P. Bodesheim, A. Freytag, E. Rodner, and J. Denzler: Local Novelty Detection in Multi-class Recognition Problems. Proc. WACV 2015, pp. 813–820.

Local Novelty Detection in Multi-class Recognition Problems

Paul Bodesheim, Alexander Freytag, Erik Rodner, Joachim Denzler
Computer Vision Group, Friedrich Schiller University Jena, Germany
{firstname.lastname}@uni-jena.de | www.inf-cv.uni-jena.de

Abstract

In this paper, we propose using local learning for multi-class novelty detection, a framework that we call local novelty detection. *Estimating the novelty of a new sample is an extremely challenging task due to the large variability of known object categories. The features used to judge on the novelty are often very specific for the object in the image and therefore we argue that individual novelty models for each test sample are important. Similar to human experts, it seems intuitive to first look for the most related images thus filtering out unrelated data. Afterwards, the system focuses on discovering similarities and differences to those images only. Therefore, we claim that it is beneficial to solely consider training images most similar to a test sample when deciding about its novelty. Following the principle of local learning, for each test sample a local novelty detection model is learned and evaluated. Our local novelty score turns out to be a valuable indicator for deciding whether the sample belongs to a known category from the training set or to a new, unseen one. With our local novelty detection approach, we achieve state-of-the-art performance in multi-class novelty detection on two popular visual object recognition datasets, Caltech-256 and ImageNet. We further show that our framework: (i) can be successfully applied to unknown face detection using the Labeled-Faces-in-the-Wild dataset and (ii) outperforms recent work on attribute-based unfamiliar class detection in fine-grained recognition of bird species on the challenging CUB-200-2011 dataset.*

1. Introduction

Novelty detection is an important aspect for recognition systems in real-life applications. Usually, the learned model captures only a fixed number of different object categories, but due to uncontrolled environments it is also possible that objects of previously unseen categories occur. Those new objects should not be labeled by the system as being one of the known categories, as standard recognition systems would do if there is no novelty detection mechanism involved. Instead, we need to assign an additional label to these new objects indicating that they are different compared to all known categories in order to avoid classification mistakes. In a lifelong learning scenario, a human annotator could then assign labels to the unknown objects such that the knowledge database of the system can be increased using incremental learning techniques.

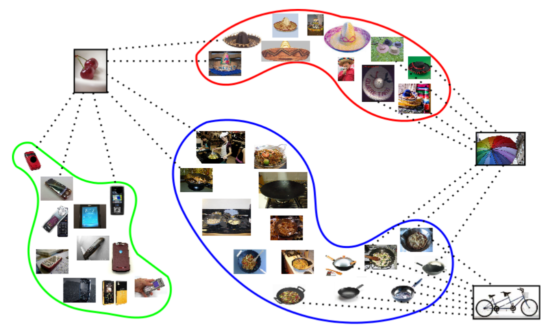


Figure 1. Our idea for novelty detection: instead of using all samples of all known categories (cellphone, wok, sombrero), we only use the k nearest neighbors (connected by dotted lines) of each test sample to learn a novelty detection model. This may result in a one-class classification scenario if all neighbors belong to the same class (as it is the case for the tandem in the lower right corner). We argue that a sample is novel if it is far away from its nearest neighbors in the training set, measured by a local novelty detection model that represents the boundaries of the known class distributions. Images of this figure are taken from ImageNet [8].

For many scenarios, it is often not possible to collect images for every object category that may occur in the application, even using large-scale databases like ImageNet [8]. If the recognition problem is at an instance level, *e.g.*, in face recognition, it is even more likely that an unknown instance appears. Imagine there is an automatic security check at the entrance of a building and only a fixed set of registered people is allowed to enter. Everybody could try to enter that building, but only the registered people should pass the test. Face verification only leads to decisions whether two images contain the same person [27], but in case of multiple images per known person, comparing the current test image with all the ones stored in the database is likely to take too much time for real-time processing.

