.68±0.45	87.84±0.84	87.68±0.82	87.60±0.89	86.96±0.64	86.66±0.94	86.70±1.16	86.63±1.03	86.46±0.53	85.68±1.11	86.53±0.72	85.09±0.90	85.24±0.88	86.61±0.99
IRMA2005-BALANCED	93.20	45.90	32.40	23.90	18.70	16.90	16.5	20.60	17.40	15.70	15.70	14.90	15.10	16.80
IRMA2005	70.3	69.8	63.5	52.2	40.8	36.2	33.60	36.70	30.80	26.80	22.5	19.90	16.90	18.30
IRMA2006-BALANCED	95.60	64.60	47.5	37.0	33.5	31.30	29.90	34.00	31.80	28.80	30.5	28.00	26.40	27.50
KTH-TIPS	63.27±1.63	22.61±1.03	23.85±2.44	27.44±1.08	27.56±0.89	28.37±1.19	28.71±1.54	25.17±1.93	26.49±1.59	27.39±1.10	26.85±1.65	26.61±2.36	26.80±2.48	31.78±1.81
KTH-TIPS2	60.24±5.87	49.81±9.07	48.81±7.02	48.32±8.66	48.56±9.18	50.72±7.81	50.50±11.20	42.77±8.41	46.94±9.90	51.33±7.15	46.25±10.74	49.64±7.83	47.42±9.32	46.97±7.27
LANDUSE	40.17±2.58	30.71±1.15	30.26±1.65	29.57±2.11	30.81±1.22	30.19±3.11	30.50±1.53	29.21±2.57	31.00±1.67	30.67±1.64	30.24±1.68	30.19±1.61	33.24±2.68	43.14±1.86
LYMPHOMA	35.41±4.90	23.51±6.95	26.22±5.55	25.41±6.64	28.11±4.55	24.86±7.91	27.03±6.51	21.35±7.20	23.51±5.68	20.81±8.64	22.43±4.84	23.51±4.53	20.54±6.86	21.35±5.73
MMODALITY	46.22	26.11	26.34	28.47	28.78	29.20	29.16	29.81	29.50	30.08	31.68	30.73	32.18	33.44
MNIST	87.59	75.86	48.42	4.88	2.97	2.62	2.43	3.76	2.69	2.28	2.36	2.11	1.84	1.82
MNIST-12000	87.13	78.03	59.17	8.27	3.84	3.15	2.82	5.15	3.30	2.71	2.87	2.55	2.35	2.57
MNIST-ROTATION	87.73	83.38	79.17	45.46	18.36	14.13	12.20	30.39	15.10	11.33	12.58	10.59	9.78	10.06
MNIST-BIMG	88.46	87.44	83.97	65.23	49.94	44.29	39.00	45.49	38.55	33.72	31.21	29.99	27.35	28.02
MNIST-BRAND	88.40	87.01	81.08	59.97	44.86	35.35	28.93	40.38	33.62	24.30	23.50	19.13	14.86	15.41
MNIST-BIMG-ROT	88.67	89.14	88.04	83.59	79.24	76.90	74.85	76.96	74.45	70.47	67.70	65.81	61.60	63.46
MSTAR-S1	63.23	51.85	45.48	42.80	41.43	41.15	40.07	39.16	40.24	38.99	38.59	36.71	32.04	41.61
NATURALSCENES	83.52±0.70	54.19±0.48	51.40±0.62	52.09±1.17	54.32±0.80	55.16±0.98	56.42±0.84	54.32±0.80	55.62±1.38	57.63±0.89	59.75±0.85	60.53±1.04	64.62±0.53	69.64±0.93
NORB-UNIFORM	73.89	54.77	29.64	19.03	16.12	13.06	10.86	14.51	14.03	9.50	10.86	8.07	5.24	5.27
NORB-JITTCLUTT	80.28	80.33	74.94	65.90	59.30	56.26	54.39	58.08	55.16	51.72	49.54	47.60	43.57	43.77
OLIVA	55.38±0.96	45.61±1.82	40.70±0.73	36.32±0.62	34.42±1.01	34.13±1.19	33.88±1.17	34.50±1.38	33.93±1.36	33.95±0.77	34.77±0.85	35.38±1.21	38.26±1.68	41.53±0.96
ORL	53.05±2.38	3.35±1.18	2.2±1.36	1.1±1.24	1.75±1.05	1.65±1.27	2.05±1.68	1.25±1.55	2.4±1.59	1.75±0.87	2.05±1.94	2.7±1.45	1.95±1.25	3.2±1.73
OUTEX	5.23±1.00	2.01±0.72	4.35±1.37	6.27±0.72	8.03±1.16	8.56±1.63	9.05±1.02	8.50±1.52	10.16±0.95	11.39±2.11	12.25±1.84	13.15±1.58	18.52±0.92	30.72±1.31
PFID	82.95±2.42	75.55±2.78	74.95±2.77	76.78±2.80	75.71±3.01	75.14±2.28	74.40±2.80	75.11±2.01	74.89±3.25	77.13±3.00	77.27±2.98	76.48±1.63	76.31±2.20	76.42±2.27
PPMI24	94.17	93.42	91.79	90.83	90.25	88.79	88.96	90.17	88.62	88.08	87.54	87.42	86.71	88.58
POLLEN	57.96±0.70	53.70±1.12	47.67±1.77	36.72±1.28	30.69±1.08	27.85±1.47	26.79±1.35	26.09±0.68	23.42±1.14	22.31±0.70	19.40±1.07	19.03±0.92	15.30±1.24	13.13±0.77
PUBFIG83	96.34±0.64	95.33 ± 0.80	93.16 ± 0.57	89.95 ± 0.53	87.11 ± 0.66	85.99 ± 0.58	85.10 ± 0.57	86.07 ± 1.11	84.23 ± 0.91	82.88 ± 0.68	82.06 ± 1.07	81.99 ± 1.48	82.01 ± 1.39	86.40 ± 1.41
RBC	60.36±1.22	51.44±1.79	40.85±2.29	36.22±1.59	35.70±2.20	35.25±1.53	35.38±1.28	31.98±1.38	33.93±1.63	34.38±1.76	31.27±1.12	33.61±1.07	34.32±0.75	37.71±1.25
RECT-BASIC	50.03	18.72	6.13	0.26	0.02	0.01	0.01	0.06	0.01	0.01	0.01	0.03	0.05	0.60
RECT-BIMG	49.86	44.94	38.09	21.50	17.59	17.56	17.48	16.25	15.78	16.58	16.00	17.07	18.47	21.19
RNAI	76.5±3.57	42.75±6.17	45.25±6.17	51.25±5.27	58.0±4.58	61.75±6.71	62.25±6.93	57.0±8.05	58.25±4.04	62.25±6.07	68.5±7.68	71.0±4.5	76.5±5.39	76.5±5.27
SEROUS	51.86±0.92		36.64±1.54	36.28±1.47	35.63±2.47	35.08±2.22	34.48±2.53	32.84±2.23	34.07±2.91	34.07±2.24	32.51±1.81	31.78±2.63	32.15±1.74	33.66±2.95
SHAPE1	59.0±5.76	52.0±4.82	39.6±4.84	29.63±2.78	22.6±2.73	22.9±3.75	21.5±3.96	21.9±3.94	22.2±4.02	19.3±2.57	18.5±2.77	19.5±4.10	15.7±4.56	17.3±3.1
SMEAR	15.05±2.60	12.75 ± 2.04	13.96 ± 2.69	12.64 ± 2.66	11.87 ± 1.95	12.09 ± 3.33	10.66 ± 2.14	10.99 ± 2.25	12.53 ± 2.75	10.44 ± 2.57	11.43 ± 1.92	10.88 ± 1.87	9.67 ± 2.24	10.77 ± 1.54
SOCCER	49.43±2.78	30.57±5.26	24.86±3.11	23.24±1.82	18.86±3.07	21.62±2.37	21.33±3.44	18.48±3.69	19.99±2.29	23.43±3.91	23.24±3.93	26.67±2.41	29.62±2.96	37.14±4.06
SPORTS	29.62±1.61	20.98±0.88	18.99±1.82	17.52±1.12	16.50±0.93	16.32±0.84	16.16±1.18	16.05±1.19	15.77±1.41	15.32±1.18	15.93±0.87	15.06±1.35	16.07±1.12	18.48±0.83
SOIL-24	27.08	1.67	0.0	0.0	0.83	5.0	20.83	0.0	0.42	11.25	3.33	19.58	42.92	62.5
SOIL-47	41.06	23.30	18.51	20.64	31.49	38.94	44.79	29.89	37.23	46.81	47.34	56.81	69.36	74.68
STOIC-101	73.94	61.97	58.94	56.97	55.45	53.18	52.42	51.21	50.45	48.18	45.15	46.36	44.70	46.97
STOMATA	45.45 ± 0.00	44.55 ± 2.73	39.09 ± 10.79	36.36 ± 10.76	37.27 ± 15.98	20.00 ± 11.35	21.82 ± 14.77	24.55 ± 11.54	21.82 ± 9.27	27.27 ± 12.86	16.36 ± 12.06	25.45 ± 11.35	23.64 ± 10.12	19.09 ± 11.82
STONEFLY9	46.34	34.32	30.14	30.49	27.96	26.57	25.26	26.22	25.96	24.39	24.30	22.91	20.03	19.51
SUBCELLULAR	65.81±3.42	37.67±5.53	35.58±3.90	23.95±4.60	22.21±3.98	21.40±4.54	19.88±5.21	20.47±4.39	19.77±3.75	21.63±3.61	20.47±3.33	19.07±3.53	19.19±3.87	20.70±3.95
SUN397	98.76±0.08	97.54±0.10	97.14±0.11	96.57±0.10	95.91±0.10	95.51±0.12	95.30±0.13	95.54±0.14	95.14±0.12	94.72±0.15	94.59±0.20	94.22±0.12	94.25±0.12	94.88±0.10
SWELEAF	63.88±2.03	25.93±1.29	16.92±1.42	12.25±1.25	10.55±1.30	8.72±0.95	7.36±0.77	10.71±1.49	9.40±0.89	7.24±0.57	7.84±1.13	6.95±1.07	6.37±0.80	8.19±1.22
TERMINALBULB	74.69±1.80	68.10±2.07	66.67±2.42	65.80±2.49	63.75±1.80	65.98±2.67	65.92±2.45	64.69±1.77	63.62±2.89	64.12±3.15	63.56±2.99	64.20±2.24	67.69±1.76	74.14±2.00
TINYGRAZ03	60.53±3.10	52.65±3.29	49.47±3.18	49.02±3.74	48.32±3.96	47.88±3.64	49.47±3.30	47.79±3.82	49.82±2.50	48.58±3.10	47.61±2.65	49.73±4.71	50.27±3.69	56.99±3.50
TSG-60	9.33±1.70	4.58±1.19	2.42±0.69	2.83±0.76	2.42±0.95	2.83±0.93	3.08±0.84	2.25±0.53	2.92±0.77	4.17±0.99	4.42±0.99	9.08±1.60	20.5±1.63	37.08±3.03
UIUCTEX	79.92±1.27	24.92±1.89	27.68±1.67	34.16±1.60	36.62±1.81	36.80±1.21	38.08±1.73	36.88±1.14	37.96±1.82	38.06±2.34	40.06±1.92	39.86±1.27	41.10±1.38	44.28±2.36
WANG	21.78±1.13	16.32±1.21	14.52±1.44	15.44±1.71	15.48±1.58	15.34±1.67	16.18±1.55	16.34±1.35	16.58±1.38	16.34±0.74	18.10±1.44	16.14±1.22	18.44±1.37	20.84±1.55
XEROX-6	26.94±3.60	17.65±3.35	17.55±1.98	17.96±3.60	14.69±2.10	16.22±3.83	17.35± 3.85	15.82±2.93	13.98±2.96	14.80±2.93	15.61±2.89	13.98±2.09	14.29±3.38	13.98±2.24
ZEBRATOXIC	50.00	54.17	54.17	41.67	15.62	7.29	6.25	10.42	4.17	7.29	7.29	7.29	10.42	12.50
ZUBUD	26.09	16.52	11.30	6.09	4.35	3.48	4.35	3.48	3.48	3.48	3.48	3.48	4.35	9.57

Datasets	HSV-8 × 8	HSV-16 × 16	HSV-32 × 32	HSV-ROTATE360-16 × 16	GRAY-16 × 16
ACROSOMES	4.49±0.75	4.02 ± 0.31	4.81 ± 1.21	6.65 ± 1.14	4.02 ± 0.31
AGEMAP-FAGING	43.71±4.39	47.23±3.02	49.52 ± 5.15	51.05±4.51	55.62±2.73
ALL-IDB2	1.15±1.76	0.96±1.29	1.73±1.35	1.54±1.68	5.77±2.72
ALOI	22.26	22.57	21.41	39.48	39.56
ALOT	31.70±0.30	31.64±0.36	31.29±0.32	41.41±0.23	58.43±0.50
APOPTOSIS	9.14 ± 2.94	8.14 ± 2.86	9.71 ± 2.37	10.29 ± 3.25	8.14 ± 2.86
BINUCLEATE	12.25±4.10	4.36±5.26	6.25 ± 3.75	11.50 ± 8.15	4.36±5.26
BIRDS	48.23±1.83	44.97±3.43	47.03±1.85	48.8±1.15	45.4±2.07
BREAST-CANCER	15.42±5.18	13.33±4.49	15.69±3.29	15.42±3.48	34.03±6.37
BUILDINGSAB	42.98±2.79	42.82±2.07	43.70±2.51	51.16±2.27	42.82±2.07
BUTTERFLIES	51.53±2.60	49.86±2.39	50.05±2.54	46.00±3.46	45.42±2.77
BUTTERFLIES-CLEAN	17.76±1.56	17.80±1.41	16.07±1.43	20.53±1.77	16.56±1.51
CALTECH-101	71.48	70.61	71.07	77.56	72.09
CALTECH-256	88.34±0.29	88.26±0.26	88.25±0.21	90.44±0.21	89.79±0.37
C.ELEGANS	39.40±4.91	38.63±4.97	37.95±4.67	39.03±5.92	38.63±4.97
C.ELEGANS-LIVE-DEAD	4.74±2.83	3.16 ± 3.49	4.74±5.49	11.05±7.61	3.16 ± 3.49
CHARS74K	71.58±1.61	71.16±1.66	71.28±1.23	79.59±1.64	71.16±1.66
CHO	3.99±2.77	2.62±1.83	3.23±1.88	2.46±2.09	2.62±1.83
CIFAR-10	50.08	48.91	49.05	55.31	55.24
COIL-100 nv1	12.15	12.15	12.20	13.89	35.75
CONVEX	9.36	8.04	8.00	9.87	8.04
ETH-80	17.35±4.73	16.83±5.17	19.70±7.35	21.19±8.50	18.96±6.33
EVENTS	35.17±2.10	36.04±1.28	35.31±1.36	46.98±2.71	49.42±1.81
FAMOUSLAND	67.45±2.13	66.85±3.55	69.65±2.06	76.65±3.02	72.95±2.11
FLOWERS17	34.41±2.22	33.24±1.16	34.47±1.95	36.12±2.82	57.74±1.54
FLOWERS102-CLEAN	56.68	57.19	56.76	62.39	73.24
GALAXYZOO	5.85±0.18	4.22±0.17	3.45±0.09	5.79±0.18	5.25±0.15
GTSRBCROP	36.52	29.01	28.31	46.37	21.01
HCC	35.00±9.81	33.21 ± 14.86	40.36±12.01	43.57±10.81	69.29±9.61
HEP2-ICPR2012	63.49	62.67	62.13	63.76	62.67
HPA	7.24 ± 1.55	6.19 ± 1.30	5.90 ± 2.29	9.90±1.91	11.52±1.78
INDOOR	85.81±0.66	85.09±0.90	85.38±0.70	88.23±0.65	88.98±0.68
IRMA2005-BALANCED	15.0	14.9	14.40	31.40	14.9
IRMA2006-BALANCED	27.7	26.40	26.10	43.60	26.40
KTH-TIPS	23.88±2.52	22.61±1.03	21.34±1.80	26.27±1.61	22.61±1.03
KTH-TIPS2	45.78±10.50	42.77±8.41	50.39±6.84	47.16±8.61	54.99±4.60
LANDUSE	27.21±1.08	29.21±2.57	28.07±1.91	40.55 ± 1.81	41.67±1.55
LYMPHOMA	23.07±4.38	20.54±6.86	22.13±5.73	24.0±5.40	42.40±5.19
MMODALITY	27.48	26.11	26.64	29.05	35.99
MNIST	2.39	1.82	1.69	4.53	1.82
MNIST-12000	2.66	2.35	2.30	6.64	2.35
MNIST-ROTATION	10.43	9.78	9.48	8.15	9.78
MNIST-BIMG	32.84	27.35	23.95	46.54	27.35
MNIST-BRAND	20.38	14.86	12.70	30.06	14.86
MNIST-BIMG-ROT	66.29	61.60	58.29	56.29	61.60
MSTAR-S1	34.66	32.04	29.99	38.70	32.04
NATURALSCENES	50.60±0.97	51.40±0.62	50.68±0.86	65.83±0.93	51.40±0.62
NORB-UNIFORM	5.65	5.24	5.64	13.31	5.24
NORB-JITTCLUTT	48.80	43.57	40.61	51.72	43.57
OLIVA	34.13±1.03	33.88±1.17	33.60±0.56	44.64±0.91	39.09±1.58
ORL	1.2±1.27	1.1±1.24	1.4±1.04	5.6±1.71	1.1±1.24
OUTEX	2.80±0.70	2.01±0.72	2.08±0.39	3.94±0.65	20.39±0.82
PFID	75.41±3.73	74.40±2.80	75.00±2.60	80.74±3.54	80.36±2.14
POLLEN	15.99±1.07	13.13±0.77	11.75±0.84	14.43±1.16	13.13±0.77
PPMI24	86.38	86.71	86.25	90.38	91.33
PUBFIG83	82.67±0.48	81.99 ± 1.48	81.93±1.32	91.59±1.02	84.82±0.98
RBC	33.31±1.48	31.27±1.12	29.93±1.83	31.66±1.30	31.27±1.12
RECT-BASIC	0.01	0.01	0.01	6.55	0.01
RECT-BIMG	15.74	15.78	15.64	38.60	15.78
RNAI	47.5±4.33	42.75±6.17	45.75±7.99	47.25±7.02	42.75±6.17
SEROUS	35.38±1.49	31.78±2.63	32.54±2.14	30.85±1.96	41.53±1.98
SHAPE1	17.2±3.31	15.7±4.56	16.7±4.20	25.8±4.28	10.7±2.79
SMEAR	13.41±3.64	9.67 ± 2.24	9.67 ± 2.07	9.45±3.58	14.18±2.8
SOCCER	19.24±3.04	18.48±3.69	18.76±3.13	23.14±3.07	51.05±4.82
SOIL-24	0.83	0.0	0.0	7.5	43.33
SOIL-47	18.94	18.51	17.23	29.26	61.06
SPORTS	15.44±1.57	15.06±1.35	14.90±1.61	19.30±1.17	35.18±1.56
STOIC101	45.61	44.70	44.70	53.79	59.55
STOMATA	25.45 ± 12.06	16.36 ± 12.06	25.45 ± 9.79	12.73 ± 10.91	16.36 ± 12.06
STONEFLY9	22.13	19.51	19.08	28.66	25.09
SUBCELLULAR	20.58±2.90	19.07±3.53	20.35±5.36	17.09±2.50	19.07±3.53
SUN397	94.37±0.16	94.22±0.12	94.20±0.17	95.90±0.13	n/a
SWELEAF	7.03±0.43	6.37±0.80	7.24±0.69	12.36±1.31	6.37±0.80
TERMINALBULB	65.87±2.33	63.56±2.99	63.37±1.46	66.09±2.79	63.56±2.99
TINYGRAZ03	51.06±4.90	47.61±2.65	47.96±3.99	52.74±3.74	66.46±4.03
TSG-60	2.58±0.45	2.25±0.53	2.33±0.73	3.49±1.22	13.00±1.50
UIUCTEX	26.34±2.43	24.92±1.89	25.02±1.42	29.50±1.15	24.92±1.89
WANG	15.36±1.21	14.52±1.44	15.22±1.44	17.10±1.05	24.12±2.08
XEROX-6	16.12±2.77	13.98±2.96	14.80±2.97	20.81±3.42	28.16±4.64
ZEBRATOXIC	5.21	4.17	4.17	4.17	6.25
ZUBUD	4.35	3.48	4.35	11.30	11.30

TABLE III
EVALUATION OF THE SUBWINDOW DESCRIPTORS WITH ET-DIC. OTHER PARAMETERS ARE CONSTANT: $n_{min} = 1, T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION WITH BEST SIZE INTERVALS.