Benefits of local learning In this paper, we propose using local learning for multi-class novelty detection. Local learning [6] is a well known technique in classification: for each test sample, a specific classification model is learned on-the-fly using its k nearest neighbors in the training set. This allows for learning complex nonlinear decision boundaries, which are necessary to model real-world image data with large diversity. Additionally, only a subset of the training data is used to build each classification model which prevents overfitting to the training set. Especially in large-scale scenarios, learning a classification model with hundreds of thousands of samples is time-consuming. Often, linear models are used for large-scale tasks, because the more powerful kernel approaches can not be exploited in the presence of huge training sets. With local learning, kernel approaches can also be applied to large-scale data, since only a subset is used to learn classification models.

Motivation for local novelty detection Typically, a sample is considered to be novel if it is far away from the training samples in feature space [5]. However, this implies that it is even far away from its closest samples. Thus, decisions do not need to take all known samples into account, but can rather be done in an exemplar-specific manner by learning a novelty detection model from the k nearest neighbors only (see Figure 1). A second argument for our local novelty detection approach is the high diversity of visual objects in images. The distribution of visual features in the feature space is very complex and local models have a higher complexity compared to their global counterparts.

Novelty detection is especially challenging for a fine-grained recognition task, in which different classes can only be distinguished by subtle details [36] and thus, even unknown classes only differ slightly from known ones. We argue that our local novelty detection approach is able to cope with those tiny differences by only considering a small neighborhood of the test sample and we show encouraging results on the CUB-200-2011 benchmark dataset for bird species recognition [37]. To the best of our knowledge, we are not aware of any other work dealing with local learning approaches for the task of novelty detection, especially in the context of visual object recognition.

Outline of this paper This paper is organized as follows. The next section summarizes related work for both local learning and novelty detection. In Sect. 3, we present our local learning approach for novelty detection. Experimental results for novel visual object detection in object recognition as well as for the specific applications of novel face detection and unknown bird species detection are presented in Sect. 4. Conclusions are drawn in the last section.

2. Related work

In this section, we review related work on local learning, multi-class novelty detection, and one-class classification.

2.1. Local learning

The idea of local learning has already been introduced in [6] for the task of handwritten digit recognition. Beside this application, the authors also study the influence of locality on the capacity of their classification model. In [40], a local learning approach is applied to image categorization. Their resulting framework, called SVM-KNN, consists of k-nearest-neighbor-search followed by multi-class SVM classification. Local learning is also used to improve classification models based on bag-of-visual-words features for facial expression recognition [18]. In recent work on relative attribute prediction [39], local ranking models are learned to accomplish fine-grained visual comparisons. Furthermore, local learning principles have been used in fine-grained recognition to learn exemplar-specific feature representations [12] and for nonparametric part transfer [13]. However, we have not found any work about local learning for multi-class novelty detection.

2.2. Multi-class novelty detection

Multi-class novelty detection is poorly studied so far. Given some training samples from a fixed number of known classes, the task is to decide for each test sample whether it belongs to one of the known classes (it does not matter to which class exactly) or to a new class that has not been observed in the training set. Most of the existing novelty detection approaches use the one-class classification paradigm and either treat multiple known classes as a single, artificial class [22], or learn a single one-class classification model for each known class independently [35]. In our previous work [5], we have shown that both strategies lead to poor performance compared to a simple one-vs-all SVM baseline. Furthermore, the kernel null space approach that we have presented in [5] based on kernel null Foley-Sammon transform (KNFST) outperforms the SVM baseline as well as all the tested approaches based on different one-class classification techniques. Therefore, we have decided to use KNFST in our local learning framework. Note that the code for KNFST is publicly available [5].

The goal of KNFST is to compute a low-dimensional embedding of the training data such that samples of the same class are mapped to the same point but samples from different classes are mapped to different points. This can be achieved by enforcing the within-class scatter of each projection direction $\boldsymbol{\varphi}$ to be zero ($\boldsymbol{\varphi}^\top \boldsymbol{S_w} \boldsymbol{\varphi} = 0$), while ensuring a positive between-class scatter ($\boldsymbol{\varphi}^\top \boldsymbol{S_b} \boldsymbol{\varphi} > 0$). Obviously, this maximizes the Fisher discriminant criterion:

$$\mathrm{F}(\boldsymbol{\varphi}) = \frac{\boldsymbol{\varphi}^\top \boldsymbol{S_b} \boldsymbol{\varphi}}{\boldsymbol{\varphi}^\top \boldsymbol{S_w} \boldsymbol{\varphi}} \rightarrow \infty \quad . \tag{1}$$

Using the kernel trick allows for estimating within-class and between-class scatter in some high-dimensional feature space, because the necessary computations depend only on inner products of the samples. After computing the so-called null projection directions $\boldsymbol{\varphi}$ and the corresponding class representations in the null space, the novelty score of a test sample is the smallest distance of the test sample to one of the class representations in the null space [5].