Datasets	1	$\sqrt{M}$	$M/8$	$M/4$	$M/2$	M
ACROSOMES	6.23 ± 1.13	4.02 ± 0.31	4.29 ± 1.27	4.18 ± 1.05	4.40 ± 0.55	4.50 ± 0.99
AGEMAP-FAGING	58.10±3.74	47.23±3.02	45.52±3.76	42.10±3.85	42.29±5.01	44.00±3.48
ALL-IDB2	2.31±2.24	0.96±1.29	1.73±1.60	1.35±1.93	1.92±1.72	2.12±1.35
ALOI	25.58	22.57	20.03	18.86	17.54	16.16
ALOT	32.85±0.46	31.64±0.36	29.96±0.36	29.41±0.24	28.42±0.28	28.81±0.35
APOPTOSIS	9.86 ± 3.41	8.14 ± 2.86	8.86 ± 3.25	8.71 ± 2.51	9.71 ± 2.91	7.86 ± 3.51
BINUCLEATE	12.50±8.80	10.00±9.55	4.36±5.26	4.75 ± 6.93	8.50 ± 6.82	7.50 ± 5.24
BIRDS	51.33±2.29	44.97±3.43	45.30±2.77	43.1±1.95	42.73±3.11	42.37±1.86
BREAST-CANCER	18.88±3.47	13.33±4.49	15.0±4.02	14.58±3.47	15.14±4.71	14.86±5.34
BUILDINGSAB	42.54±2.15	42.82±2.07	43.37±2.31	42.65±1.76	40.88±1.95	43.20±2.32
BUTTERFLIES	60.02±2.03	49.86±2.39	45.56±1.65	44.49±3.24	41.28±3.14	39.91±3.18
BUTTERFLIES-CLEAN	24.93±2.61	17.80±1.41	16.31±2.12	15.89±1.40	17.02±1.83	16.58±2.29
CALTECH-101	72.07	70.61	70.82	70.77	71.07	71.73
CALTECH-256	88.32±0.29	88.26±0.26	88.18±0.43	88.07±0.29	88.34±0.28	88.12±0.37
C.ELEGANS	44.27±5.55	38.63±4.97	39.66±5.55	38.29±4.32	38.12±4.36	38.72±4.73
C.ELEGANS-LIVE-DEAD	2.11 ± 3.49	3.16 ± 3.49	5.79 ± 4.97	6.84 ± 5.29	5.79 ± 4.37	5.26 ± 5.26
CHARS74K	73.01±1.22	71.16±1.66	71.17±1.26	71.65±1.02	71.51±1.30	71.40±1.25
CHO	4.92±2.26	2.62±1.83	3.08±1.82	3.54±1.55	3.23±2.22	3.38±2.04
CIFAR-10	53.50	48.91	47.84	47.24	47.00	46.33
COIL-100 nv1	13.13	12.15	12.37	12.25	12.20	12.75
CONVEX	11.35	8.04	7.92	7.89	7.79	7.87
ETH-80	19.42±5.16	16.83±5.17	14.94±6.35	14.97±6.20	19.86±9.47	16.20±5.93
EVENTS	37.69±1.95	36.04±1.28	34.42±1.56	33.02±1.71	34.96±1.70	35.06±1.89
FAMOUSLAND	64.6±3.35	66.85±3.55	70.05±2.35	70.65±3.32	72.65±1.61	73.4±4.27
FLOWERS17	40.24±1.97	33.24±1.16	33.44±1.75	31.82±1.93	32.12±2.11	30.91±1.97
FLOWERS102-CLEAN	59.16	57.19	56.20	55.81	55.36	55.65
GALAXYZOO	15.87±0.55	4.22±0.17	3.20±0.11	2.94±0.13	2.73±0.07	2.64±0.13
GTSRBCROP	47.30	29.01	23.59	21.15	19.50	18.73
HCC	48.93 ± 15.20	33.21 ± 14.86	35.71 ± 10.83	30.71 ± 15.17	50.00±11.52	41.56±8.78
HEP2-ICPR2012	64.71	62.67	62.53	61.85	62.40	61.85
HPA	11.71 ± 2.00	6.19 ± 1.30	5.24 ± 0.88	6.29±2.45	5.90±1.75	5.03±2.14
INDOOR	85.49±1.29	85.09±0.90	85.19±0.51	84.78±0.57	84.54±0.91	84.94±0.69
IRMA2005-BALANCED	17.00	14.90	13.90	14.20	14.40	14.20
IRMA2006-BALANCED	29.60	26.40	26.60	25.40	25.30	25.80
KTH-TIPS	26.80±1.94	22.61±1.03	23.46±1.96	20.41±1.66	20.95±1.78	21.37±2.02
KTH-TIPS2	48.89±10.96	42.77±8.41	47.77±7.28	47.86±8.27	48.83±8.99	45.52±8.39
LANDUSE	32.43 ± 1.57	29.21±2.57	27.07 ± 2.67	26.10 ± 1.96	25.21±1.67	25.33 ± 2.42
LYMPHOMA	32.53±3.78	20.54±6.86	21.47±3.40	22.67±3.48	16.53±3.22	17.99±5.54
MMODALITY	28.70	26.11	31.68	26.64	25.88	25.50
MNIST	2.95	1.82	1.65	1.54	1.52	1.63
MNIST-12000	3.81	2.35	2.26	2.26	2.20	2.25
MNIST-ROTATION	12.72	9.78	9.53	9.20	9.08	8.94
MNIST-BIMG	54.90	27.35	23.92	21.27	19.79	18.61
MNIST-BRAND	39.39	14.86	12.64	11.33	10.59	10.05
MNIST-BIMG-ROT	78.66	61.60	58.65	55.45	52.55	50.66
MSTAR-S1	37.28	32.04	29.77	26.98	23.96	21.63
NATURALSCENES	53.77±1.23	51.40±0.62	50.69±0.86	49.49±0.88	48.99±0.74	49.00±0.75
NORB-UNIFORM	8.87	5.24	5.46	6.33	6.65	7.08
NORB-JITTCLUTT	61.28	43.57	40.54	38.68	37.22	36.14
OLIVA	37.67±1.36	33.88±1.17	33.64±0.67	33.56±0.89	32.50±1.48	31.93±1.21
ORL	1.25±0.90	1.1±1.24	1.05±1.23	0.95±0.79	1.75±0.78	1.2±0.81
OUTEX	2.85±0.88	2.01±0.72	2.31±0.85	1.90±0.85	1.67±0.63	1.74±0.78
PFID	76.53±3.35	74.40±2.80	75.22±3.29	74.10±3.37	76.31±2.54	75.60±3.97
POLLEN	20.47±1.46	13.13±0.77	11.75±1.10	10.47±0.79	9.54±0.77	9.51±0.80
PPMI24	88.79	86.71	85.46	85.83	85.38	85.04
PUBFIG83	84.86±1.26	81.99 ± 1.48	80.86±1.35	80.63±0.96	80.49±1.25	79.58±1.56
RBC	34.74±1.37	31.27±1.12	29.96±1.61	30.20±1.63	34.47±1.01	29.86±1.66
RECT-BASIC	0.01	0.01	0.01	0.01	0.01	0.01
RECT-BIMG	17.80	15.78	15.67	15.55	15.48	15.47
RNAI	45.0±6.61	42.75±6.17	43.5±4.89	40.75±3.54	44.0±6.82	40.25±7.37
SEROUS	37.32±2.69	31.78±2.63	31.42±2.21	32.92±1.32	31.20±1.67	31.23±2.45
SHAPE1	18.3±4.15	15.7±4.56	17.7±5.16	13.9±2.74	14.6±2.97	15.9±4.18
SMEAR	14.62 ± 3.37	9.67 ± 2.24	11.87 ± 3.10	9.23 ± 3.12	10.00 ± 1.59	8.48±2.01
SOCCER	24.95	18.48±3.69	17.33±2.20	17.33±2.17	16.76±2.93	18.48±3.08
SOIL-24	6.67	0.0	0.0	0.0	0.42	0.0
SOIL-47	28.72	18.51	17.13	16.38	16.60	15.85
SPORTS	24.75±1.19	15.06±1.35	14.02±1.62	14.80±0.74	12.79±1.29	13.07±1.24
STOIC101	49.39	44.70	42.73	42.12	41.36	39.39
STOMATA	40.00 ± 8.33	16.36 ± 12.06	20.00 ± 7.93	26.36 ± 10.33	21.82 ± 13.61	20.00 ± 6.80
STONEFLY9	31.10	19.51	18.55	18.03	18.38	16.99
SUBCELLULAR	20.93±3.08	19.07±3.53	19.42±2.33	19.53±3.28	16.63±4.19	18.14±3.61
SUN397	94.76±0.15	94.22±0.12	94.11±0.15	94.09±0.09	93.86±0.14	93.78±0.20
SWELEAF	6.69±0.96	6.37±0.80	7.08±0.64	6.95±1.20	6.28±0.96	6.71±0.94
TERMINALBULB	64.99±1.48	63.56±2.99	64.25±1.90	63.85±2.84	62.49±2.62	61.82±1.56
TINYGRAZ03	52.12±3.44	47.61±2.65	48.05±3.43	47.08±3.03	46.99±3.10	46.37±5.67
TSG-60	2.42±0.45	2.25±0.53	2.58±0.69	2.50±0.91	2.67±0.73	2.08±0.77
UIUCTEX	27.5±2.05	24.92±1.89	25.04±1.03	24.62±2.09	24.80±0.95	23.04±1.99
WANG	15.82±0.77	14.52±1.44	15.06±1.87	14.58±1.36	14.60±1.34	14.58±0.99
XEROX-6	16.53±3.76	13.98±2.96	15.20±3.14	13.88±3.17	13.78±3.14	13.57±2.66
ZEBRATOXIC	5.21	4.17	4.17	4.17	4.17	3.12
ZUBUD	4.35	3.48	3.48	3.48	3.48	3.48

TABLE IV

EVALUATION OF THE NUMBER OF RANDOM TESTS IN A TREE NODE (K) WHERE M IS THE TOTAL NUMBER OF ATTRIBUTES ($M = 256$ IN GRAYSCALE IMAGES, 768 IN COLOR IMAGES WITH ET-DIC). OTHER PARAMETERS ARE CONSTANT: $n_{min} = 1$, $T = 10$, $N_{ls} = 1$ MILLION WITH BEST SIZE INTERVALS AND 16×16 DESCRIPTORS.

Datasets	1	5	10	20
ACROSOMES	5.14 ± 0.85	4.23 ± 0.65	4.02 ± 0.31	4.40 ± 1.16
AGEMAP-FAGING	47.05±3.49	46.10±3.39	47.23±3.02	49.52±3.13
ALL-IDB2	1.73±1.35	2.12±1.60	0.96±1.29	1.15±1.76
ALOI	36.89	24.41	22.57	21.09
ALOT	40.58±0.29	32.98±0.41	31.64±0.36	30.98±0.44
APOPTOSIS	12.14 ± 4.48	9.71 ± 2.91	8.14 ± 2.86	7.43 ± 3.37
BINUCLEATE	12.5±8.80	10.00±9.55	4.36±5.26	7.00±8.93
BIRDS	48.43±2.70	46.97±3.16	44.97±3.43	47.67±1.53
BREAST-CANCER	19.31±3.19	16.25±2.78	13.33±4.49	16.53±3.01
BUILDINGSAB	43.15±2.07	42.32±3.05	42.82±2.07	41.10±2.50
BUTTERFLIES	53.27±2.57	52.43±3.24	49.86±2.39	50.66±2.47
BUTTERFLIES-CLEAN	19.60±1.61	17.29±1.86	17.80±1.41	18.47±3.56
CALTECH-101	76.92	72.68	70.61	71.01
CALTECH-256	93.17±0.20	89.32±0.27	88.26±0.26	87.45±0.32
C.ELEGANS	37.86±4.73	40.17±4.54	38.63±4.97	37.86±3.40
C.ELEGANS-LIVE-DEAD	3.68 ± 4.11	7.37 ± 4.21	3.16 ± 3.49	7.37 ± 5.86
CHARS74K	75.94±1.81	72.62±1.66	71.16±1.66	71.09±1.03
CHO	4.00±1.71	3.85±1.85	2.62±1.83	3.38±1.51
CIFAR-10	62.04	50.66	48.91	47.77
COIL-100 nv1	13.32	12.21	12.15	12.0
CONVEX	8.79	8.10	8.04	8.01
ETH-80	14.82±4.61	16.34±6.87	16.83±5.17	16.19±5.96
EVENTS	36.25±1.82	36.88±1.42	36.04±1.28	35.72±2.69
FAMOUSLAND	72.75±2.40	67.55±2.69	66.85±3.55	67.95±2.07
FLOWERS17	35.47±2.21	35.29±1.13	33.24±1.16	32.29±2.78
FLOWERS102-CLEAN	59.37	57.10	57.19	56.93
GALAXYZOO	5.92±0.28	4.42±0.15	4.22±0.17	4.03±0.12
GTSRBCROP	39.81	31.11	29.01	28.14
HCC	38.93 ± 11.17	42.50 ± 9.17	33.21 ± 14.86	36.07 ± 11.40
HEP2-ICPR2012	63.35	63.08	62.67	62.67
HPA	7.62 ± 1.70	7.05 ± 3.11	6.19 ± 1.30	6.57 ± 2.19
INDOOR	88.75±0.57	86.19±0.52	85.09±0.90	84.51±0.88
IRMA2005-BALANCED	18.50	15.60	14.90	14.60
IRMA2006-BALANCED	31.50	27.0	26.40	26.80
KTH-TIPS	23.51±1.80	25.05±2.62	22.61±1.03	23.15±2.50
KTH-TIPS2	47.46±9.33	49.44±6.23	42.77±8.41	50.60±6.63
LANDUSE	31.12±2.94	29.14±1.62	29.21±2.57	28.79±1.83
LYMPHOMA	23.6±3.68	22.27±3.91	20.54±6.86	23.6±3.04
MMODALITY	29.24	26.49	26.11	26.49
MNIST	2.66	1.96	1.82	1.72
MNIST-12000	3.18	2.51	2.35	2.37
MNIST-ROTATION	13.20	10.24	9.78	9.55
MNIST-BIMG	40.27	29.80	27.35	26.02
MNIST-BRAND	29.73	16.72	14.86	13.79
MNIST-BIMG-ROT	73.94	65.21	61.60	59.84
MSTAR-S1	35.23	32.95	32.04	31.53
NATURALSCENES	57.31±0.62	51.45±0.81	51.40±0.62	50.64±0.87
NORB-UNIFORM	10.66	5.67	5.24	5.13
NORB-JITTCLUTT	52.55	44.85	43.57	42.77
OLIVA	36.00±0.71	33.47±0.89	33.88±1.17	33.25±1.31
ORL	1.25±1.19	1.85±0.78	1.1±1.24	1.65±0.74
OUTEX	1.92±1.09	2.27±0.64	2.01±0.72	2.43±0.97
PFID	76.69±3.92	74.95±5.88	74.40±2.80	75.98±3.59
POLLEN	14.46±1.75	13.42±0.42	13.13±0.77	14.13±0.77
PPMI24	90.12	88.12	86.71	86.04
PUBFIG83	91.59±0.70	84.33±1.10	81.99 ± 1.48	80.41±0.88
RBC	32.41±1.99	32.39±0.58	31.27±1.12	31.74±1.18
RECT-BASIC	0.02	0.01	0.01	0.01
RECT-BIMG	19.76	16.25	15.78	15.54
RNAI	49.25±6.99	46.0±8.67	42.75±6.17	45.25±6.75
SEROUS	34.95±1.73	33.50±1.53	31.78±2.63	33.93±1.78
SHAPE1	20.8±4.70	17.0±2.72	15.7±4.56	15.3±3.52
SMEAR	11.21 ± 3.18	10.11 ± 2.90	9.67 ± 2.24	9.78 ± 1.73
SOCCER	20.76±2.12	20.38±2.26	18.48±3.69	20.00±3.43
SOIL-24	9.58	0.42	0.0	0.0
SOIL-47	33.51	21.81	18.51	18.19
SPORTS	16.90±1.40	15.67±1.01	15.06±1.35	15.07±1.83
STOIC101	50.0	44.70	44.70	45.15
STOMATA	26.36 ± 12.50	26.36 ± 13.76	16.36 ± 12.06	23.64 ± 7.27
STONEFLY9	22.74	21.08	19.51	20.12
SUBCELLULAR	18.49±2.67	18.84±3.99	19.07±3.53	20.12±4.23
SUN397	97.35±0.08	95.07±0.08	94.22±0.12	93.54±0.13
SWELEAF	6.87±0.55	6.71±1.38	6.37±0.80	6.87±0.97
TERMINALBULB	65.05±2.80	64.92±1.11	63.56±2.99	64.94±1.92
TINYGRAZ03	52.65±4.12	51.50±3.64	47.61±2.65	49.82±5.13
TSG-60	2.75±1.06	2.17±0.67	2.25±0.53	1.67±0.0
UIUCTEX	33.82±1.72	27.80±1.20	24.92±1.89	24.92±1.77
WANG	15.02±1.12	14.92±1.46	14.52±1.44	16.36±0.68
XEROX-6	16.53±4.25	15.20±3.01	13.98±2.96	16.02±1.94
ZEBRATOXIC	4.17	4.17	4.17	4.17
ZUBUD	4.35	4.35	3.48	4.35

TABLE V

EVALUATION OF THE NUMBER OF TREES (T) WITH ET-DIC. OTHER PARAMETERS ARE CONSTANT: $n_{min} = 1$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION WITH BEST SIZE INTERVALS AND 16×16 DESCRIPTORS.

Datasets	2	5	10	50	100	500	1000
ACROSOMES	4.02 ± 0.31	4.13 ± 0.72	4.63 ± 1.16	6.16 ± 0.77	6.63 ± 1.23	9.57 ± 2.22	11.66 ± 1.42
AGEMAP-FAGING	47.23±3.02	47.81±4.01	49.24±3.22	52.29±3.26	59.24±4.01	60.19±2.51	61.52±3.49
ALL-IDB2	0.96±1.29	2.12±2.01	1.35±1.73	2.31±1.15	2.69±2.14	2.88±3.47	1.73±1.35
ALOI	22.57	22.14	22.21	24.90	27.02	34.40	39.22
ALOT	31.64±0.36	31.90±0.41	32.48±0.40	37.38±0.30	40.14±0.42	48.34±0.36	52.54±0.28
APOPTOSIS	8.14 ± 2.86	9.14 ± 3.57	9.14 ± 3.51	9.29 ± 3.95	10.00 ± 2.47	9.86 ± 4.26	8.43 ± 2.16
BINUCLEATE	4.36±5.26	6.50 ± 6.14	9.25 ± 6.43	8.25 ± 7.83	10.25 ± 6.17	12.50 ± 6.80	14.00 ± 6.24
BIRDS	44.97±3.43	47.93±3.34	48.13±2.06	47.00±2.29	47.40±2.61	47.87±3.63	48.53±2.98
BREAST-CANCER	13.33±4.49	16.53±2.44	15.97±6.34	15.28±3.51	18.33±3.22	18.19±3.59	16.39±4.72
BUILDINGSAB	42.82±2.07	43.15±2.78	43.81±2.20	45.52±1.88	45.52±2.06	50.94±2.75	53.98±2.82
BUTTERFLIES	49.86±2.39	48.15±3.36	52.11±2.25	50.07±3.96	50.98±3.14	49.89±2.80	50.37±2.21
BUTTERFLIES-CLEAN	17.80±1.41	19.89±2.14	21.22±2.97	21.78±2.08	23.58±1.75	27.18±2.35	28.89±2.22
CALTECH-101	70.61	71.11	71.16	72.31	72.29	74.28	76.96
CALTECH-256	88.26±0.26	88.13±0.34	87.92±0.38	88.34±0.21	88.87±0.39	90.15±0.28	90.57±0.21
C.ELEGANS	38.63±4.97	35.47±4.10	38.20±6.89	41.79±5.33	39.91±5.51	46.15±4.77	45.47±4.72
C.ELEGANS-LIVE-DEAD	3.16 ± 3.49	3.16 ± 4.21	4.74 ± 4.97	6.84 ± 4.11	5.26 ± 4.08	6.84 ± 5.79	5.79 ± 9.25
CHARS74K	71.16±1.66	71.28±1.32	71.08±1.25	71.70±1.18	70.89±1.37	73.82±1.46	74.94±1.77
CHO	2.62±1.83	4.15±3.44	3.38±2.04	4.62±2.57	4.77±3.73	7.69±2.18	7.23±3.44
CIFAR-10	48.91	48.79	48.97	50.68	51.26	54.28	56.19
COIL-100 nv1	12.15	12.55	12.47	13.23	13.48	16.30	18.30
CONVEX	8.04	7.96	7.81	8.20	8.46	9.94	10.81
ETH-80	16.83±5.17	16.70±8.29	16.04±6.31	19.12±6.98	22.13±4.51	21.65±7.59	23.63±5.00
EVENTS	36.04±1.28	35.94±2.49	36.65±2.29	36.17±1.62	37.54±1.45	39.63±1.04	39.40±1.46
FAMOUSLAND	66.85±3.55	68.0±2.28	68.6±2.08	71.4±2.60	73.65±2.50	74.7±2.87	76.6±1.07
FLOWERS17	33.24±1.16	32.82±1.90	34.03±1.76	35.62±1.59	36.76±2.45	36.59±1.69	40.35±2.28
FLOWERS102-CLEAN	57.19	57.09	58.01	60.31	62.62	66.92	68.68
GALAXYZOO	4.22±0.17	4.19±0.11	4.29±0.15	4.65±0.09	5.11±0.13	6.69±0.21	7.62±0.18
GTSRBCROP	29.01	29.45	28.67	30.79	32.23	36.47	38.23
HCC	33.21 ± 14.86	40.00 ± 8.27	42.86 ± 8.75	42.14 ± 11.04	46.79 ± 12.47	47.50 ± 13.03	54.64 ± 13.03
HEP2-ICPR2012	62.67	62.81	62.81	63.49	63.76	64.71	64.85
HPA	6.19 ± 1.30	6.76 ± 2.02	5.81 ± 2.46	7.81 ± 2.62	10.38 ± 3.03	9.81 ± 2.00	12.57 ± 2.51
INDOOR	85.09±0.90	85.07±0.86	84.95±0.73	85.89±1.01	86.71±0.69	87.99±0.73	88.46±0.74
IRMA2005-BALANCED	14.90	15.40	16.30	18.60	20.70	25.70	28.60
IRMA2006-BALANCED	26.40	26.70	27.10	30.20	33.00	39.10	43.70
KTH-TIPS	22.61±1.03	22.89±1.88	24.32±2.58	29.46±2.27	30.12±2.03	34.78±1.88	36.88±2.05
KTH-TIPS2	42.77±8.41	50.77±11.91	45.34±10.01	48.27±9.57	47.63±9.26	50.86±9.36	51.15±8.99
LANDUSE	29.21±2.57	29.21±2.30	30.43 ±1.18	33.38±2.25	37.31±1.72	43.37±2.13	49.95±2.05
LYMPHOMA	20.54±6.86	22.53±4.44	25.60±5.16	26.93±3.26	28.80±5.07	32.80±5.41	35.99±6.98
MMODALITY	26.11	26.41	26.98	29.05	29.39	32.02	33.85
MNIST	1.82	1.66	1.92	2.37	2.61	3.65	4.23
MNIST-12000	2.35	2.42	2.44	2.99	3.48	4.91	5.81
MNIST-ROTATION	9.78	9.87	9.84	11.37	12.56	16.67	19.48
MNIST-BIMG	27.35	26.93	26.74	26.37	27.32	30.74	32.56
MNIST-BRAND	14.86	14.47	14.24	14.44	16.26	18.64	21.16
MNIST-BIMG-ROT	61.60	61.31	60.64	60.30	60.89	64.44	66.48
MSTAR-S1	32.04	31.93	30.90	33.52	35.52	37.91	40.07
NATURALSCENES	51.40±0.62	50.68±1.00	52.00±0.77	54.89±0.71	56.40±0.93	59.49±1.00	62.18±1.18
NORB-UNIFORM	5.24	5.63	5.67	7.12	8.15	13.51	15.45
NORB-JITTCLUTT	43.57	43.30	43.49	44.99	45.92	49.21	50.88
OLIVA	33.88±1.17	33.94±1.43	34.58±0.58	35.83±1.13	36.21±0.94	38.98±1.19	39.85±1.31
ORL	1.1±1.24	1.6±1.3	1.7±1.36	2.65±1.45	2.65±1.83	4.9±2.0	9.65±2.36
OUTEX	2.01±0.72	2.34±1.11	2.55±0.95	4.21±0.76	4.98±0.67	9.31±0.90	12.20±0.82
PFID	74.40±2.80	76.04±3.05	76.97±2.38	77.79±2.27	78.63±2.46	80.52	81.39±1.79
POLLEN	13.13±0.77	12.79±1.01	13.86±1.14	15.39±0.77	15.88±1.04	19.40±1.00	22.84±1.19
PPMI24	86.71	86.21	85.88	86.62	86.33	85.62	85.96
PUBFIG83	81.99 ± 1.48	80.78±1.36	81.48±1.28	82.77±0.52	83.05±1.27	84.93±1.10	85.82±0.89
RBC	31.27±1.12	36.19±1.26	37.31±1.30	41.08±1.59	43.03±1.50	45.62±2.33	46.99±0.92
RECT-BASIC	0.01	0.01	0.01	0.01	0.01	0.04	0.11
RECT-BIMG	15.78	15.84	16.04	17.10	17.65	19.07	19.65
RNAI	42.75±6.17	47.5±6.32	47.25±6.84	47.25±4.93	48.25±5.48	50.25±4.25	56.25±6.64
SEROUS	31.78±2.63	32.79±2.81	32.68±2.35	33.47±2.05	33.14±1.75	37.57±2.53	39.75±2.82
SHAPE1	15.7±4.56	16.8±2.82	14.5±3.14	16.2±2.75	16.6±2.29	16.8±3.28	15.7±1.79
SMEAR	9.67 ± 2.24	10.11 ± 2.59	11.32 ± 3.51	11.98 ± 4.07	9.89 ± 1.10	11.32 ± 2.74	13.52 ± 2.69
SOCCER	18.48±3.69	19.14±2.64	20.19±3.56	20.38±2.42	20.76±3.15	22.67±2.88	24.67±2.27
SOIL-24	0.0	0.0	0.0	0.0	0.0	4.58	7.5
SOIL-47	18.51	17.55	18.09	20.64	23.09	29.47	34.47
SPORTS	15.06±1.35	16.36±1.30	16.03±1.29	17.60±1.00	18.27±1.09	20.00±1.46	21.08±1.06
STOIC101	44.70	44.24	45.91	46.52	55.00	57.73	67.12
STOMATA	16.36 ± 12.06	20.00 ± 7.93	20.00 ± 9.79	21.82 ± 13.61	22.73 ± 14.80	25.45 ± 8.91	29.09 ± 11.35
STONEFLY9	19.51	19.34	19.95	20.56	21.08	27.00	27.35
SUBCELLULAR	19.07±3.53	18.95±3.08	19.07±4.45	21.16±3.32	21.86±3.24	25.93±3.37	30.0±3.23
SUN397	94.22±0.12	93.85±0.18	93.54±0.18	93.71±0.14	93.97±0.16	94.63±0.16	94.99±0.15
SWELEAF	6.37±0.80	7.39±0.76	7.37±0.51	8.01±0.57	7.37±0.56	9.73±1.07	10.95±0.54
TERMINALBULB	63.56±2.99	65.89±2.61	63.85±2.37	64.98±2.10	64.67±2.66	66.70±1.79	66.37±3.17
TINYGRAZ03	47.61±2.65	51.15±4.66	52.12±2.04	48.05±2.91	50.00±4.80	52.74±5.07	55.49±2.99
TSG-60	2.25±0.53	2.25±0.53	2.00±0.85	3.17±0.82	2.83±0.93	3.67±1.07	6.50±1.74
UIUCTEX	24.92±1.89	25.12±1.43	27.34±2.27	31.40±0.98	32.6±2.50	39.10±1.57	41.50±1.05
WANG	14.52±1.44	15.48±1.26	14.66±1.77	16.36±1.82	16.36±0.94	17.82±1.31	18.4±1.50
XEROX-6	13.98±2.96	15.92±2.00	14.49±3.70	16.94±2.96	18.37±3.29	18.78±3.54	20.92±1.95
ZEBRATOXIC	4.17	4.17	5.21	5.21	5.21	6.25	12.50
ZUBUD	3.48	4.35	4.35	5.22	6.96	13.91	20.0

TABLE VI
EVALUATION OF THE MINIMUM NODE SIZE (n_{min}) WITH ET-DIC. OTHER PARAMETERS ARE CONSTANT: $T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION WITH BEST SIZE INTERVALS AND 16×16 DESCRIPTORS

TABLE VII
INFLUENCE OF THE SUBWINDOW SIZE INTERVALS WITH ET-FL.
OTHER PARAMETERS ARE CONSTANT: $n_{min} = 5000$, $T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, AND TERMINAL FREQUENCY IMAGE REPRESENTATION. IN ITALIC THE SIZE EVALUATION SCENARIO FOR BIG