Furthermore, recent papers on open set recognition [19, 32] also deal with unknown categories that are not present in the training set. However, since novelty detection in general is often treated as a one-class problem, we additionally review related work on one-class classification in the next section for the sake of completeness. Additionally, there exist overview papers about different approaches towards novelty detection [24, 25, 26, 28].

2.3. One-class classification

The most popular techniques for one-class classification are one-class SVM (1–SVM) [33] and support vector data description (SVDD) [34]. Both are based on optimization problems: whereas 1–SVM separates the training data from the origin of the feature space using a hyperplane with maximum margin, SVDD encloses the training data with a hypersphere of minimum volume. The two methods can implicitly be applied in a nonlinear and high-dimensional feature space using kernel functions to compute inner products.

Another one-class approach is based on Gaussian process (GP) regression [20]. Assuming a zero-mean GP prior and setting the labels of all training samples to a non-zero constant (*e.g.*, 1), a novelty score can be inferred from the GP regression framework using either the predictive mean or the predictive variance of the estimated distribution.

Further one-class classification techniques range from prototype based approaches [1, 2], Parzen density estimation [4], convex hull algorithms [7], as well as one-class random forests [9] to kernel PCA [16] and kernel Fisher discriminants [30]. However, one-class methods are inferior to the multi-class KNFST approach in multi-class settings [5].

3. Local learning for novelty detection

Local learning has been introduced to tackle the trade-off between the capacity of a learning system and the amount of training data [6]. A local model is learned separately for each test sample using its k nearest neighbors in the training set, which is in contrast to the traditional approach of learning one global model for all test samples using the whole training set. In this paper, we are interested in applying local learning to novelty detection. We argue that novelty can be inferred locally for each test sample by only considering the most similar training samples. For a novel sample, those nearest training samples form the boundary to the known classes. The common assumption in novelty detection is that a test sample is novel if it is far away in feature space from all known categories observed so far. In this case, it is especially far away from its k nearest neighbors (see Figure 1). We exploit this observation using local learning.

Moreover, local novelty detection models are more suitable to manage the huge complexity of data distributions induced by features of real-world images. Especially for the task of novelty detection, the separation of known and unknown classes is highly nonlinear and specific for each class and each local region in the feature space. We therefore learn exemplar-specific novelty detection models to achieve a higher flexibility in the decision process particularly tailored to the sample currently under observation.

For each test sample, we first compute the nearest neighbors. Since the squared Euclidean distance of two samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be computed using inner products only:

$$\begin{aligned} \mathrm{d}_{Euclidean}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)^2 &= \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^\mathsf{T}\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right) \\ &= \boldsymbol{x}_i^\mathsf{T}\boldsymbol{x}_i + \boldsymbol{x}_j^\mathsf{T}\boldsymbol{x}_j - 2\cdot\boldsymbol{x}_i^\mathsf{T}\boldsymbol{x}_j \quad , \end{aligned} \tag{2}$$

we determine the k nearest neighbors of a test sample based on the squared Euclidean distance in some high-dimensional kernel feature space by computing inner products with a kernel function κ:

$$\mathrm{d}_{\kappa}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)^2 = \kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_i\right) + \kappa\left(\boldsymbol{x}_j, \boldsymbol{x}_j\right) - 2\cdot\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \quad . \tag{3}$$

Having determined the nearest neighbors, we learn a local novelty detection model using those neighbors only. Note that computing such a model is fast compared to learning a model using the whole training set with probably thousands of samples. The predicted novelty score of the local model is then used to rank the test sample.