Datasets	1×1	$0 - 10\%$	$0 - 25\%$	$0 - 50\%$	$0 - 75\%$	$0 - 90\%$	$0 - 100\%$	$25 - 50\%$	$25 - 75\%$	$25 - 100\%$	$50 - 75\%$	$50 - 100\%$	$75 - 100\%$	$90 - 100\%$
ACROSOMES	36.43 ± 1.48	6.02 ± 0.58	3.28 ± 0.53	3.21 ± 0.45	3.19 ± 0.43	3.37 ± 0.85	3.80 ± 0.74	2.90 ± 0.68	3.15 ± 0.58	3.10 ± 0.78	3.84 ± 0.81	4.49 ± 0.81	4.85 ± 0.62	5.71 ± 0.85
AGEMAP-FAGING	4.19±1.77	8.95±3.36	16.57±3.02	24.10±3.54	24.10±3.01	26.10±5.25	25.90±4.28	29.14±3.08	27.81±3.92	28.86±3.51	27.05±4.43	27.05±4.31	31.24±4.80	33.24±3.75
ALL-IDB2	0.19 ± 0.58	0.38 ± 0.77	1.15 ± 0.94	1.35 ± 1.23	1.15 ± 1.28	0.96 ± 0.96	1.54 ± 1.15	1.73 ± 2.18	1.35 ± 1.23	2.12 ± 2.18	1.92 ± 1.92	2.31 ± 1.68	3.46 ± 1.44	12.88 ± 6.20
ALOI	70.16	58.36	47.14	32.79	26.42	25.78	27.62	23.78	*23.03*	25.55	26.12	28.30	32.15	35.07
ALOT	30.86±0.34	24.72±0.32	25.50±0.29	26.62±0.29	27.36±0.43	27.71±0.32	27.05±0.38	26.10±0.29	26.62±0.51	26.80±0.25	26.69±0.49	26.73±0.36	26.63±0.38	27.54±0.50
APOPTOSIS	40.71 ± 7.09	8.00 ± 3.39	3.71 ± 1.71	4.71 ± 2.56	5.71 ± 2.56	5.71 ± 1.81	5.57 ± 3.35	5.71 ± 2.12	5.71 ± 2.86	5.00 ± 3.21	7.86 ± 1.84	7.29 ± 2.59	10.00 ± 3.26	12.29 ± 2.32
BINUCLEATE	26.16±7.59	3.0±2.45	1.25±2.02	2.50±3.35	2.75±5.18	5.00±5.02	1.75±3.17	2.00±3.32	3.25±4.88	3.75±5.15	5.00±4.03	6.00±6.34	10.00±8.06	10.25±6.56
BIRDS	65.23±2.13	60.43±2.33	54.4±1.64	49.60±1.61	45.13±2.22	43.53±3.21	44.77±1.80	46.37±3.03	44.0±2.08	43.17±2.82	*41.23±2.85*	42.0±2.27	43.43±1.89	47.7±3.17
BREAST-CANCER	11.53±2.78	7.08±3.59	9.31±3.11	8.47±2.10	12.22±2.83	12.22±2.39	12.64±1.01	10.83±2.47	*11.94±4.22*	12.22±3.82	12.08±2.24	14.31±4.21	14.17±3.77	18.47±5.12
BUILDINGSAB	70.11±2.51	54.36±2.73	*52.09±2.83*	53.92±2.39	55.91±2.75	54.92±2.28	58.18±1.42	51.27±3.97	51.66±2.87	55.64±2.88	55.97±2.99	60.33±2.62	68.67±2.88	81.33±2.48
BUTTERFLIES	65.54±2.30	57.62±1.33	53.07±2.34	48.81±3.07	43.62±3.42	44.21±3.36	43.32±3.09	44.37±3.04	41.97±2.73	41.42±3.58	39.61±2.80	*39.22±2.17*	42.75±3.93	46.89±3.48
BUTTERFLIES-CLEAN	33.56±1.80	31.69±1.98	25.27±2.27	21.6±2.30	20.42±2.20	20.98±2.72	21.09±2.19	*15.62±2.34*	17.31±1.60	18.22±2.45	16.93±1.36	19.78±1.99	20.91±1.44	24.56±2.79
CALTECH-101	84.16	83.25	80.15	78.56	76.43	75.78	74.96	75.20	73.95	72.80	72.49	72.74	*73.63*	76.04
CALTECH-256	94.11±0.37	93.81±0.16	92.36±0.31	90.79±0.35	90.13±0.30	89.95±0.29	89.77±0.26	89.29±0.21	89.09±0.27	88.88±0.41	88.76±0.33	89.11±0.43	89.78±0.29	91.26±0.21
C.ELEGANS	65.21±3.46	26.15±3.36	27.78±4.78	34.53±4.53	38.55±4.50	44.96±1.95	41.63±4.12	40.68±4.24	41.45±5.10	43.59±4.98	44.79±2.60	46.24±3.83	52.39±4.43	54.87±4.45
C.ELEGANS-LIVE-DEAD	21.58±4.97	1.58 ± 2.41	3.16 ± 3.49	2.63 ± 2.63	4.74 ± 3.68	4.21 ± 6.57	3.16 ± 3.49	2.63 ± 2.63	3.16 ± 3.49	3.16 ± 4.21	3.16 ± 3.49	3.16 ± 4.82	4.74 ± 3.68	13.68 ± 7.52
CHARS74K	96.85±0.56	92.51±0.97	90.51±0.92	87.78±1.38	84.82±1.24	83.09±0.97	83.03±1.30	85.49±1.01	81.62±1.31	80.14±1.55	77.76±0.97	77.54±1.29	75.34±1.50	78.66±0.63
CHO	7.99±3.76	4.77±3.73	2.92±1.28	*2.31±1.24*	3.23±2.43	3.54±3.00	4.77±2.33	5.23±2.68	5.69±2.84	5.85±2.65	7.99±3.14	6.77±3.01	16.0±3.01	31.54±5.85
CIFAR-10	74.44	68.31	66.80	63.47	60.42	58.57	57.05	61.22	58.05	54.95	55.86	54.33	*53.24*	55.75
COIL-100 nv1	23.11	32.15	27.63	*25.04*	26.55	26.99	26.65	23.35	23.79	25.33	33.52	41.14	60.42	67.35
CONVEX	50.10	33.93	21.01	15.76	12.55	11.29	11.35	12.53	10.35	10.23	8.96	*9.93*	13.28	17.30
ETH-80	17.67±5.57	18.45±6.95	20.41±9.08	16.62±7.16	17.82±5.86	16.45±5.92	17.10±7.62	16.94±5.24	17.07±7.63	16.43±5.69	18.61±8.39	17.01±6.62	17.59±5.91	18.49±6.51
EVENTS	48.48±2.23	36.99±2.54	37.13±1.16	35.33±1.96	35.17±2.41	35.48±1.86	37.08±1.89	35.625±1.09	36.89±2.27	36.54±2.90	38.04±2.23	*37.52±2.48*	40.85±2.05	43.44±1.29
FAMOUSLAND	82.55±2.15	76.8±1.65	74.7±2.23	72.55±1.94	70.55±2.16	72.3±2.24	72.15±2.16	69.25±3.04	71.35±2.29	72.05±2.35	*70.45±2.83*	72.65±4.58	73.95±2.70	77.4±2.57
FLOWERS17	47.47±1.85	37.5±1.82	34.97±2.30	34.06±2.60	33.38±2.40	33.06±2.08	34.56±2.10	33.08±2.08	33.29±1.91	34.88±1.56	33.82±1.69	*34.38±1.70*	37.15±1.37	40.5±1.61
FLOWERS102-CLEAN	78.30		63.06	61.78	62.08	62.80	62.69	59.64	59.09	60.16	58.01	59.69	62.88	67.02
GALAXYZOO	17.26±0.18	19.39±0.27	14.62±0.16	11.49±0.15	10.30±0.24	9.31±0.11	8.48±0.27	9.66±0.26	9.15±0.19	7.61±0.15	8.23±0.17	6.71±0.16	4.96±0.23	3.88±0.15
GTSRBCROP	78.71	72.84	67.90	59.34	51.21	47.07	43.24	51.08	45.66	38.69	37.89	33.29	29.69	28.19
HCC	13.21 ± 8.53	11.79 ± 7.24	17.86 ± 10.71	20.36 ± 11.36	15.00 ± 8.72	16.79 ± 12.22	20.00 ± 11.43	21.43 ± 13.36	22.14 ± 8.72	22.14 ± 9.01	21.07 ± 8.89	21.43 ± 8.45	22.50 ± 8.23	27.50 ± 13.03
HEP2-ICPR2012	74.80	61.85	59.54	60.08	60.49	60.08	61.04	57.22	58.72	60.63	60.08	62.53	65.53	67.57
HPA	4.76±2.92	3.14 ± 1.13	3.43 ± 1.06	3.81 ± 1.81	4.95 ± 1.94	5.71 ± 1.54	7.62 ± 1.81	5.43 ± 2.17	5.52 ± 3.15	8.95 ± 3.02	7.52 ± 2.71	9.71 ± 2.58	12.57 ± 4.08	16.57 ± 2.22
INDOOR	89.05±0.76	85.13±0.90	84.13±0.63	84.45±0.87	84.20±0.84	84.12±0.45	84.52±1.06	84.27±0.96	84.84±0.96	84.72±0.82	85.21±1.05	85.44±0.55	86.95±1.30	88.95±0.68
IRMA2005	54.9	26.90	21.10	16.50	15.40	14.0	13.5	13.80	13.90	13.30	13.10	15.40	*16.20*	19.40
IRMA2006	66.5	40.3	30.9	26.2	23.9	23.1	23.5	23.1	22.8	25.1	24.0	23.8	*25.70*	27.5
KTH-TIPS	51.07±2.97	*7.44±1.86*	7.44±1.39	8.66±1.87	8.12±0.88	10.41±1.43	10.22±2.03	8.10±1.11	8.98±1.69	10.46±1.61	10.71±1.63	12.05±1.69	16.24±2.23	24.39±1.55
KTH-TIPS2	55.79±8.30	48.38±6.68	46.90±8.20	47.87±6.27	48.42±9.01	50.09±7.74	46.67±8.29	*49.39±7.13*	49.73±8.59	51.88±7.38	52.87±	52.21±9.63	54.52±9.37	51.49±9.54
LANDUSE	23.71±2.18	21.88±1.74	21.48±1.38	21.55±2.12	23.17±2.73	25.90±2.17	27.74±1.37	23.33±1.00	25.07±2.10	28.52±2.40	26.74±1.87	33.05±1.53	41.71±2.17	50.81±2.34
LYMPHOMA	11.35±5.64	7.84±3.90	5.95±3.97	9.19±4.71	8.65±3.97	10.27±3.15	8.38±5.85	7.87±3.18	8.38±2.82	11.08±5.19	10.27±4.80	11.62±4.69	*12.43±5.43*	11.35±4.32
MMODALITY	49.73	23.51	23.32	26.03	26.68	26.41	26.87	26.60	26.76	29.01	29.81	29.85	32.37	36.15
MNIST	77.19	52.29	14.74	3.65	2.96	2.58	2.49	2.4	2.36	1.87	1.86	1.77	1.64	*1.98*
MNIST-12000	78.77	65.69	36.86	6.59	4.09	3.48	3.27	3.29	2.87	2.75	2.41	2.30	*2.63*	3.28
MNIST-ROTATION	78.46	78.92	68.58	42.16	28.65	22.98	20.90	26.69	20.41	16.35	14.82	13.18	*12.65*	16.00
MNIST-BIMG	85.72	78.65	69.06	48.24	38.58	32.87	29.55	36.39	31.90	26.91	27.24	23.41	*22.10*	24.91
MNIST-BRAND	85.69	78.65	70.20	42.03	29.83	21.22	16.70	24.73	21.27	13.03	14.23	11.00	*9.15*	12.15
MNIST-BIMG-ROT	86.84	86.94		78.23	74.216	70.57		72.60	69.63	62.56	64.90	58.93	*57.31*	61.07
MSTAR-S1	43.03	41.61	35.40	30.56	30.68	30.11	30.85	23.16	25.55	25.61	27.09	24.42	*19.12*	17.36
NATURALSCENES	86.50±0.59	40.17±0.56	*40.04±1.11*	43.89±0.63	47.18±1.05	49.72±0.57	52.60±0.81	46.98±0.57	50.87±0.88	55.94±1.10	56.09±0.60	59.97±0.87	65.82±0.84	71.02±0.77
NORB-UNIFORM	66.35	54.47	37.23	27.67	23.54	19.51	17.83	22.21	19.83	15.32	15.72	12.49	*13.28*	14.88
NORB-JITTCLUTT	80.04	78.36	71.00	58.39	50.23	44.93	43.07	51.85	45.73	40.03	40.46	35.38	*31.68*	32.96
OLIVA	51.99±1.30	33.97±1.12	29.99±1.21	29.74±1.43	31.27±0.85	32.56±0.95	*34.37±1.25*	30.63±1.05	32.93±0.99	35.31±1.45	35.59±1.18	38.49±0.78	43.43±1.64	49.35±1.43
ORL	7.1±1.45	1.35±1.10	2.25±1.08	*1.95±1.01*	2.05±1.33	2.1±1.53	2.2±1.05	2.05±1.39	1.95±1.42	2.5±1.57	3.2±2.11	3.4±1.51	3.65±1.47	5.75±2.88
OUTEX	1.04±0.28	*2.06±1.09*	1.46±0.55	4.33±1.28	4.70±1.54	6.71±1.16	7.89±0.87	5.60±1.03	8.52±1.19	10.39±1.04	10.99±1.50	15.09±0.94	24.05±1.89	38.22±2.56
PFID	77.02±2.12	75.82±0.92	74.73±2.36	72.73±3.12	75.00±2.27	74.43±2.49	74.97±2.23	75.00±2.99	74.67±3.58	77.73±2.98	76.42±3.45	77.87±2.28	79.64±3.05	81.37±2.68
POLLEN	33.13±1.88	12.48±1.24	9.42±0.93	6.21±0.88	6.32±1.17	5.36±0.97	5.15±1.07	5.17±0.90	4.64±1.03	4.48±0.57	3.83±0.69	3.99±0.74	3.42±0.37	*4.55±0.70*
PPMI24	93.875	92.21	90.00	89.21	87.29	87.12	87.71	87.88	87.21	86.71	85.79	86.12	86.71	88.58
PUBFIG83	94.37±0.71	91.72 ± 0.89	88.88 ± 1.20	85.83 ± 1.27	83.36 ± 0.88	82.69 ± 0.89	82.87 ± 0.81	82.01 ± 1.75	81.13 ± 1.37	81.22 ± 0.79	79.45 ± 0.64	80.96 ± 1.17	82.37 ± 1.62	85.28 ± 1.03
RBC	55.09±1.55	56.54±1.43	48.33±1.21	42.90±1.56	41.04±1.07	39.86±1.07	39.92±1.52	35.93±1.70	37.33±1.58	36.08±1.27	*32.95±1.82*	33.08±2.01	33.41±0.96	36.61±0.96
RECT-BASIC	50.03	12.07	7.43	0.93	0.48	0.35	0.43	0.39	0.18	0.28	*0.20*	0.14	0.32	0.92
RECT-BIMG	49.84	46.23			24.12	24.10	24.87	22.99	*22.26*	22.58	22.16	23.56	25.48	27.30
RNAI	32.5±5.12	14.5±3.32	23.25±5.81	39.5±5.79	55.25±4.93	58.75±6.35	58.75±6.54	48.75±4.07	62.0±4.30	70.25±6.27	68.25±10.00	72.75±7.07	76.75±5.92	80.75±8.66
SEROUS	36.23±2.36	29.21±1.38	28.91±1.42	27.65±2.09	27.81±2.36	29.67±2.25	29.59±2.05	29.29±1.94	29.37±1.91	30.11±3.11	30.79±2.33	*30.27±2.56*	33.14±2.12	35.82±1.43
SHAPE1	53.9±4.50	40.9±2.34	24.8±4.09	21.8±4.11	17.1±3.05	17.4±4.22	19.0±3.35	16.8±3.31	15.4±2.69	15.7±4.1	18.3±2.33	20.4±5.10	*26.4±5.48*	37.0±5.18
SMEAR	8.13±2.37	7.69±1.47	7.25±2.92	7.58±1.73	8.46±3.61	9.23±1.92	10.22±2.91	6.26±1.91	7.58±3.16	9.78±2.57	8.35±3.15	9.67±2.01	12.97±2.73	16.92±2.47
SOCCER	44.86±3.77	24.95±3.48	25.33±2.34	21.90±1.41	20.95±2.52	22.29±1.83	22.86±2.98	*20.86±2.90*	22.67±4.95	24.09±3.07	23.71±3.91	27.14±3.11	29.43±2.14	39.71±3.38
SOIL-24	67.92	61.67	*57.08*	35.42	34.17	32.08	38.75	22.5	20.83	47.08	27.5	66.67	76.67	92.5
SOIL-47	83.09	75.74	55.53	52.21	52.34	52.98	59.30	65.85	74.47	86.28	57.66	61.17	70.53	84.04
SPORTS	17.47±1.43	10.10±1.68	9.40±0.97	9.41±0.93	9.59±1.28	9.93±0.69	11.08±1.05	10.00±1.11	10.52±0.96	11.84±1.10	10.72±0.90	*11.70±1.20*	14.77±0.94	16.34±1.27
STOIC101	27.73	27.42	25.76	26.36	28.48	28.64	30.00	26.52	30.00	32.88	32.88	38.79	43.94	51.97
STOMATA	37.27 ± 13.14	30.00 ± 11.54	32.73 ± 7.27	28.18 ± 12.50	24.55 ± 12.89	22.73 ± 10.16	28.18 ± 14.91	20.91 ± 12.23	20.00 ± 12.73	22.73 ± 12.36	27.27 ± 9.09	21.82 ± 8.33	26.36 ± 13.14	34.55 ± 9.79
STONEFLY9	17.42	15.59	15.59	14.72	16.11	15.77	15.94	15.68	16.81	16.03	17.68	18.12	18.03	18.73
SUBCELLULAR	34.99±5.21	16.28±4.26	15.35±2.69	14.07±2.46	13.95±2.21	11.63±2.90	14.77±3.64	12.33±3.68	13.60±3.12	14.19±2.98	13.84±3.73	*15.35±3.78*	18.37±4.79	18.72±4.14
SUN397	97.96±0.10	97.14±0.11	96.57±0.17	95.99±0.15	95.51±0.14	95.42±0.23	95.28±0.19	95.54±0.16	95.27±0.15	95.07±0.09	95.19±0.11	95.17±0.19	95.45±0.18	96.04±0.11
SWELEAF	39.73	14.48±2.04	7.35±0.93	5.64±0.96	5.00±0.73	5.24±0.58	4.65±0.70	3.33±0.84	4.08±0.65	3.95±0.68	4.12±0.75	4.04±0.55	*4.60±0.70*	n/a
TERMINALBULB	80.22±0.51	58.31±1.52	57.45±0.66	56.20±0.85	55.35±2.05	54.87±2.61	56.07±1.67	55.80±1.81	54.89±1.61	56.08±2.06	55.35±1.16	56.79±1.20	59.91±2.03	66.30±1.39
TINYGRAZ03	65.13±4.23	66.73±2.63	65.31±2.09	64.25±3.24	62.30±2.88	63.36±4.32		63.72±2.99	61.42±5.19	60.00±3.19	*57.61±2.84*	57.88±3.27	55.84±3.27	58.76±3.17
TSG-60	4.75±0.65	19.33±1.38	14.75±4.84	4.08±1.51	4.92±1.60	5.08±0.87	5.75±1.08	*5.58±1.18*	7.75±1.75	7.17±1.59	16.58±2.80	39.75±2.79	73.17±3.82	73.5±4.34
UIUCTEX	49.72±1.73	*12.32±1.53*	12.92±1.32	15.42±1.58	18.62±2.10	19.7±1.82	21.96±1.53	17.16±1.21	20.64±1.20	22.44±1.66	24.02±2.12	26.14±1.69	31.68±1.23	36.98±1.42
WANG	17.48±1.54	12.14±1.64	*12.0±0.83*	13.22±0.58	14.48±1.26	15.44±0.99	15.32±1.45	15.46±1.43	16.12±1.44	17.56±1.88	16.72±1.45	18.54±1.33	22.38±1.80	23.72±1.47
XEROX-6	17.55±2.18	7.35±2.08	10.20±1.94	9.18±1.44	8.78±4.21	8.78±2.38	11.43±1.87	11.84±4.36	*9.80±4.64*	10.61±2.38	9.39±3.73	14.90±4.21	13.88±2.63	17.34±4.08
ZEBRATOXIC	50.00	34.75	26.04	13.54	5.21	5.21	6.26	9.38	4.17	5.21	4.17	8.33	10.42	13.54
ZUBUD	24.35	26.96	17.39	6.96	7.83	6.96	7.83	7.83	7.83	7.83	7.83	8.70	12.17	17.39

TABLE VIII
INFLUENCE OF THE n_{min} FOR ET-FL (LARGE VALUES). OTHER PARAMETERS ARE CONSTANT: $T = 10, k = sqrt(M), N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, BEST SUBWINDOW SIZE INTERVALS, AND TERMINAL FREQUENCY IMAGE REPRESENTATION. THE NUMBER OF VISUAL WORDS IS DECREASING FROM LEFT TO RIGHT.

Datasets	1000	2500	5000	10000	25000	50000
ACROSOMES	2.77 ± 0.49	2.52 ± 0.29	2.90 ± 0.68	3.26 ± 0.71	4.34 ± 0.79	5.86 ± 0.70
AGEMAP-FAGING	6.29±2.18	4.29±0.88	4.19±1.77	3.90±1.24	4.86±1.56	5.24±0.98
ALL-IDB2	0.38±0.77	0.96±0.96	0.19 ± 0.58	0.96±1.29	0.96±1.97	0.96±1.29
ALOI	23.15	22.74	23.03	24.51	25.85	26.97
ALOT	28.88±0.36	23.92±0.41	24.72±0.32	26.27±0.35	29.96±0.50	34.82±0.41
APOPTOSIS	3.43 ± 1.46	4.00 ± 2.10	3.71 ± 1.71	4.57 ± 3.05	5.00 ± 2.58	7.14 ± 2.47
BINUCLEATE	0.25±0.75	1.25±3.01	1.25±2.02	2.00±2.18	0.75±1.15	0.75±2.25
BIRDS	41.63±2.66	41.0±3.08	41.23±2.85	43.1±2.45	43.9±2.70	48.27±2.57
BREAST-CANCER	10.14±2.71	9.17±3.73	7.08±3.59	8.89±2.26	9.31±2.33	10.14±3.73
BUILDINGSAB	47.62±2.68	49.99±3.43	52.09±2.83	54.92±3.19	52.98±3.24	52.38±2.66
BUTTERFLIES	43.32±2.94	41.05±2.37	39.22±2.17	40.05±3.44	40.23±3.13	42.38±1.88
BUTTERFLIES-CLEAN	14.13±1.22	15.93±1.82	15.62±2.34	16.49±2.03	18.78±2.26	20.6
CALTECH-101	71.27	72.91	72.49	73.64	75.33	76.95
CALTECH-256	87.96±0.22	88.32±0.33	88.76±0.33	89.59±0.44	90.52±0.19	91.08±0.38
C.ELEGANS	27.52±4.72	27.09±3.27	26.15±3.36	27.35±2.65	27.69±2.87	30.09±3.33
C.ELEGANS-LIVE-DEAD	4.21±3.94	3.16 ± 3.49	1.58 ± 2.41	3.16 ± 3.49	4.74 ± 5.49	3.68 ± 2.41
CHARS74K	73.76±1.17	73.71±1.77	75.34±1.50	77.09±1.43	78.33±1.41	79.95±1.09
CHO	3.08±1.95	2.92±2.33	2.31±1.24	4.46±2.33	6.31±2.96	8.77±2.39
CIFAR-10	55.60	55.28	53.24	54.78	55.93	58.83
COIL-100 nv1	25.46	24.83	23.35	19.79	22.48	25.48
CONVEX	9.56	8.41	8.96	10.49	10.27	11.00
ETH-80	17.65±6.39	17.35±7.15	16.43±5.69	16.19±6.56	15.70±3.93	17.83±5.51
EVENTS	34.27±1.79	35.73±2.37	37.46±1.48	40.08±1.49	42.46±2.54	43.17±1.60
FAMOUSLAND	67.75±2.76	68.75±3.51	69.25±3.04	71.05±3.19	72.8±2.91	75.0±2.25
FLOWERS17	33.32±3.44	32.41±3.29	33.06±2.08	33.62±2.09	36.74±1.72	39.99±1.61
FLOWERS102-CLEAN	59.54	56.90	58.01	59.83	62.93	64.58
GALAXYZOO	2.84±0.13	3.49±0.19	3.88±0.15	4.09±0.22	4.84±0.42	5.97±0.31
GTSRBCROP	24.44	25.55	28.19	29.20	33.52	38.03
HCC	17.14 ± 6.93	10.71 ± 10.71	11.79 ± 7.24	13.21 ± 9.92	16.07 ± 9.55	16.43 ± 9.61
HEP2-ICPR2012	58.04	59.54	57.22	59.13	60.08	61.31
HPA	2.95 ± 1.50	2.38 ± 0.98	3.14 ± 1.13	3.62 ± 1.80	4.48 ± 1.65	6.00 ± 2.04
INDOOR	85.11±0.64	84.03±0.88	84.12±0.45	85.07±0.91	86.74±0.68	87.19±1.30
IRMA2005	11.3	12.60	13.10	13.90	15.70	16.80
IRMA2006	23.10	22.40	22.80	23.60	25.60	26.90
KTH-TIPS	9.24±1.54	7.29±1.00	7.44±1.86	7.78±1.47	7.76±1.59	8.80±0.94
KTH-TIPS2	52.44±10.32	54.55±9.35	46.67±8.29	52.83±6.25	52.76±9.29	53.15±7.50
LANDUSE	17.76±1.90	17.55±1.60	21.48±1.38	19.31±1.70	22.79±2.01	26.90±2.00
LYMPHOMA	9.19±4.71	8.38±4.90	5.95±3.97	7.57±3.38	10.00±4.53	11.08±4.59
MMODALITY	22.06	21.79	23.32	24.05	28.89	28.28
MNIST	1.44	1.4	1.64	1.88	2.66	3.64
MNIST-12000	2.72	2.18	2.30	2.90	3.51	5.05
MNIST-ROTATION	9.42	10.58	12.65	15.83	19.62	24.25
MNIST-BIMG	20.39	21.65	22.10	24.29	26.46	26.12
MNIST-BRAND	8.21	8.40	9.15	11.77	14.08	12.76
MNIST-BIMG-ROT	55.05	58.22	57.31	60.232	60.67	60.26
MSTAR-S1	16.73	18.10	17.36	20.15	26.64	26.35
NATURALSCENES	40.00±0.69	39.31±0.81	40.04±1.11	41.91±1.09	45.13±0.77	48.21±0.78
NORB-UNIFORM	9.95	10.67	12.49	14.74	19.75	23.47
NORB-JITTCLUTT	24.71	28.65	31.68	35.20	40.02	43.83
OLIVA	28.01±0.68	28.36±0.86	29.74±1.43	31.90±1.03	34.81±1.33	38.49±1.38
ORL	2.1±1.66	1.75±0.78	1.35±1.10	1.45±1.11	1.3±1.17	1.8±1.58
OUTEX	1.20±0.62	0.90±0.59	1.04±0.28	1.46±0.57	2.66±1.08	3.01±0.88
PFID		72.87±2.43	72.73±3.12	73.72±3.00	75.57±2.67	74.67±2.78
POLLEN	3.88±0.66	3.53±0.55	3.42±0.37	3.95±0.45	5.37±0.59	5.81±0.90
PPMI24	85.00	85.33	85.79	86.83	87.75	86.96
PUBFIG83	78.54	78.08±0.89	79.45±0.64	81.27 ± 1.09	83.16 ± 1.17	85.04 ± 0.91
RBC	28.76±1.18	30.04±1.28	32.95±1.82	36.88±1.20	42.91±1.20	41.15±0.70
RECT-BASIC	0.09	0.21	0.14	0.20	0.19	0.27
RECT-BIMG	17.78	20.07	22.16	20.95	20.76	20.90
RNAI	14.25±5.48	9.5±3.67	14.5±3.32	5.5±4.15	11.75±4.34	15.5±7.14
SEROUS	24.04±2.44	27.57±2.52	27.65±2.09	30.27±1.93	32.76±1.52	31.94±2.08
SHAPE1	17.5±4.59	16.1±3.99	15.4±2.69	16.6±3.75	18.3±3.93	16.9±3.01
SMEAR	7.58±2.43	7.25 ± 2.32	6.26±1.91	8.57 ± 2.73	9.45 ± 3.34	12.20 ± 2.11
SOCCER	17.52±2.86	20.0±2.33	20.86±2.90	22.10±4.30	24.38±2.77	27.05±4.47
SOIL-24	9.58	19.17	20.83	31.25	35.0	55.0
SOIL-47	48.72	53.19	52.21	57.02	64.04	65.85
SPORTS	9.46±1.02	8.94±1.00	9.40±0.97	9.61±0.95	12.35±1.03	13.66±1.31
STOIC101	31.67	27.58	25.76	25.76	29.55	33.48
STOMATA	20.91 ± 11.54	20.91 ± 7.10	20.00 ± 12.73	25.45 ± 11.35	23.64 ± 12.33	25.45 ± 10.60
STONEFLY9	14.98	15.42	14.72	15.77	16.72	16.03
SUBCELLULAR	11.86±2.84	13.02±3.32	11.63±2.90	11.63±2.49	16.51±3.63	15.35±2.98
SUN397	95.18±0.19	94.91±0.13	95.07±0.09	95.45±0.15	95.42±0.16	95.48±0.19
SWELEAF	4.25±0.55	4.43±0.96	3.33±0.84	4.33±1.02	5.03±0.74	6.11±0.68
TERMINALBULB	55.20±1.41	54.23±1.61	54.87±2.61	53.86±1.47	55.08±1.88	56.26±1.80
TINYGRAZ03	53.89±1.65	54.16±3.51	55.84±3.27	53.54±2.28	62.21±2.69	68.05±1.88
TSG-60	3.58±1.49	8.83±1.72 ?	4.08±1.51	3.92±1.24	4.58±1.13	6.08±1.40
UIUCTEX	13.38±1.54	12.5±1.47	12.32±1.53	11.82±1.0	12.5±0.93	13.84±0.96
WANG	11.44±0.77	11.7±1.50	12.0±0.83	12.0±0.69	14.8±0.85	16.68±1.06
XEROX-6	11.43±2.18	10.41±1.50	7.35±2.08	11.84±2.29	15.51±1.63	14.29±3.87
ZEBRATOXIC	3.12	4.17	4.17	5.21	8.33	8.33
ZUBUD	9.57	11.30	6.96	7.83	7.83	9.57

TABLE IX
INFLUENCE OF LOWER n_{min} FOR E T-FL (SMALL VALUES). OTHER PARAMETERS ARE CONSTANT: $T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, BEST SUBWINDOW SIZE INTERVALS, AND TERMINAL FREQUENCY IMAGE REPRESENTATION. THE NUMBER OF VISUAL WORDS IS DECREASING FROM LEFT TO RIGHT. FINAL CLASSIFIER WAS LIBLINEAR.

Datasets	2	10	100	500	1000
ACROSOMES	3.68 ± 0.70	3.41 ± 0.47	2.50 ± 0.55	2.79 ± 0.82	2.68 ± 0.62
AGEMAP-FAGING	42.29±3.52	38.10±2.69	18.95±3.00	7.62±0.95	4.86±1.44
ALL-IDB2	1.15±1.54	0.77±1.28	1.15±0.94	0.38±0.77	0.38±0.77
ALOI	n/a	n/a	n/a	n/a	n/a
ALOT	44.21±0.28	38.43±0.42	30.92±0.33	28.49±0.24	28.88±0.36
APOPTOSIS	12.57 ± 3.31	12.14 ± 3.34	6.14 ± 1.44	5.71 ± 2.47	3.86 ± 1.81
BINUCLEATE	12.0±6.96	6.50±5.61	3.00±3.50	1.50±3.20	0.25±0.75
BIRDS	47.0±2.97	46.27±3.05	40.33±2.47	39.9±2.92	40.13±1.37
BREAST	13.61±6.11	13.33±6.34	12.64±3.75	9.44±3.27	9.44±3.28
BUILDINGSAB	41.71±2.39	41.77±2.45	43.31±5.53	44.53±2.50	43.31±2.06
BUTTERFLIES	52.27±2.24	49.06±2.44	44.14±2.54	41.12±2.65	40.78±3.12
BUTTERFLIES-CLEAN	19.53±6.96	17.8±1.79	13.98±1.45	12.13±4.19	12.18
CALTECH-101	74.12	71.75	70.35	69.74	70.48
CALTECH-256	89.76±0.44	88.68±0.29	87.77±0.31	88.17±0.33	88.53±0.36
C.ELEGANS	40.09±4.84	41.97±14.49	30.17±4.17	30.17±5.32	26.58±3.71
C.ELEGANS-LIVE-DEAD	8.42 ± 5.37	4.74 ± 4.37	1.58 ± 2.41	3.16 ± 3.49	3.68 ± 3.37
CHARS74K	73.43±1.30	73.29±1.94	71.83±1.44	71.24±2.05	71.96±1.81
CHO	4.46±1.45	5.23±2.59	3.69±3.31	2.92±2.70	2.46±2.09
CIFAR-10	50.71	49.93	50.26	53.66	55.60
COIL-100	22.83	22.0	22.31	25.31	27.0
CONVEX	10.54	10.37	8.79	7.72	7.47
ETH-80	22.36±8.18	23.09±7.88	18.22±6.83	17.44±6.75	14.01±6.71
EVENTS	43.31±1.17	39.77±1.18	33.23±2.26	32.56±1.14	32.73±2.86
FLOWERS17	37.03±2.39	33.91±1.79	29.59±1.99	29.59±2.00	29.0±1.88
FLOWERS102-CLEAN	58.55	55.83	51.96	52.00	53.31
FAMOUS	72.95±2.14	68.6±3.38	63.2±2.40	62.25±2.70	60.95±2.13
GALAXYZOO	3.79±0.16	3.62±0.12	2.82±0.06	2.69±0.10	2.84±0.13
GTSRBCROP	28.86	27.86	25.24	25.10	25.00
HCC	43.93 ± 12.43	37.86 ± 13.57	24.29 ± 9.95	19.29 ± 8.48	14.99± 7.11
HEP2-ICPR2012	62.67	61.17	61.58	59.67	59.40
HPA	13.90 ± 3.16	10.67 ± 3.09	5.52 ± 2.51	3.62 ± 1.58	3.71 ± 2.02
INDOOR	87.33±0.92	85.22±0.82	84.46±0.64	84.55±0.79	85.11±0.64
IRMA2005	21.3	17.3	13.6	11.5	11.3
IRMA2006	39.5	31.1	24.5	22.8	22.4
KTH-TIPS	23.51±1.46	18.68±2.31	11.49±1.93	8.46±2.95	7.66±1.79
KTH-TIPS2	50.99±9.29	50.05±9.92	49.61±9.20	47.71±9.27	50.76±11.39
LANDUSE	32.5±1.89	27.69±1.97	20.33±1.96	17.14±1.77	17.19±1.58
LYMPHOMA	27.2±4.31	21.2±8.31	12.53±3.38	9.33±3.27	9.60±4.37
MMODALITY	25.76	24.50	21.64	21.64	22.06
MNIST	1.64	1.47	1.28	1.22	1.42
MNIST-12000	2.56	2.32	1.89	1.90	1.92
MNIST-ROTATION	10.25	9.61	8.14	8.57	9.24
MNIST-BIMG	28.572	26.28	20.98	19.36	20.39
MNIST-BRAND	14.63	13.14	8.48	7.94	8.21
MNIST-BIMG-ROT	64.06	62.02	54.46	54.91	55.05
MSTAR-S1	41.32	38.99	22.20	16.79	17.47
NATURALSCENES	50.89±1.45	47.66±0.60	42.77±0.76	40.21±0.83	39.44±0.65
NORB-UNIFORM	8.56	7.99	8.07	9.21	9.95
NORB-JITTCLUTT	19.68	18.57	19.66	22.69	24.71
OLIVA	36.44±1.64	33.78±0.51	28.08±0.89	26.42±1.06	27.13±1.23
ORL	1.35±1.25	1.2±1.12	0.95±0.79	1.45±1.33	1.35±0.79
OUTEX	3.33±1.04	2.25±0.94	0.83±0.48	0.79±0.47	0.74±0.42
PFID	75.66±1.62	74.84±3.15	71.09±3.06	69.56±3.29	69.73±4.50
POLLEN	12.04±0.97	10.24±1.34	5.32±0.97	4.23±0.87	4.19±0.83
PPMI24	87.79	85.54	84.83	84.25	85.00
PUBFIG83		81.05 ± 1.70	77.48 ± 0.92	77.81 ± 1.35	78.54 ± 1.00
RBC	41.12±1.10	38.55±2.22	32.19±1.54	29.80±0.81	29.80±0.87
RECT-BASIC	0.04	0.03	0.03	0.04	0.07
RECT-BIMG	15.81	15.60	14.91	15.88	16.83
RNAI	42.75±3.78	38.75±4.22	24.25±4.04	18.25±5.81	16.75±5.37
SEROUS	34.91±0.93	33.31±1.86	27.90±1.40	27.19±2.20	27.38±2.59
SHAPE1	21.0±7.95	21.4±3.20	19.2±7.74	16.2±6.97	16.1±6.24
SMEAR	13.41±2.98	11.87 ± 2.64	8.13 ± 2.10	6.37 ± 2.19	7.47 ± 2.35
SOCCER	19.33±7.30	21.14±3.32	14.76±4.13	16.0±3.09	17.43±3.16
SOIL-24	5.0	4.58	6.67	11.25	16.67
SOIL-47	29.26	34.26	35.32	43.62	48.08
SPORTS	18.14±1.24	14.54±4.91	10.65±3.72	8.87±3.08	8.51±1.17
STOIC101	62.12	46.06	30.91	23.48	21.52
STOMATA	25.45 ± 11.35	30.00 ± 15.77	21.82 ± 7.27	26.36 ± 16.98	22.73 ± 10.16
STONEFLY9	25.17	20.30	16.03	16.03	16.11
SUBCELLULAR	20.58±5.72	17.91±4.07	13.60±4.19	14.19±2.19	13.26±3.79
SUN397	n/a	94.38±0.12	94.09±0.10	94.21±0.13	94.45±0.10
SWELEAF	12.03±0.84	9.88±1.06	5.27±0.85	4.07±0.78	4.08±0.66
TERMINALBULB	64.63±2.47	63.01±2.55	57.89±1.87	55.35±2.01	55.38±1.70
TINYGRAZ03	60.97±4.69	59.38±2.75	54.96±3.78	61.33±3.24	59.73±73
TSG-60	n/a	1.83±1.38	1.17±0.76	1.33±1.0	2.33±1.53
UIUCTEX	25.54±1.62	20.70±1.39	15.9±1.29	13.82±1.42	12.46±1.62
XEROX-6	18.06±3.26	15.31±2.19	12.04±1.70	9.08±2.43	9.90±4.01
WANG	15.68±0.99	14.32±1.90	12.1±1.65	10.07±1.93	10.04±1.39
ZEBRATOXIC	6.25	7.29	5.21	5.21	3.12
ZUBUD	6.96	8.70	7.83	7.83	9.56

TABLE X
INFLUENCE OF THE NUMBER OF RANDOM TESTS FOR ET-FL. OTHER PARAMETERS ARE CONSTANT: $T = 10$, $n_{min} = 5000$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, BEST SUBWINDOW SIZE INTERVALS, AND TERMINAL FREQUENCY IMAGE REPRESENTATION.

Datasets	1	$\sqrt{M}$	$M/8$	$M/4$	$M/2$	M
ACROSOMES	2.97 ± 0.51	2.90 ± 0.68	2.95 ± 0.67	2.86 ± 0.68	3.23 ± 0.56	3.12 ± 0.79
AGEMAP-FAGING	3.62±1.94	4.19±1.77	4.10±1.21	4.67±2.11	4.19±1.77	4.19±1.77
ALL-IDB2	1.54±1.68	0.19±0.58	1.54±1.88	1.54±1.88	1.54±1.88	1.54±1.88
ALOI	31.58	23.03	22.25	21.39	20.27	20.18
ALOT	32.33±0.43	24.72±0.32	25.32±0.42	25.66±0.45	25.69±0.34	25.62±0.36
APOPTOSIS	4.00 ± 2.10	3.71 ± 1.71	4.29 ± 3.00	4.29 ± 2.39	5.23±2.13	4.43 ± 2.51
BINUCLEATE	2.25±3.78	1.25±2.02	3.0±2.69	5.00±3.16	4.75±3.78	4.17±3.99
BIRDS	47.4±2.15	41.23±2.85	41.83±2.30	39.97±1.78	39.10±2.70	38.37±2.04
BREAST-CANCER	10.14±3.17	7.08±3.59	9.44±2.96	10.55±2.86	9.99±3.21	17.92±2.74
BUILDINGSAB	49.89±2.07	52.09±2.83	52.38±2.22	50.99±2.53	51.44±2.29	52.76±2.07
BUTTERFLIES	54.21±1.66	39.22±2.17	34.85±3.23	36.34±1.70	33.99±3.44	30.92±2.45
BUTTERFLIES-CLEAN	21.58±2.74	15.62±2.34	15.20±2.26	13.87±1.35	16.11±2.00	15.04±1.65
CALTECH-101	71.89	72.49	72.28	72.79	72.71	72.30
CALTECH-256	88.29±0.31	88.76±0.33	88.93±0.36	88.73±0.34	88.70±0.30	88.95±0.26
C.ELEGANS	27.69±4.57	26.15±3.36	26.58±3.55	25.47±3.84	30.59±5.67	28.80±6.10
C.ELEGANS-LIVE-DEAD	2.63 ± 4.85	1.58 ± 2.41	3.68 ± 4.11	2.63 ± 4.85	3.16 ± 3.49	3.16 ± 3.49
CHARS74K	76.87±1.44	75.34±1.50	75.54±1.75	75.56±1.64	76.26±1.61	74.25±1.74
CHO	2.77±2.04	2.31±1.24	3.85±3.25	3.23±2.62	3.23±1.88	2.92±1.75
CIFAR-10	56.29	53.24	53.19	52.80	53.48	53.72
COIL-100 nv1	28.51	23.35	24.20	24.13	25.58	28.85
CONVEX	12.49	8.96	8.88	8.91	8.79	8.88
ETH-80	15.70±6.49	16.43±5.69	17.38±9.06	15.63±8.51	15.66±5.27	n/a
EVENTS	35.08±2.19	35.17±2.41	35.46±2.39	37.42±1.66	36.77±2.55	35.83±2.65
FAMOUSLAND	63.7±2.39	69.25±3.04	70.75±1.74	71.25±5.42	71.17±2.35	71.05±2.41
FLOWERS17	34.32±1.57	33.06±2.08	33.38±1.57	33.21±2.94	34.62±2.43	32.76±1.99
FLOWERS102-CLEAN	57.82	58.01	57.83	57.44	56.99	57.55
GALAXYZOO	9.65±0.39	3.88±0.15	3.28±0.14	3.10±0.08	2.90±0.12	2.90±0.09
GTSRBCROP	45.30	28.19	24.18	23.02	21.92	20.32
HCC	17.86 ± 8.89	11.79 ± 7.24	16.79 ± 8.23	15.36 ± 9.66	14.29 ± 8.14	15.58± 8.24
HEP2-ICPR2012	60.90	57.22	58.45	57.77	56.95	58.72
HPA	3.33 ± 1.3	3.14 ± 1.13	3.71 ± 1.31	4.29 ± 1.36	3.71 ± 2.19	3.49±1.05
INDOOR	83.13±0.72	84.12±0.45	84.22±0.62	83.99±0.55	83.84±0.54	84.36±0.95
IRMA2005	13.80	13.10	13.10	14.10	13.90	13.4
IRMA2006	24.2	22.8	23.5	23.90	23.20	25.4
KTH-TIPS	9.05±0.74	7.44±1.86	8.37±1.26	7.46±1.02	6.75±1.51	8.02±0.93
KTH-TIPS2	52.96±10.11	46.67±8.29	54.26±11.04	51.83±7.96	50.96±7.94	52.13±9.23
LANDUSE	18.62±1.31	21.48±1.38	17.24±0.95	18.07 ± 1.81	18.43 ± 1.22	16.74 ± 1.53
LYMPHOMA	6.76±4.72	5.95±3.97	8.13±2.42	8.93±2.92	7.60±2.67	7.60±2.07
MMODALITY	22.71	23.32	23.44	22.90	23.32	23.74
MNIST	1.98	1.64	1.63	1.56	1.81	1.71
MNIST-12000	2.49	2.30	2.39	2.61	2.58	2.67
MNIST-ROTATION	14.31	12.19	12.36	12.16	12.18	11.93
MNIST-BIMG	47.06	22.10	20.30	18.59	18.26	17.54
MNIST-BRAND	19.52	9.15	8.44	8.63	8.51	8.33
MNIST-BIMG-ROT	74.65	57.31	57.35	55.69	54.16	50.16
MSTAR-S1	16.28	17.36	15.82	13.49	17.99	15.48
NATURALSCENES	43.35±0.75	40.04±1.11	39.62±1.04	38.84±0.63	38.75±0.74	38.54±0.78
NORB-UNIFORM	14.88	12.49	12.35	13.24	13.61	14.89
NORB-JITTCLUTT	48.97	31.68	29.21	28.16	27.08	26.54
OLIVA	27.90±0.85	29.74±1.43	29.54±1.14	30.54±0.81	30.56±0.94	30.60±0.97
ORL	2.95±1.21	1.35±1.10	2.7±2.12	2.4±1.71	2.45±1.25	2.3±1.27
OUTEX	1.18±0.33	1.04±0.28	1.34±0.64	2.04±0.91	1.43±0.51	1.04±0.28
PFID	74.45±3.11	72.73±3.12	75.41±2.58	74.67±2.66	73.88±3.75	
POLLEN	4.77±0.75	3.42±0.37	4.57±0.65	5.06±0.64	5.01±0.86	5.23±1.00
PPMI24	88.33	85.79	85.04	85.62	85.71	85.83
PUBFIG83	81.34±1.05	79.45 ± 0.64	78.70 ± 1.33	79.90 ± 1.12	79.94 ± 1.42	80.59 ± 1.26
RBC	45.57±1.62	32.95±1.82	32.64±1.18	30.74±0.96	29.50±1.43	30.30±1.97
RECT-BASIC	0.22	0.14	0.20	0.19	0.20	0.23
RECT-BIMG	23.89	22.16	22.15	22.18	21.90	21.73
RNAI	17.5±6.52	14.5±3.32	11.0±5.15	13.0±4.97	10.83±3.33	15.25±5.96
SEROUS	26.78±1.98	27.65±2.09	26.09±2.35	25.55±1.66	26.67±1.42	28.17±1.81
SHAPE1	17.0±3.44	15.4±2.69	19.8±2.99	17.7±3.07	17.8±4.49	19.3±2.45
SMEAR	8.68 ± 2.88	6.26±1.91	7.36 ± 2.36	8.68 ± 3.16	9.45 ± 1.85	7.69 ± 1.84
SOCCER	24.29±5.51	20.86±2.90	19.14±2.23	20.76±5.07	18.66±2.49	18.86±4.17
SOIL-24	52.08	20.83	22.5	24.58	19.58	23.75
SOIL-47	61.06	52.21	54.36	54.68	53.94	54.89
SPORTS	10.49±0.70	9.40±0.97	9.30±1.14	8.99±1.18	8.77±0.77	8.71±0.97
STOIC101	30.76	25.76	26.97	26.36	25.91	24.09
STOMATA	21.82 ± 10.91	20.00 ± 12.73	25.45 ± 9.79	22.73 ± 6.10	24.55 ± 10.00	25.45 ± 14.55
STONEFLY9	15.16	14.72	14.46	14.98	15.16	15.42
SUBCELLULAR	14.19±3.11	11.63±2.90	13.72±3.56	13.60±3.72	15.0±4.04	13.72±1.93
SWELEAF	3.24±0.37	3.33±0.84	4.91±0.96	5.27±0.82	5.76±1.01	4.29±0.72
SUN397	94.83±0.18	95.07±0.09	95.18±0.19	95.16±0.13	95.25±0.17	95.19±0.12
TERMINALBULB	55.28±1.84	54.87±2.61	53.37±2.07	55.22±1.71	55.28±1.22	53.85±1.74
TINYGRAZ03	61.50±2.66	55.84±3.27	61.15±4.52	60.53±4.95	60.80±3.36	59.03±5.37
TSG-60	5.25±1.40	4.08±1.51	n/a	n/a	n/a	n/a
UIUCTEX	11.82±1.47	12.32±1.53	11.38±1.10	12.16±1.09	12.06±1.13	12.18±0.58
WANG	10.74±0.89	12.0±0.83	12.14±1.62	12.2±1.41	12.66±1.56	11.95±1.58
XEROX-6	9.59±3.07	7.35±2.08	9.69±1.95	11.22±4.33	10.92±3.76	11.43±3.95
ZEBRATOXIC	5.21	4.17	4.17	3.12	5.21	4.17
ZUBUD	8.70	6.96	8.70	8.70	7.83	7.83

TABLE XI
INFLUENCE OF THE GLOBAL FEATURE REPRESENTATION FOR ET-FL: TERMINAL FREQUENCIES (TF) OR BINARY (TB), HIERARCHICAL FREQUENCY (HF) OR BINARY (HB). OTHER PARAMETERS ARE CONSTANT: $T = 10$, $k = sqrt(M)$, $n_{min} = 5000$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, BEST SUBWINDOW SIZE INTERVALS.