In this paper, we use the multi-class novelty detection approach based on kernel null space methods that we have presented in our previous work [5]. We have shown that this approach is superior to other novelty detection approaches [5], which are usually formulated for one-class classification problems (*e.g.*, one-class SVM [33], one-class Gaussian processes [20]). This observation can also be verified by the experiments in this paper (see Sect. 4).

It is interesting to note that we compute a feature transformation leading to an *exemplar-specific subspace*, in our case an exemplar-specific null space. Here, only the most similar training samples of a test image contribute to the computations and we achieve a local transformation representing the neighborhood of the test sample under consideration. Moreover, only samples of those categories are involved that share features with the test image or are visually similar. Unrelated categories with visually dissimilar objects are excluded explicitly at an early stage reducing the number of visual concepts to compare with and the number of training samples to build the local model from. This means nothing but discarding the easy-to-differentiate categories first based on k-nearest-neighbor-search and decide about novelty using a more complex kernel-based novelty detection model learned from the training samples of the categories most similar to the test sample.

In our proposed local learning approach it is possible that all of the k nearest neighbors of a test sample have the same class label leading to a one-class classification scenario, especially if k is very small. Hence, we should have a novelty detection method applicable to both one-class and multi-class settings for our local learning approach. Although other combinations of one-class classifiers and multi-class novelty detectors are possible, we use kernel null space methods for both scenarios since a one-class formulation is also provided in [5]. Thus, our local models are either one-class null space models or multi-class null space models depending on the class distribution of the nearest neighbors. Note that the idea of using local models can also be applied to other novelty detection algorithms beside null space methods, which makes it possible to incorporate our framework in many other techniques and applications[1].