Datasets	TF	TB	HF	HB
ACROSOMES	2.90 ± 0.68	2.68 ± 0.71	3.30 ± 0.51	3.84 ± 0.73
AGEMAP-FAGING	4.19±1.77	13.24±2.81	4.10±1.42	12.38±2.95
ALL-IDB2	0.19±0.58	0.77±0.94	0.96±0.96	2.31±1.44
ALOI	23.03	95.04	23.20	99.90
ALOT	24.72±0.32	24.46±0.37	24.02±0.44	24.12±0.36
APOPTOSIS	3.71 ± 1.71	5.86 ± 3.41	3.57 ± 1.32	6.86 ± 3.25
BINUCLEATE	1.25±2.02	41.75±6.13	1.50±3.00	32.75±11.64
BIRDS	41.23±2.85	41.0±2.47	43.03±2.54	41.3±3.64
BUILDINGSAB	52.09±2.83	69.17±2.73	55.64±2.84	72.76±3.37
BREAST-CANCER	7.08±3.59	18.61±3.63	9.72±3.23	18.06±3.51
BUTTERFLIES	39.22±2.17	45.59±2.62	39.34±3.08	45.29±4.22
BUTTERFLIES-CLEAN	15.62±2.34	23.56±1.21	17.4±2.57	23.6±1.48
CALTECH-101	72.49	74.38	73.73	73.38
CALTECH-256	88.76±0.33	88.28±0.39	89.26±0.30	88.34±0.40
C.ELEGANS	26.15±3.36	28.72±3.16	25.90±3.65	30.09±2.78
C.ELEGANS-LIVE-DEAD	1.58 ± 2.41	6.32 ± 3.16	0.0 ± 0.0	7.37 ± 5.86
CHARS74K	75.34±1.50	75.18±1.53	77.95±1.71	76.24±1.19
CHO	2.31±1.24	30.62±5.90	3.54±3.37	23.85±5.60
CIFAR-10	53.24	55.61	54.14	60.29
COIL-100 nv1	23.35	87.66	24.25	83.61
CONVEX	8.96	8.15	9.16	8.95
ETH-80	16.43±5.69	14.95±7.77	19.41±6.79	20.58±5.43
EVENTS	35.17±2.41	37.35±2.24	36.94±1.18	38.79±1.59
FAMOUSLAND	69.25±3.04	73.15±2.18	72.05±2.33	75.7±1.49
FLOWERS17	33.06±2.08	33.68±2.28	36.15±1.82	33.79±1.82
FLOWERS102-CLEAN	58.01	58.82	59.63	60.26
GALAXYZOO	3.88±0.15	4.48±0.19	3.97 ± 0.18	4.30 ± 0.16
GTSRBCROP	28.19	29.99	28.50	28.43
HCC	11.79 ± 7.24	37.50 ± 14.08	15.00 ± 8.42	39.29 ± 11.63
HEP2-ICPR2012	57.22	61.99	79.29	63.22
HPA	3.14 ± 1.13	3.62 ± 1.33	4.00 ± 1.75	4.19 ± 1.22
INDOOR	84.12±0.45	84.47±0.75	84.39±0.39	85.01±1.06
IRMA2005	13.10	23.0	13.70	22.60
IRMA2006	22.8	36.90	22.20	37.5
KTH-TIPS	7.44±1.86	22.07±1.50	7.39±1.33	21.51±1.98
KTH-TIPS2	46.67±8.29	46.38±6.59	53.26±8.06	54.41±6.71
LANDUSE	21.48±1.38	25.67±1.82	19.62±2.03	24.79±2.11
LYMPHOMA	5.95±3.97	12.97±5.38	6.49±3.24	14.59±5.43
MMODALITY	23.32	24.31	23.36	25.15
MNIST	1.64	2.2	1.61	2.53
MNIST-12000	2.30	2.48	2.70	2.73
MNIST-ROTATION	12.65	14.57	13.46	16.45
MNIST-BIMG	22.10	22.00	23.04	24.03
MNIST-BRAND	9.15	9.55	11.04	11.95
MNIST-BIMG-ROT	57.31	59.36	59.67	63.27
MSTAR-S1	17.36	19.46	24.36	21.91
NATURALSCENES	40.04±1.11	43.63±1.00	41.51±0.77	44.90±0.77
NORB-UNIFORM	12.49	17.88	12.89	19.83
NORB-JITTCLUTT	31.68	57.25	31.73	50.09
OLIVA	29.74±1.43	33.84±0.94	30.50±1.07	34.60±0.88
ORL	1.35±1.10	6.2±2.01	2.15±2.00	8.25±1.01
OUTEX	1.04±0.28	1.97±0.61	1.50±0.66	2.80±0.43
PFID	72.73±3.12	75.66±2.54	73.17±2.58	75.25±4.60
POLLEN	3.42±0.37	3.79±1.07	3.88±0.33	4.32±0.73
PPMI24	85.79	85.92	86.42	85.50
PUBFIG83	79.45 ± 0.64	81.10 ± 1.04	79.71 ± 1.19	81.47 ± 1.57
RBC	32.95±1.82	32.39±1.73	33.79±0.85	32.93±0.76
RECT-BASIC	0.14	0.60	0.14	0.79
RECT-BIMG	22.16	24.81	22.94	23.80
RNAI	14.5±3.32	25.75±4.34	13.0±4.0	28.5±6.34
SEROUS	27.65±2.09	28.06±2.20	29.67±2.13	29.04±1.89
SHAPE1	15.4±2.69	23.6±4.8	17.8±2.6	15.9±4.83
SMEAR	6.26±1.91	9.45 ± 2.75	8.02 ± 2.83	8.24 ± 1.13
SOCCER	20.86±2.90	34.48±4.42	21.62±5.41	30.67±3.09
SOIL-24	20.83	69.58	32.08	69.58
SOIL-47	52.21	78.94	68.19	80.32
SPORTS	9.40±0.97	9.75±1.10	9.22±1.19	9.95±0.82
STOIC101	25.76	56.97	28.64	57.42
STOMATA	20.00 ± 12.73	23.64 ± 11.64	28.18 ± 8.58	20.91 ± 14.69
STONEFLY9	14.72	17.16	78.72	77.00
SUBCELLULAR	11.63±2.90	15.0±4.04	14.07±2.52	12.91±1.42
SUN397	95.07±0.09	95.42±0.15	95.26±0.21	95.29±0.11
SWELEAF	3.33±0.84	11.29±1.68	4.47±1.32	11.67±1.29
TERMINALBULB	54.87±2.61	56.63±1.17	54.62±1.81	57.19±1.20
TINYGRAZ03	55.84±3.27	54.34±2.05	58.32±3.83	56.46±3.87
TSG-60	4.08±1.51	75.5±2.92	3.33±1.18	70.08±3.32
UIUCTEX	12.32±1.53	25.78±1.89	13.34±1.05	25.30±1.40
WANG	12.0±0.83	13.0±1.43	23.98±1.22	23.66±0.77
XEROX-6	7.35±2.08	11.84±2.63	9.99±3.19	13.67±3.33
ZEBRATOXIC	4.17	7.29	8.17	7.29
ZUBUD	6.96	12.17	7.83	12.17

Datasets	1	5	10	20	40
ACROSOMES	5.91 ± 0.70	3.68 ± 0.88	2.90 ± 0.68	2.92 ± 0.69	2.72 ± 0.55
AGEMAP-FAGING	6.19±1.15	4.29±1.87	4.19±1.77	3.71±1.50	3.81±2.09
ALL-IDB2	1.15±1.76	0.96±1.29	0.19±0.58	0.58±1.23	0.96±1.29
ALOI	46.41	26.54	23.03	21.96	21.05
ALOT	39.57±0.63	27.76±0.30	24.72±0.32	22.54±0.34	21.28±0.19
APOPTOSIS	8.14 ± 2.31	2.57 ± 1.67	3.71 ± 1.71	4.43 ± 2.34	4.29 ± 2.21
BINUCLEATE	4.50±5.10	5.25±3.61	1.25±2.02		2.5±2.50
BIRDS	55.03±2.22	45.8±2.16	41.23±2.85	45.03±2.88	44.0±2.72
BREAST-CANCER	13.06±2.42	7.50±1.78	7.08±3.59	6.39±3.12	9.58±3.37
BUILDINGSAB	52.43±1.92	52.04±2.15	52.09±2.83	52.15±2.42	52.76±2.14
BUTTERFLIES	49.57±3.42	42.13±2.78	39.22±2.17	40.07±1.74	38.03±3.25
BUTTERFLIES-CLEAN	20.96±1.28	16.64±1.85	15.62±2.34	14.78±1.76	15.13±1.49
CALTECH-101	80.57	75.22	72.49	71.46	71.18
CALTECH-256	92.81±0.31	90.12±0.30	88.76±0.33	88.25±0.25	87.71±0.29
C.ELEGANS	30.17±3.12	25.47±4.50	26.15±3.36	28.55±3.92	27.69±4.34
C.ELEGANS-LIVE-DEAD	3.68 ± 3.37	1.05 ± 3.16	1.58 ± 2.41	2.11 ± 2.58	1.58 ± 2.41
CHARS74K	84.18±1.30	76.59±1.55	75.34±1.50	71.45±0.77	69.68±1.36
CHO	6.46±2.46	3.23±1.45	2.31±1.24	2.46±2.09	2.77±1.79
CIFAR-10	62.27	54.98	53.24	54.76	55.69
COIL-100 nv1	28.42	24.24	23.35	23.06	21.62
CONVEX	10.92	10.32	8.96	8.13	7.41
ETH-80	22.84±7.26	19.0±7.03	16.43±5.69	15.11±5.81	15.38±7.46
EVENTS	44.85±1.72	36.90±2.22	35.17±2.41	35.42±2.11	34.08±2.04
FAMOUSLAND	77.1±2.3	70.3±2.59	69.25±3.04	68.35±3.07	67.7±2.72
FLOWERS17	42.03±2.95	36.03±2.5	33.06±2.08	33.35±1.97	31.82±3.52
FLOWERS102-CLEAN	67.92	59.48	58.01	57.38	56.62
GALAXYZOO	5.73±0.41	4.18±0.11	3.88±0.15	3.70±0.10	3.65±0.13
GTSRBCROP	45.86	33.01	28.19	23.32	21.00
HCC	18.93 ± 10.17	11.79 ± 10.66	11.79 ± 7.24	16.79 ± 8.53	15.36 ± 8.83
HEP2-ICPR2012	58.17	58.86	57.22	58.04	57.90
HPA	5.52 ± 2.55	3.14 ± 0.96	3.14 ± 1.13	2.95 ± 1.97	2.57 ± 1.65
INDOOR	89.73±0.62	86.26±0.93	84.12±0.45	83.02±0.88	81.88±1.07
IRMA2005	21.5	14.7	13.1	12.0	11.8
IRMA2006	27.70	23.10	22.80	22.90	22.40
KTH-TIPS	12.22±1.73	7.95±1.20	7.44±1.86	7.46±1.36	7.24±1.34
KTH-TIPS2	53.44±7.87	50.27±7.91	46.67±8.29	53.29±8.60	47.25±9.36
LANDUSE	32.17±1.65	22.69±1.79	21.48±1.38	17.83±2.43	17.10±0.69
LYMPHOMA	10.81±4.19	6.22±4.84	5.95±3.97	4.05±3.02	6.93±3.20
MMODALITY	33.40	26.41	23.32	21.41	20.95
MNIST	6.87	2.26	1.64	1.26	1.06
MNIST-12000	7.32	3.18	2.30	2.20	1.94
MNIST-ROTATION	29.75	17.04	12.65	10.01	8.61
MNIST-BIMG	36.89	27.69	22.10	18.19	15.55
MNIST-BRAND	25.13	12.76	9.15	7.75	6.59
MNIST-BIMG-ROT		57.31	55.38	54.81	50.06
MSTAR-S1	27.43	24.70	17.36	18.84	16.62
NATURALSCENES	51.94±0.81	41.82±0.74	40.04±1.11	39.22±0.91	38.36±0.90
NORB-UNIFORM	26.93	15.37	12.49	10.49	8.77
NORB-JITTCLUTT	49.75	36.90	31.68	26.36	21.05
OLIVA	39.58±1.08	31.94±1.10	29.74±1.43	28.08±1.05	27.75±1.06
ORL	2.45±1.82	2.15±1.76	1.35±1.10	0.9±0.86	1.7±1.35
OUTEX	2.04±0.47	1.69±0.80	1.04±0.28	1.32±0.65	1.30±0.91
PFID	76.28±1.02	77.08±3.55	72.73±3.12	70.57±3.81	69.89±3.96
POLLEN	9.89±1.36	5.28±0.74	3.42±0.37	3.37±0.69	3.10±0.83
PPMI24	89.54	87.75	85.79	87.21	87.17
PUBFIG83	88.96 ± 0.94	83.01 ± 1.22	79.45 ± 0.64	77.02 ± 2.06	75.81 ± 1.36
RBC	47.16±1.67	37.17±1.14	32.95±1.82	31.65±1.25	29.14±1.03
RECT-BASIC	1.11	0.34	0.14	0.12	0.09
RECT-BIMG	22.94	21.59	22.16	20.79	17.25
RNAI	13.5±5.02	14.25±4.48	14.5±3.32	12.25±4.93	13.0±5.57
SEROUS	34.89±2.01	30.76±2.02	27.65±2.09	26.48±2.59	25.19±1.90
SHAPE1	21.9±3.33	16.2±3.82	15.4±2.69	13.6±3.61	16.3±4.03
SMEAR	12.42 ± 3.26	8.24 ± 2.96	6.26±1.91	6.15 ± 1.85	6.04 ± 1.65
SOCCER	25.81±2.57	20.95±1.28	20.86±2.90	19.52±2.60	17.90±3.58
SOIL-24	42.92	17.08	20.83	29.58	20.83
SOIL-47	65.64	56.17	52.21	55.53	55.53
SPORTS	15.74±1.52	10.38±0.59	9.40±0.97	8.48±0.97	7.71±0.90
STOIC101	35.00	27.88	25.76	25.76	26.67
STOMATA	20.91 ± 10.79	27.27 ± 9.96	20.00 ± 12.73	15.45 ± 8.18	24.55 ± 14.11
STONEFLY9	21.08	15.42	14.72	15.33	14.11
SUBCELLULAR	16.98±4.60	14.30±3.33	11.63±2.90	12.79±2.49	13.84±3.31
SUN397	96.21±0.12	95.76±0.13	95.07±0.09	94.26±0.19	93.09±0.16
SWELEAF	5.71±1.11	4.16±0.63	3.33±0.84	3.92±0.33	4.77±0.72
TERMINALBULB	59.07±2.48	54.57±1.24	54.87±2.61	53.04±1.93	53.79±2.09
TINYGRAZ03	66.99±4.39	58.94±4.64	57.17±4.53	54.69±5.05	53.89±3.72
TSG-60	7.08±4.12	4.25±0.95	4.08±1.51	4.33±1.22	
UIUCTEX	16.84±1.02	12.72±1.40	12.32±1.53	12.04±1.57	11.50±1.33
WANG	15.82±1.14	11.50±1.41	12.0±0.83	11.42±1.53	11.98±1.14
XEROX-6	14.29±3.16	10.61±2.20	7.35±2.08	10.61±3.33	11.22±3.53
ZEBRATOXIC	14.58	5.21	4.17	4.17	4.17
ZUBUD	8.70	8.70	6.96	6.96	8.70

TABLE XII

INFLUENCE OF THE NUMBER OF TREES FOR ET-FL. OTHER PARAMETERS ARE CONSTANT: $k = sqrt(M)$, $n_{min} = 5000$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, BEST SUBWINDOW SIZE INTERVALS, AND FREQUENCY TERMINAL IMAGE REPRESENTATION.

B. Results with simple optimizations

In Table XIII we report results using simple optimizations described in main text (Section 4.2). DIFFTWOPIXNEIGH is related to the optimization of Section 4.2.3 and means Extremely Randomized Trees are built with node tests that threshold the difference of one pixel with one of its 8 direct neighbor (choosen randomly). Filters (nb_{fil}) means we applied nb_{fil} filters (389 in grey images, for color images filters are applied on each RGB channel so a total of 1167 filtered images) and then described subwindows by $16 \times 16 \times nb_{fil}$ intensity values corresponding to pixel values of all these filtered images. In-depth analysis and extension of these optimizations will be subject of future study.

TABLE XIII
ERROR RATE RESULTS USING OPTIMIZATIONS DESCRIBED IN MAIN TEXT SECTION 4.2.

Datasets	Best	Best with optimizations	Description of optimizations
ACROSOMES	2.50±0.55	**1.48±0.49**	Align images horizontally, DIFFTWOPIXNEIGH, ET-FL
ALOT	21.28	**14.32**	10 millions training subwindows, ET-FL
CHARS74K	69.68 ± 1.36	**52.10±1.44**	TRGB, DIFFTWOPIXNEIGH, ET-FL, $T = 100$
CIFAR10	46.33	**25.69**	Filters (1167), T=750, k=1, ET-FL
GTSRB	18.73	**1.19**	10 million training subwindows, TRGB, DIFFTWOPIXNEIGH, $T = 100$, $k = 28$, $n_{min} = 10000$, ET-FL
HEP2	56.95	**43.5**	$8,6$ million training subwindows (15%-45% size intervals + rotation + mirroring), TRGB, $T = 80$, $n_{min} = 6000$, $k = 28$, DIFFTWOPIXNEIGH, ET-FL
IRMA-2005	11.3	**5.90**	Filters (190), DIFFTWOPIXNEIGH, 5 million training subwindows, $T = 100$, $n_{min} = 25000$, ET-FL
IRMA-2006	22.2	**14.80**	10 million training subwindows, DIFFTWOPIXNEIGH, $T = 100$, $n_{min} = 10000$, ET-FL
LANDUSE	16.74±1.53	**13.31±1.60**	10 million training subwindows, $T = 100$, $k = 1$, $n_{min} = 50000$, ET-FL
MMODALITY	20.95	**10.00**	ET-FL, Combination with bags of textual descriptions [68]
NATURALSCENES	38.36±0.90	**31.35 ± 0.41**	Filters (389), 150000 training subwindows, $T = 100$, $k = 28$, $nmin = 1000$, ET-FL
NORB-JITTCLUTT	18.57	**12.94**	10 million training subwindows, $T = 40$, $n_{min} = 100$, ET-FL
RBC	29.14	**19.31**	Subwindows described by raw pixels and WND-CHARM features (328), ET-DIC

C. Comparison with other methods

1) Comparison with Extremely Randomized Trees on global pixel image representation: In Table XIV we compare on several datasets (with fixed image sizes) our approach to the direct application of Extra-Trees without subwindow extraction, ie. where each image is represented by a single input vector encoding all its pixel values. These results show the benefits of using variants with random subwindow extraction as they clearly outperforms the global approach on datasets with variabilities such as translation (CONVEX, HPA, LANDUSE, OUTEX, RECT-BASIC), in-plane rotation (MNIST-ROTATION), and viewpoint changes (COIL-100). On MNIST-BIMG-ROT, increasing the number of trees with ET-DIC actually yields results comparable to ET-Global. Moreover, it has to be noted that through the optimization of ET-DIC and ET-FL parameters, one obtains very significant improvements for some datasets (e.g. 12.94% on NORB-JITTCLUTT with ET-FL, while optimizing decision trees parameters for ET-Global could not bring such improvments). Still, ET-Global is only slightly inferior on some datasets where optimal size intervals with ET-DIC and ET-FL are large (e.g. GALAXYZOO, CIFAR-10).