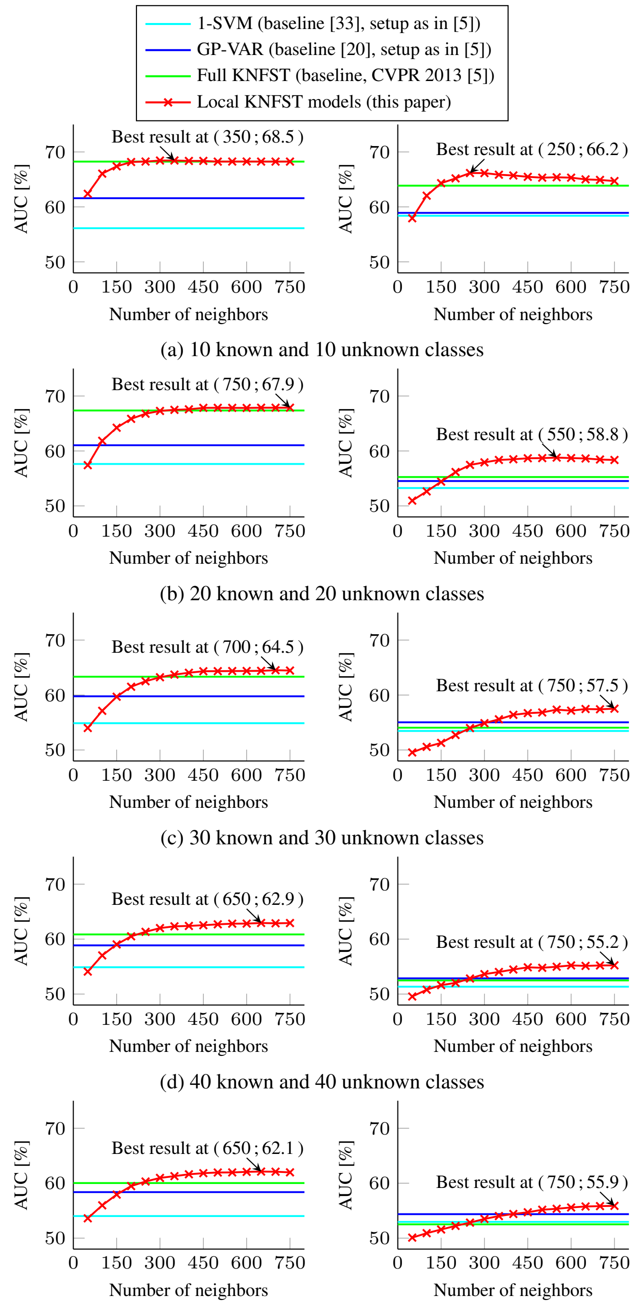
4. Experiments

To show the usefulness of local learning for novelty detection, we have conducted several experiments using different datasets. In Sect. 4.1, we evaluate our method on several class selections from the Caltech-256 [14] and the ImageNet [8] dataset. These experiments show the novelty detection performance for general object recognition tasks. Results in a face identification application using the Labeled-Faces-in-the-Wild dataset [17] and in unknown bird species detection using the CUB-200-2011 dataset [37] are presented in Sect. 4.2 and Sect. 4.3, respectively.

4.1. Multi-class novelty detection results

Features For both datasets, Caltech-256 and ImageNet, we have chosen the same publicly available feature representations of images as in [5]: bag-of-visual-words histograms of densely sampled SIFT descriptors. To measure similarities between histograms, we apply the histogram intersection kernel, a well-known and widely used kernel for histograms obtained from image data [3, 5, 11, 23, 29].

Setup In each experiment, we randomly select two subsets of classes. One subset represents the known classes used during learning, the other subset represents new classes whose samples should be detected as unknown during testing. Both subsets contain the same number of classes and we have tested several settings, ranging from 10 known and 10 unknown classes to 50 known and 50 unknown classes. We use the same evaluation metric as in [5] and results in terms of AUC [10] are averaged over 20 different random class selections for each setting. Following the experimental protocol of [5], samples of Caltech-256 categories that are used as known classes are split in two sets of the same size such that we have the same number of train and test samples per known category. All samples



(a) 10 known and 10 unknown classes

(b) 20 known and 20 unknown classes

(c) 30 known and 30 unknown classes

(d) 40 known and 40 unknown classes

(e) 50 known and 50 unknown classes

Figure 2. Results of our local learning approach (depending on the number of nearest neighbors) compared to the KNFST baseline as well as one-class classifier approaches (1–SVM and GP–VAR) on the Caltech-256 dataset (left column) and the ImageNet dataset (right column) for a varying number of known classes (a-e).

of Caltech-256 categories that are considered as unknown classes are used as new samples in the testing phase. From the ImageNet dataset, 100 images of known categories are selected to build a training set and the test set contains 50 images of each known and unknown category.

[1]Source code available: `www.inf-cv.uni-jena.de/novelty_detection`

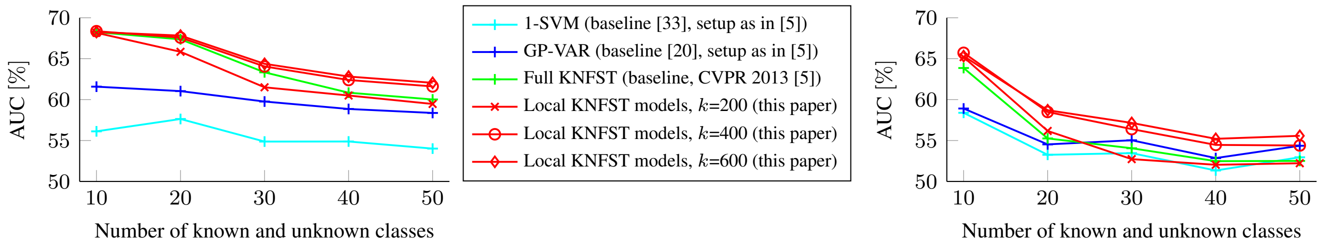


Figure 3. Novelty detection performance of selected models for the Caltech-256 dataset (left) and the ImageNet dataset (right).

Method comparison Detailed results for both datasets are shown in Figure 2. We compare our local learning approach using kernel null space methods (KNFST) with the baseline approach of learning a single full KNFST model [5] with all training samples. Additionally, we learn for each known class a one-class classifier to have further competitors as in [5]. For the one-class models, we apply one-class SVM (1–SVM) [33] and uncertainties from Gaussian processes (GP–VAR) [20]. For a small number of neighbors (*e.g.*, 50 or 100), the baseline methods achieve better performance than our local learning approach. This is due to the fact that with such a small number of neighbors, it is likely the case that only few samples of a small number of known classes are used to learn local novelty detection models. Thus, the learned model is not that accurate, since we do not expect a sufficient number of samples for each class within a small neighborhood of a test sample. Increasing the neighborhood leads to clear performance gains with results superior to those of the baseline methods proposed in [5]. Especially if we increase the number of known classes, the gap between the results of our local learning approach and the baselines increases as well. We clearly observe that our local learning model is most beneficial for an increasing number of known classes. It achieves state-of-the-art performance for various settings on both datasets.