TABLE XIV
COMPARISON OF ET-DIC ($T = 10$, $K = \sqrt{M}$, $n_{min} = 1$, OPTIMAL SIZE INTERVALS), ET-FL ($T = 10$, $K = \sqrt{M}$, $n_{min} = 5000$), AND EXTREMELY RANDOMIZED TREES ($T = 500$, $K = sqrt$, $n_{min} = 1$) ON GLOBAL IMAGE PIXELS ON SEVERAL DATASETS WITH FIXED-SIZE IMAGES.

Datasets	ET-DIC	ET-FL	ET-Global
APOPTOSIS	8.14 ± 2.86	3.71 ± 1.71	15.42±3.55
CIFAR-10	48.91	49.93	49.13
COIL-100	12.15	23.35	32.61
CONVEX	8.04	8.96	19.67
GALAXYZOO	4.22±0.17	3.88±0.15	4.23±0.11
LANDUSE	29.21±2.57	21.48±1.38	47.12±3.67
HPA	6.19 ± 1.30	3.14 ± 1.13	16.29±2.90
MNIST-12000	2.35	1.89	4.36
MNIST-ROTATION	9.78	12.65	14.48
MNIST-BIMG-ROT	61.60	57.31	55.11
NORB-UNIFORM	5.24	12.49	8.47
NORB-JITTCLUTT	43.57	31.68	38.90
ORL	1.1±1.24	1.35±1.10	1.25±1.68
OUTEX	2.01±0.72	1.04±0.28	65.05
RECT-BASIC	0.01	0.14	4.59
RECT-BIMG	15.78	22.16	20.74
STOMATA	16.36 ± 12.06	20.00 ± 12.73	28.18±14.35

2) Comparison with baseline methods: In Table XV we compared our results with previously published baselines on 24 datasets for which baseline results were provided once the dataset was first released. These baselines are either global approaches using various classifiers on downscaled image versions, or approaches using orientations filters (GIST [39] and V1S [69]), or methods based on color histograms.

3) Detailed comparisons: We first provide result comparisons in our preferred field of application (bioimaging data). Then we propose comparisons on a subset of other datasets. Overall these comparisons suggest to consider our approach as a first try on new datasets as it performs better than many published methods, although obviously not reaching state-of-the-art performances on every dataset.

a) Detailed comparisons on bioimaging datasets: On ACROSOMES, we perform better (1.48%) than a neural network classifier using 20 texture features [1] that yields 3.93% error rate. On APOPTOSIS, we obtain 2.57 ± 1.67 while [6] yields 1.6% error rate by extracting 295 features, then follow an iterative process (called CARTA) that involves a human user in order to apply a genetic algorithm for feature selection, and finally use a SVM classifier. On STOMATA, [6] states that the original CARTA features were not sufficient so they included 130 additional rotation-invariant features to better separate the two classes into different cluster regions, but no final classification rate is provided. On the HPA dataset, we obtain 2.38% error rate while [28] evaluated 6 approaches and obtain between 33.40% (using GLCM features and a SVM classifier) downto 10% (using hundreds of features from CHARM [75] combined with PCA and LDA). On SMEAR we perform better than an nearest-neighbor classifier using 20 nuclear/cytoplasmic features but worse than an optimized approach using genetic algorithm to perform feature selection before applying the nearest neighbor classifier [48]. On HCC, we obtain 10.71% error rate (???) while a SVM classifier using SIFT features yields 63.93% error rate, but our approach is not as good than the best variant of [26] that uses wavelet-based covariance descriptors (published results vary from 22.1% downto 0.7%). On BINUCLEATE, C.ELEGANS, CHO, RNAi, SUBCELLULAR, TERMINALBULB, and LYMPHOMA, our ET-FL results using similar test protocols are on par with the computationaly intensive WND-CHARM method [75] which computes a large set (> 1000) of features and compounds of image transforms. On AGEMAP dataset we perform much better (3.62% error rate) than WND-CHARM (51%). Additional experiments (not included) also show excellent results with our method on the two other liver gender datasets (with caloric restriction diet and ad-libitum diet) of the IICBU 2008 benchmark. During the IRMA-2005 challenge, 41 runs were submitted with error rates ranging from 73.3% downto 12.6% [30], while we now obtain 5.90% using optimizations. On IRMA-2006, 27 runs were submitted with results ranging from 31.7% (using weighted k-nearest neighbor on rescaled images of 16×16 pixels) downto 16.2% (with sparse histogram of image patches and a maximum entropy classifier) [72], while we now obtain 14.80% using optimizations. During the HEP2 ICPR 2012 challenge, 28 recognition algorithms were submitted, error rate results varied from 81.5% downto 31.5% (using SVM with descriptors that encode spatial relations among adjacent LBP features) while we obtain 43.5% error rate with optimizations and rank 9th. On RBC, our optimized 19.31% error rate is better than expected human error rate ($> 20\%$) and than various approaches using gaussian mixture densities or nearest neighbor classifiers, but less good than the best variants of [45] (using combination of gaussian mixtures and LDA, or combination of nearest neighbors with several processing steps) where results vary from 31% error rate downto 15.3%. On MMODALITY ImageCLEF 2010 challenge, our best run using optimizations yielded 10% error rate and ranked 6th among 45 runs (ranging from 88% error rate downto 6%). On ZEBRATOXIC we obtain better results than a phenotype recognition model that implements various intensity and texture descriptors and a support vector machine classifier, following the same two protocols [62]: internal cross-validation on the training set ($1.12\% \pm 1.09$ vs $2.6\% \pm 0.95$ error rate), and evaluation on an independant test set (3.12% vs 6.25% error rate).

b) Detailed comparisons on other datasets: On NORB-UNIFORM we obtain better results than a combination of Convolutional Nets and SVM [38], restricted Boltzman machines [76], and a huge convolutional network with five hidden layers and hundreds of maps per layer [77] although their result is better than ours when adding a translation and contrast-extracting layer. On NORB-JITTCLUTT, we obtain comparable results to Convolutional Network "LeNet7" but lower results than a combination of unsupevised Convolutional Nets and SVM. On MNIST variants, our results are comparable with 3 hidden layer Deep Belief Network and 3 hidden layer Stacked Autoassociator Network [18], but lesser than the best results

TABLE XV
COMPARISON WITH BASELINE METHODS ON SEVERAL DATASETS (CH: COLOR HISTOGRAMS, NNL1/2: NEAREST NEIGHBOR WITH L1/2 NORM, MLR: MULTINOMIAL LOGISTIC REGRESSION, SVM: SUPPORT VECTOR MACHINES).

Datasets	Our best error rate	Baseline error rate	Baseline method
CALTECH-101	69.74±	**42.45±0.74**	V1S [69]
CALTECH-256	87.71±	92.4±0.7 / **76**	NNL1 / V1S [69]
CIFAR-10	**25.69**	58.87 / 45.3	MLR on 32×32 / Gist [70]
COIL-100	**12.0**	$>$ 34.69	NNL1 [71]
CONVEX	**7.41**	19.13	SVM RBF [18]
INDOOR	81.88±1.07	79	Gist + RBF SVM [29]
IRMA2005	**5.90**	22.8	NNL2 on 32×32 [30]
IRMA2006	**14.80**	32.1	NNL2 on 32×32 [72]
LANDUSE	**13.31±1.60**	18.81	CH (HLS) [73]
MNIST	**1.06**	5.0	NNL2 on 28×28 [35]
MNIST-12000	**1.89**	3.03	SVM RBF on 28×28 [18]
MNIST-ROTATION	**8.14**	11.11	SVM RBF on 28×28 [18]
MNIST-BIMG	**15.55**	22.61	SVM RBF on 28×28 [18]
MNIST-BRAND	**6.59**	14.58	SVM RBF on 28×28 [18]
MNIST-BIMG-ROT	**50.06**	55.18	SVM RBF on 28×28 [18]
NORB-UNIFORM	**5.13**	18.4 / 11.6	NNL2 / SVM Gaussian on 32×32 [38]
NORB-JITTCLUTT	**12.94**	43.3	SVM Gaussian on 108×108 [38]
OLIVA	26.42	**17.3**	Gist [39]
PFID	**69.56**	88.7	CH [42]
RECT-BASIC	**0.01**	2.15	SVM RBF [18]
RECT-BIMG	**14.91**	24.04	SVM RBF [18]
SUN-397	93.54	94.5 / 91.8 / **83.7**	Tinyimages / CH / Gist [55]
WANG	**10.04**	47.2 / 16.9	NNL2 32×32 / CH [74]
ZUBUD	**3.48**	27.0 / 7.8	NNL2 32×32 / CH [74]

on the regular MNIST benchmark. On SPORTS, our ET-DIC and ET-FL results (less than 10% misclassification error rate and average class error rate) are much better than the bag-of-features model (about 55% average class error rate) used in [50] (using a combination of DoG, MSER, and affine-Harris detectors, SIFT descriptor, K-means clustering, and a linear SVM classifier) and also compared to their specific method using various types of features and selective hidden random field (about 35% average class error rate), a result probably explained by the fact that they discard too much information (color information is not used in their bag-of-feature model, and their specific method only computes a few color features for super-pixels). On SWELEAF, our results are slightly better than an extension of shape contexts with inner-distance and matching with dynamic programming [78] and much better than the original work using specifically designed features and neural networks [56]. We obtain better results than the local affine frame method [79] using only one training view on COIL-100, and we tied on SOIL-24 (perfect recognition). We are slightly inferior on ZUBUD (4 misclassified images while they obtain perfect classification when they represent local appearance by low-frequency coefficients of the discrete cosine transformation) on which several other methods perform worst (results range from 59% error rate by the first published method [80] downto 0%). On CHARS74KEIMG, we obtain better results than nearest neighbors and SVM approaches using SIFT, Spin images, MR8 descriptors, Shape Contexts and a commercial OCR system, results on par with nearest neighbors and Geometric Blur descriptors [14] but slighly inferior results than SVM and a combination of all descriptors. On PFID, we obtain significantly better results than a bag of SIFT features + SVM, and our results are on par with a specific method using pixel segmentation combined with statistics of pairwise local features [42]. On the recent GTSRB dataset, our results with optimizations (98.81%) are better than combining histograms of oriented with LDA (95.68%) and random forests (96.14%), slightly better than multi-scale convolutional neural networks (98.31%), comparable to human performance (98.84%), but slightly inferior to a committee of large convolutional neural networks (99.46%) [25]. On several other datasets (KTH-TIPS, UIUCTEX, XEROX6), we obtain with ET-FL results competitive with other approaches where various interest point detectors, invariant descriptors, and SVM kernels were combined and optimized [81], and for the two texture datasets we obtain better results than other methods specifically designed for texture classification cited in [81] although not reaching the state of the art. On EVENTS, ET-DIC and ET-FL are slightly better than the "scene only" model of [20] (ie. a dense grid sampling scheme, the use of rotation invariant SIFT descriptor to represent patches as well as some geometry/layout information, and K-Means clustering). On SHAPE1, our classification accuracy results are competitive with a localized bag-of-features model using shape features [47]. On LANDUSE, we obtain significantly better results (86.69 % recognition rate) than variants of spatial pyramid image representations involving co-occurence of visual words obtained by K-means clustering of SIFT features (77.38% recognition rate) [33], and than simpler HSL color histograms

(81.19%) [73]. On some other problems (such as BIRDS, BUTTERFLIES, CALTECH-101, CALTECH-256, CIFAR-10, EVENTS, FLOWERS102-CLEAN, INDOOR, KTH-TIPS2, NATURALSCENES, PUBFIG83, SUN397) our results using raw pixel values from original images are significantly inferior than more elaborated approaches, e.g. architectures that combine various steps of normalization, filtering and spatial pooling ([37], [82], [29], [55], [77]), or methods combining numerous image descriptors [83]. Our optimizations improve performances as discussed in main text but still do not yield state-of-the-art recognition rates.

D. Interpretability and dataset bias

As discussed in main text Section 5.2, the approach can be used to detect dataset biases.

1) Visualization of confidence maps: In Figure 8, we first illustrate the capability of the method to effectively detect specific, relevant patterns for a dataset of images of sports.

2) Dataset bias: Table XVI shows the absence of background bias for some datasets (e.g. RBC) while other datasets (e.g. ALL-IDB2, C.ELEGANS-LIVE-DEAD, and ORL) have strong bias as using only background information yields less than 10% error rates. In between, e.g. on WANG dataset, the overall error rate using small cropped regions is not good but inspection of confusion matrices reveal that two classes (Horses and Dinosaurs) are well classified using only background data, which is confirmed by Figure 9 that shows correctly classified images where correctly classified subwindows are related to artefacts/background features. On the dataset of swedish leaves, almost half of the images of a specific class are well classified by using only background regions (while only 6.67% should be well classified using random voting). On the dataset of C.elegans live/dead assays and the dataset of acute lymphoblastic leukemia, we obtain roughly 10% error rate using small background areas (see Figure 10 and 11 for examples of cropped background areas). While our results using whole images are significantly better than using only background area for all of these datasets, similar variabilities than in background regions might affect foreground regions and ease their classification. Overall, we suggest that acquisition protocols should be better described once a dataset is released so that they can also be peer-reviewed, and suggestions of appropriate preprocessing steps (normalization, illumination correction, ...) or specific evaluation protocols that take into account these potential biases should accompany their publication.

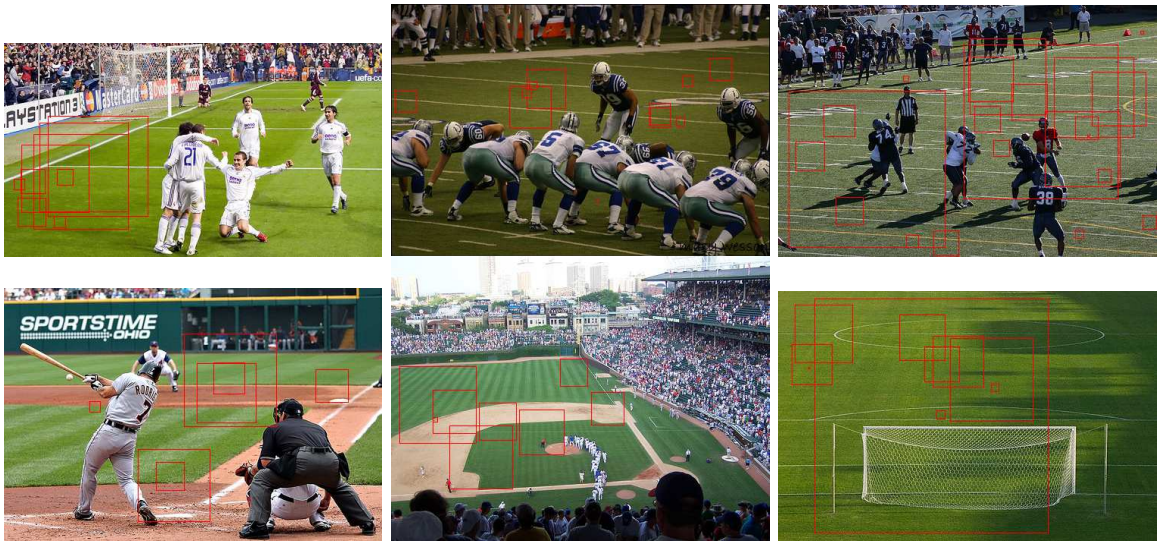

Fig. 8. Illustration of subwindows with highest probability estimates for the correct class on a dataset of sports images (classes correspond to different sports). Many of these subwindows implicitly capture local textured patterns with oriented lines, which was the purpose of the specific method developed in [50] that tries to explicitly segment the playing surfaces and derive line features.

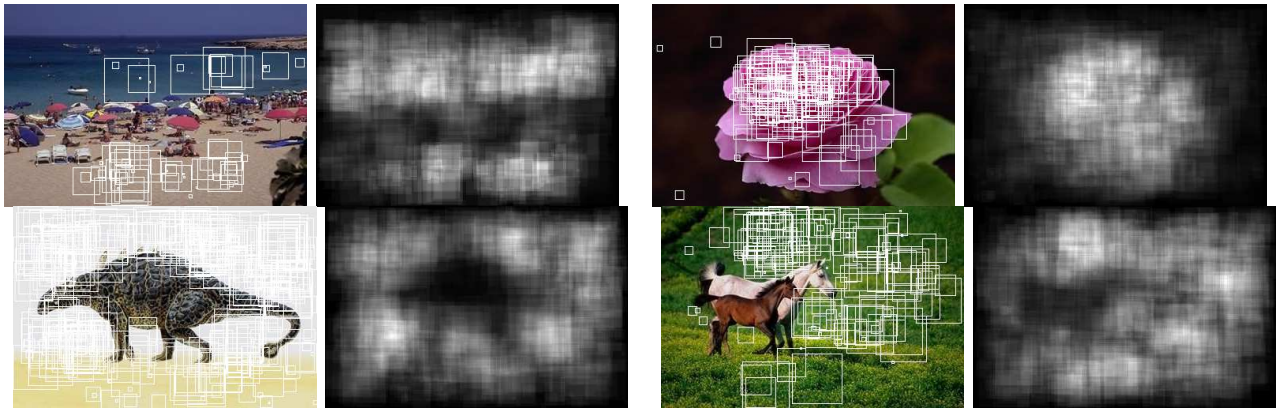

Fig. 9. Illustration of image confidence maps (lighter pixels correspond to pixels which obtain the highest probability estimates for the correct class). Although these regions might be considered as contextual information, these regions would not allow correct classificationn if the dataset is extended with new images examples with different acquisition conditions or new classes (e.g. other objects on white backgrounds, other animals on grass).

TABLE XVI
ASSESSING THE ROLE OF IMAGE BACKGROUNDS IN CLASSIFICATION PERFORMANCES. MAJORITY VOTING ER IS THE ERROR RATE ONE WOULD OBTAIN BY ALWAYS VOTING FOR THE MAJORITY CLASS. BACKGROUND ER IS THE ERROR RATE WE OBTAIN WITH ET-FL WORKING ONLY WITH SUBWINDOWS EXTRACTED IN THE CROPPED BACKGROUND AREA ($n_{min} = 5000$, $T = 10$, $k = sqrt(M)$, $N_{ls} = 1$ MILLION, 16×16 DESCRIPTORS, AND TERMINAL FREQUENCY IMAGE REPRESENTATION)).

Datasets	Background area	Subwindow sizes	Majority voting ER	Background ER
AGEMAP-FAGING	50×50 (top left)	0-25%	69.52	42.95±4.16
ALL-IDB2	50×50 (top left)	0-10%	50	9.62±3.21
C.ELEGANS-LIVE-DEAD	50×50 (top left)	0-10%	47.37	8.42 ± 7.14
HCC	50×50 (top left)	1×1	92.86	48.93± 13.03
ORL	20×20 (bottom right)	0-25%	97.5	9.35± 2.16
RBC	20×20 (top left)	0-10%	66.67	64.15±1.75
SEROUS	10×10 (top left)	0-90%	80.87	41.45±2.50
SPORTS	50×50 (top left)	0-50%	66.99	44.40±1.19
SWELEAF	100×100 (top left)	0-50%	93.33	80.61±1.33
TSG-60	50×50 (top left)	1×1	98.33	64.83 ± 5.02
WANG	20×20 (top left)	0-10%	90	56.0
ZUBUD	50×50 (top left)	0-100%	99.50	70.43
ZEBRATOXIC	50×50 (top left)	0-75%	54.17	41.67

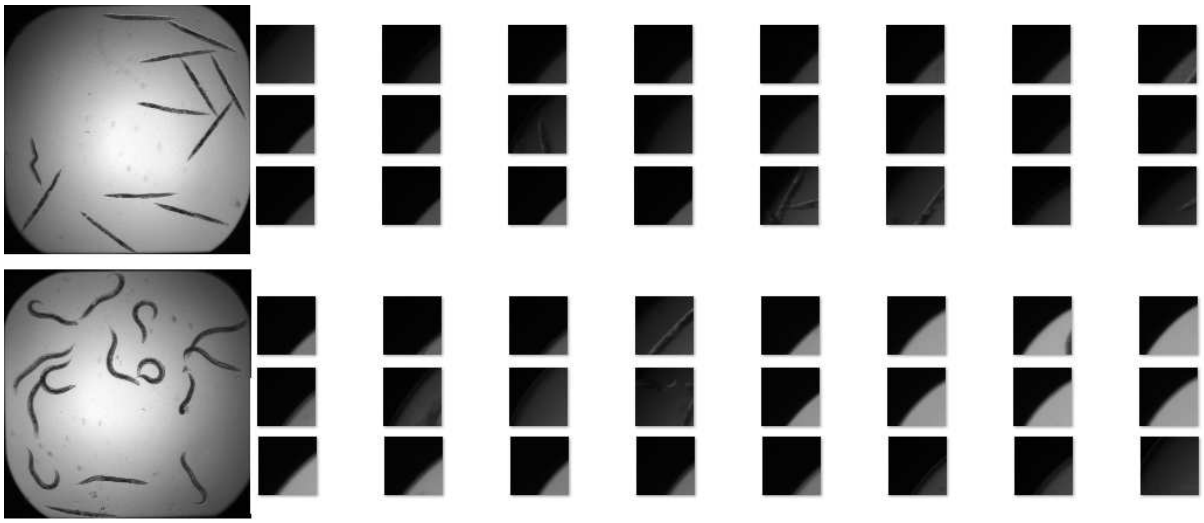

Fig. 10. Illustration of illumination bias in unprocessed images from the dataset describing C.elegans live/dead assays [13]. The two large images (left) are two images of a well from each class. Small images are cropped subimages (top left 50×50 corner) from 24 images for each class. Many samples of the top class have darker background than samples of the bottom class. Extracting individual pixels in these cropped areas and classifying them using the ET-FL variant yields less than 10% error rate.

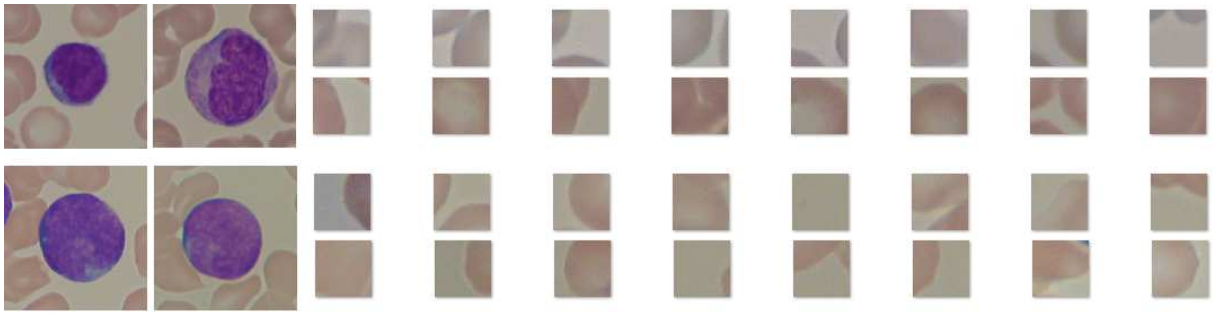

Fig. 11. Illustration of illumination/saturation bias in unprocessed images from the dataset describing normal and lymphoblast cells [3]. The large images (left) are two images from each class. Small images are cropped subimages (top left 50×50 corner) from 16 images for each class. Although differences are not visually striking (slight saturation effects), extracting individual pixels in these cropped areas and classifying them using the ET-FL variant yields less than 10% error rate.

REFERENCES

[1] González-Castro V, Aláiz-Rodríguez R, Alegre E: **Class Distribution Estimation based on the Hellinger Distance**. *Information Sciences* 2012.

[2] L S, T M, N O, M ED, G GI: **IICBU 2008 - A Benchmark Suite for Biological Imaging**. In *3rd Workshop on Bio-Image Informatics: Biological Imaging, Computer Vision and Data Mining* 2008.

[3] Labati RD, Piuri V, Scotti F: **ALL-IDB: the Acute Lymphoblastic Leukemia Image Database for Image Processing**. In *IEEE International Conference on Image Processing (ICIP)* 2011.

[4] Geusebroek JM, Burghouts GJ, Smeulders AWM: **The Amsterdam Library of Object Images**. *Int. J. Comput. Vis.* 2005, **61**:103–112.

[5] Burghouts GJ, Geusebroek JM: **Material-specific adaptation of color invariant features**. *Pattern Recognition Letters* 2009, **30**:306–313.

[6] Kutsuna Na: **Active learning framework with iterative clustering for bioimage classification**. *Nature Communications* 2012, **3**(1032).

[7] Lazebnik S, Schmid C, Ponce J: **A Maximum Entropy Framework for Part-Based Texture and Object Recognition**. In *International Conference on Computer Vision* 2005[http://lear.inrialpes.fr/pubs/2005/LSP05a]. [To appear.].

[8] Brook A, El-Yaniv R, Isler E, Kimmel R, Meir R, Peleg D: **Breast Cancer Diagnosis From Biopsy Images Using Generic Features and SVMs**. Tech. Rep. CS-2008-07, Technion, Isreal 2008.

[9] Zhang W, Kosecka J: **Hierarchical building recognition**. *Image and Vision Computing* 2007, :704–716.

[10] Lazebnik S, Schmid C, Ponce J: **Semi-local Affine Parts for Object Recognition**. In *British Machine Vision Conference* 2004[http://lear.inrialpes.fr/pubs/2004/LSP04].

[11] Fei-Fei L, Fergus R, Perona P: **Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories**. In *IEEE. Workshop on Generative-Model Based Vision* 2004.

[12] Griffin G, Holub A, Perona P: **Caltech-256 Object Category Dataset**. Tech. Rep. 7694, California Institute of Technology 2007, [http://authors.library.caltech.edu/7694].

[13] Moy TI, Conery AL, Larkins-Ford J, Wu G, Mazitschek R, Casadei G, Lewis K, Carpenter AE, Ausubel FM: **High-Throughput Screen for Novel Antimicrobials using a Whole Animal Infection Model**. *ACS Chemical Biology* 2009, **4**(7):527–533.

[14] de Campos TE, Babu BR, Varma M: **Character recognition in natural images**. In *Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal* 2009.

[15] Boland MV, Markey MK, Murphy RF: **Automated Recognition of Patterns Characteristic of Subcellular Structures in Fluorescence Microscopy Images**. *Cytometry* 1998, **33**:366–375.

[16] Krizhevsky A: **Learning Multiple Layers of Features from Tiny Images**. Tech. rep., University of Toronto, Department of Computer Science 2009.

[17] Nene S, Nayar S, Murase H: **Columbia Object Image Library: COIL-100**. Tech. Rep. CUCS-006-96, Department of Computer Science, Columbia University 1996.

[18] Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y: **An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation**. In *Proceedings of the 24th international conference on Machine learning (ICML)* 2007:473–480.

[19] Leibe B, Schiele B: **Analyzing Appearance and Contour Based Methods for Object Categorization**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, Madison, WI 2003.

[20] Li LJ, Fei-Fei L: **What, where and who? Classifying event by scene and object recognition**. In *Proc. IEEE Intern. Conf. in Computer Vision (ICCV)* 2008.

[21] Yeh T, Lee J, Darrell T: **Adaptive Vocabulary Forests br Dynamic Indexing and Category Learning**. In *Proc. ICCV* 2007.

[22] Nilsback ME, Zisserman A: **A Visual Vocabulary for Flower Classification**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Volume 2* 2006:1447–1454.

[23] Nilsback ME, Zisserman A: **Automated flower classification over a large number of classes**. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing* 2008.

[24] Gauci A, Zarb Adami K, Abela J: **Machine Learning for Galaxy Morphology Classification**. *ArXiv e-prints* 2010.

[25] Stallkamp J, Schlipsing M, Salmen J, Igel C: **Man vs. Computer: Benchmarking Machine Learning Algorithms for Traföè¡c Sign Recognition**. *Neural Networks* 2012, **32**:323–332.

[26] Keskin F, Suhre A, Kose K, Ersahin T, Cetin A, Cetin-Atalay R: **Image Classification of Human Carcinoma Cells Using Complex Wavelet-Based Covariance Descriptors**. *PLoS ONE* 2013, **8**.

[27] Percannella G, Foggia P, Soda P: **Contest on HEp-2 Cells Classification**. *http://mivia.unisa.it/hep2contest/index.shtml*.

[28] Niwas I, Kårsnäs A, Uhlmann V, P P, Kampf C, Simonsson M, Wählby C, Strand R: **Automated classification of immunostaining patterns in breast tissue from the Human Protein Atlas**. In *Workshop Histopathology Image Analysis (HIMA): Image Computing in Digital Pathology* 2012.

[29] Quattoni A, ATorralba: **Recognizing Indoor Scenes**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2009.

[30] Deselaers T, Müller H, Clough P, Ney H, Lehmann TM: **The CLEF 2005 Automatic Medical Image Annotation Task**. *International Journal of Computer Vision* 2007, **74**:51–58.

[31] Hayman E, Caputo B, Fritz M, Eklundh JO: **On the Significance of Real-World Conditions for Material Classification**. In *Proceedings of the 8th European Conference on Computer Vision (ECCV)*. Edited by LNCS 2004:253–266.

[32] Caputo B, Hayman E, Mallikarjuna P: **Class-Specific Material Categorisation**. In *Proc. International Conference on Computer Vision* 2005:1597–1604.

[33] Yang Y, Newsam S: **Spatial Pyramid Co-occurrence for Image Classification**. In *Proc. International Conference on Computer Vision (ICCV)* 2011.

[34] Pauly O, Mateus D, Navab N: **ImageCLEF 2010 Working Notes on the Modality Classification Subtask**. Tech. rep., Technische Universitat Munchen 2010.

[35] LeCun Y, Bottou L, Bengio Y, Haffner P: **Gradient-based learning applied to document recognition**. *Proc. of the IEEE* 1998, **86**(11):2278–2324, [citeseer.nj.nec.com/article/lecun98gradientbased.html].

[36] **Moving and Stationary Target Acquisition and Recognition (MSTAR) Public Dataset**. *https://www.sdms.afrl.af.mil/datasets/mstar/*.

[37] Lazebnik S, Schmid C, Ponce J: **Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories**. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Volume 2* 2006:2169–2178.

[38] Huang FJ, LeCun Y: **Large-Scale Learning with SVM and Convolutional Nets for Generic Object Categorization**. In *Proc. Computer Vision and Pattern Recognition Conference (CVPR)* 2006.

[39] Oliva A, Torralba A: **Modeling the shape of the scene: a holistic representation of the spatial envelope**. *International Journal of Computer Vision* 2001, **42**(3):145–175.

[40] Samaria F, Harter A: **Parameterisation of a Stochastic Model for Human Face Identification**. In *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision* 1994.

[41] Ojala T, Maenpaa T, Pietikainen M, Viertola J, Kyllonen J, Huovinen S: **Outex, New framework for empirical evaluation of texture analysis algorithms**. In *Proc. 16th International Conference on Pattern Recognition (ICPR), Volume 1* 2002:701–706.

[42] Yang S, Chen M, Pomerleau D, Sukthankar R: **Food recognition using statistics of pairwise local features**. In *CVPR* 2010:2249–2256.

[43] Yao B, Fei-Fei L: **Grouplet: A Structured Image Representation for Recognizing Human and Object Interactions**. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, USA 2010.

[44] Pinto N, Ston Z, Zickler T, Cox DD: **Scaling-up Biologically-Inspired Computer Vision: A Case-Study on Facebook**. In *Proc. Workshop on Biologically Consistent Vision (in conjunction with CVPR)* 2011.

[45] Keysers D, Dahmen J, Ney H: **Invariant Classification of Red Blood Cells: A Comparison of Different Approaches.** In *Bildverarbeitung für die Medizin'01* 2001:367–371.

[46] Lezoray O, Elmoataz A, Cardot H: **A Color object recognition scheme: application to cellular sorting**. *Machine Vision and Applications* 2003, **14**:166–171.

[47] Stark M, Schiele B: **How Good are Local Features for Classes of Geometric Objects**. In *Proc. International Conference on Computer Vision* 2007.

[48] Yannis Marinakisa JJ Georgios Douniasb: **Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification**. *Computers in Biology and Medicine* 2009, **39**:69–78.

[49] van de Weijer J, Schmid C: **Coloring Local Feature Extraction**. In *European Conference on Computer Vision, Volume Part II*, Springer 2006:334–348, [http://lear.inrialpes.fr/pubs/2006/VS06].

[50] Jain V, Singhal A, Luo J: **Selective Hidden Random Fields: Exploiting Domain Specific Saliency for Event Classification**. In *Proc. CVPR* 2008.

[51] J Burianek JK A Ahmadyfard: **SOIL-47, The Surrey Object Image Library**. Tech. rep., Centre for Vision, Speach and Signal processing, Univerisity of Surrey 2000.

[52] Image Perception A, 2955) LLICU: **The STOIC101 (2007-05) Dataset**. Tech. rep., Institute for Infocomm Research 2007.

[53] Martinez-Munoz, Zhang G, Payet W, Todorovic N, Larios S, Yamamuro N, Lytle A, Moldenke D, Mortensen A, Paasch E, Shapiro R, L, Dietterich T: **Dictionary-Free Categorization of Very Similar Objects via Stacked Evidence Trees**. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)* 2009.

[54] Boland M, Murphy RF: **A Neural Network Classifier Capable of Recognizing the Patterns of all Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells**. *Bioinformatics* 2001, **17**:1213–1223.

[55] Xiao J, Hays J, Ehinger K, Oliva A, Torralba A: **SUN Database: Large-scale Scene Recognition from Abbey to Zoo**. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010)* 2010.

[56] Söderkvist OJO: **Computer Vision Classification of Leaves from Swedish Trees**. *Master's thesis*, Linköping University, SE-581 83 Linköping, Sweden 2001. [LiTH-ISY-EX-3132].

[57] Wendel A, Pinz A: **Scene Categorization from Tiny Images**. In *Annual Workshop of the Austrian Association for Pattern Recognition* 2007.

[58] **Tourist Sights Graz 60 dataset**. *http://dib.joanneum.at/cape/TSG-60/*.

[59] Lazebnik S, Schmid C, Ponce J: **A Sparse Texture Representation Using Local Affine Regions**. *IEEE. Trans. PAMI* 2005, **27**(8):1265–1278.

[60] Wang JZ, Li J, Wiederhold G: **SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries**. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 2001, **23**(9):947–963.

[61] Dance C, Willamowski J, Fan L, Bray C, Csurka G: **Visual categorization with bags of keypoints**. In *ECCV International Workshop on Statistical Learning in Computer Vision* 2004.

[62] Liu R, Lin S, Rallo R, Zhao Y, Damoiseaux R, Xia T, Lin S, Nel A, Cohen Y: **Automated Phenotype Recognition for Zebrafish Embryo Based In Vivo High Throughput Toxicity Screening of Engineered Nano-Materials**. *PLoS ONE* 2012, **7**(4).

[63] Shao H, Svoboda T, Van Gool L: **ZuBuD - Zurich Building Database for Image Based Recognition**. Tech. Rep. TR-260, Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland 2003.

[64] Chang CC, Lin CJ: **LIBSVM : a library for support vector machines**. Tech. rep., Computer Science and Information Engineering, National Taiwan University 2003.

[65] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ: **LIBLINEAR: A library for large linear classification**. *Journal of Machine Learning Research* 2008, **9**:1871–1874.

[66] Marée R, Stevens B, Rollus L, Rocks N, Moles-Lopez X, Salmon I, Cataldo D, Wehenkel L: **A rich internet application for remote visualization and collaborative annotation of digital slide images in histology and cytology**. *BMC Diagnostic Pathology* 2013, **8 S1**.

[67] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E: **Scikit-learn: Machine Learning in Python** . *Journal of Machine Learning Research* 2011, **12**:2825–2830.

[68] Marée R, Stern O, Geurts P: **Biomedical Imaging Modality Classification Using Bags of Visual and Textual Terms with Extremely Randomized Trees**. In *Working Notes, ImageCLEF LAB Notebook, Multilingual and Multimodal Information Access Evaluation (CLEF)*, CELCT, Springer 2010.

[69] Pinto N, Cox DD, DiCarlo JJ: **Why is Real-World Visual Object Recognition Hard?** *PLoS Comput Biol* 2008.

[70] Ranzato M, Krizhevsky A, Hinton GE: **Factored 3-Way Restricted Boltzmann Machines For Modeling Natural Images**. *Journal of Machine Learning Research - Proceedings Track* 2010, **9**:621–628.

[71] Yang MH, Ahuja N, Roth D: **View-Based 3D Object Recognition Using SNoW**. In *Proc. European Conference on Computer Vision* 2000[citeseer.nj.nec.com/280370.html].

[72] Müller H, Deselaers T, Deserno T, Clough P, Kim E, Hersh W: **Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks**. In *In: CLEF 2006 Proceedings. Lecture Notes in Computer Science (2007* 2006:595–608.

[73] Yang Y, Newsam S: **Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification**. In *Proc. 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 2010.

[74] Deselaers T, Keysers D, Ney H: **Features for Image Retrieval: An Experimental Comparison**. *Information Retrieval* 2008, **11**:77–107.

[75] Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg I: **WND-CHARM: Multi-purpose Image Classification Using Compound Transforms**. *Pattern Recognition Letters* 2008, **29**(11):1684–1693.

[76] Nair V, Hinton GE: **3-d object recognition with deep belief nets**. In *Advances in Neural Information Processing Systems 22* 2009.

[77] Ciresan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J: **Flexible, High Performance Convolutional Neural Networks for Image Classification**. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)* 2011.

[78] Haibin L, DWJacobs: **Shape Classification Using the Inner-Distance**. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2007, **29**(2):286–299.

[79] Obdržálek S, Matas J: **Object Recognition using Local Affine Frames on Distinguished Regions**. In *Electronic Proceedings of the 13th British Machine Vision Conference, University of Cardiff* 2002.

[80] Shao H, Svoboda T, Ferrari V, Tuytelaars T, Van Gool L: **Fast Indexing for Image Retrieval Based on Local Appearance with Re-ranking**. In *Proc. IEEE International Conference on Image Processing (ICIP)* 2003:737–749.

[81] Zhang J, Marszalek M, Lazebnik S, Schmid C: **Local features and kernels for classification of texture and object categories: a comprehensive study**. *International Journal of Computer Vision* 2007, **73**(2):213–238.

[82] Bosch A, Zisserman A, Munoz X: **Image Classification using Random Forests and Ferns**. In *Proc. ICCV* 2007.

[83] Gehler PV, Nowozin S: **On Feature Combination for Multiclass Object Classification**. In *IEEE International Conference on Computer Vision (ICCV)* 2009.