Computation time One important question is how much computation time is needed to compute the novelty score of a test sample using our local learning approach. Compared to the baselines, we expect that our approach takes more time, because a local model is learned for each test sample. Naturally, computation times are directly coupled with the number of retrieved neighbors. However, the average testing time that is spent to compute the novelty score of a single sample is still below 1 second for neighborhoods of size $k \leq 700$ and up to $5{,}000$ training samples (tested on a 64-bit machine with AMD Opteron processor and 2.8 GHz). Note that the testing time involves computing similarities to the training data using the kernel function, sorting those values to obtain the k closest samples in the kernel feature space, as well as learning and evaluating a local novelty detection model estimated with only a small subset of the training data. Additional evaluations of computation times can be found in the supplemental material.

Evaluation of selected models We also summarize the novelty detection ability for a selection of the different models in Figure 3. Our local learning approach with neighborhoods of sizes 200, 400, and 600 is compared to the baselines. We observe an increasing performance for an increasing neighborhood of our local models in almost all settings and for both datasets. In our experiments with the ImageNet dataset, detection performances of all methods drop faster for an increasing number of known classes and training samples compared to the Caltech-256 dataset. Nevertheless, local learning even outperforms a global model trained on all data in a significant number of cases. Note that the performances of local models converges to the results of a global model for increasing neighborhood sizes. Interestingly, for more than ten known classes, the gap between the performances of models with 200 and 400 neighbors is larger than the improvement achieved by increasing the neighborhood size from 400 to 600. However, there is a trade-off between high performance and low computation time such that the number of neighbors should be selected wisely. Although it can be estimated using cross-validation and a leave-one-class-out strategy, we suggest to specify the size of the neighborhood based on the requirements of the application, *e.g.*, time constraints.

Further analysis Additionally, we observe from Figure 2 and Figure 3 that (i) the performance on the Caltech-256 dataset is better compared to the ImageNet dataset, which is mainly due to larger variances of object size and pose in the ImageNet dataset (see also Figure 5), and (ii) the task of multi-class novelty detection becomes harder if the number of known classes increases. For the Caltech-256 dataset, the performance decreases from about 68% AUC to about 60% AUC when the number of classes increases from 10 to 50. A similar trend can be observed by considering the more challenging ImageNet dataset, for which the performance decreases from about 65% AUC to about 55% AUC. This originates from the fact that dealing with more known classes increases the probability that a sample from an unknown class is accidentally classified as one of the known classes. This corresponds to analyzing the "openness" of a classification problem in open set recognition [19, 32]. Nevertheless, our local learning approach achieves state-of-the-art performance for all the tested numbers of known classes.

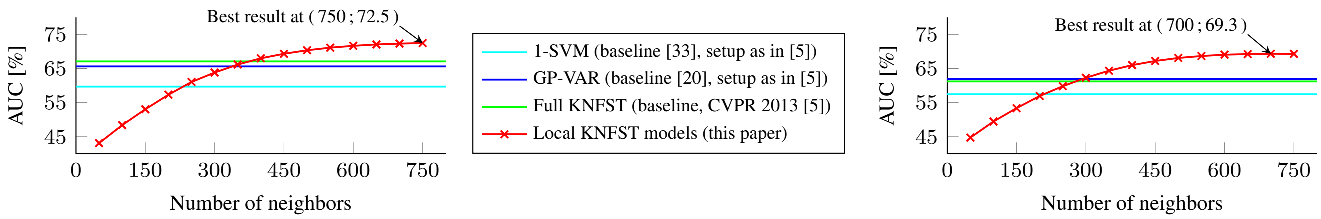


Figure 4. Unknown face detection performance on the Labeled-Faces-in-the-Wild dataset. In our experiments, either 158 categories with more than 10 samples (left) or 423 categories with more than 5 samples (right) are treated as known.

4.2. Application 1: unknown face detection

We further test our proposed method for the task of unknown face detection using the Labeled-Faces-in-the-Wild dataset [17]. This is a recognition task at an instance level, meaning that each person in the dataset is treated as a single category. In the first experiment, those 158 categories with at least 10 images are used as the known classes and random splits into 50% training data and 50% test data are computed. The remaining 5,591 categories, each with less than 10 images, are considered to be unknown. In a second experiment, we use the same setup but with a minimum of 5 images per known category leading to 423 known and 5,326 unknown categories. To represent a face, we use SIFT descriptors at three different scales and nine detected landmark positions provided by [15]. The final face descriptor is achieved via concatenation of the 27 SIFT descriptors followed by L1-normalization. Again, we use the histogram intersection kernel to compute pairwise similarities.

The outcome of our unknown face detection experiments is shown in Figure 4. Results have been averaged over 20 random splits for the samples of known categories. Again, we observe that the detection performance increases if we increase the size of the neighborhood of each test sample. Achieving a value of 68.0% AUC with 400 neighbors in the first experiment (left plot), we outperform the full KNFST baseline [5] (67.1% AUC) and the score further increases to 72.5% AUC using 750 neighbors. The absolute performance slightly decreases in the second setup (right plot), but the same trend as well as the gap between our approach and the baselines can be observed.

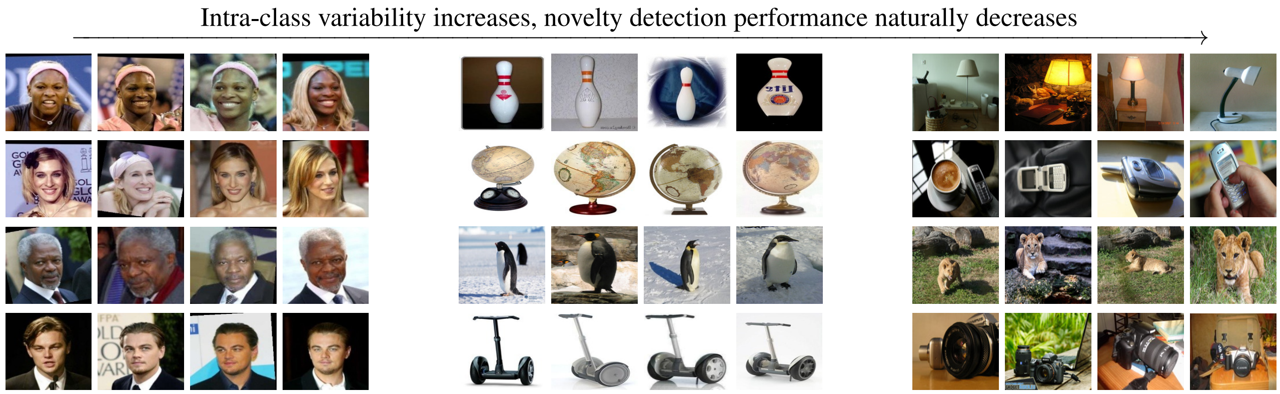
Note that for more than 150 categories, 72.5% AUC is really impressive compared to the decreasing performance trend on the Caltech-256 and the ImageNet dataset, when the number of known categories increases from 10 to 50 (see Sect. 4.1). Please also notice that the huge number of unknown face categories does not lead to a bias, since AUC is invariant to imbalanced class distributions [10]. The main difference is that in face recognition, features are highly tuned for this task, *e.g.*, using specific detectors for landmark positions, since knowledge about the shape and the structure of the objects (faces) can be exploited. For visual object recognition in general, identifying suitable landmarks and discriminating features is hard due to a higher intra-class variance compared to the instance-level task of face recognition (see Figure 5).

4.3. Application 2: unknown bird species detection

The CUB-200-2011 dataset [37] contains images of 200 different bird species and it is the benchmark for our experiments in a challenging fine-grained recognition task. We use the same experimental setup proposed by [36] and randomly select 100 from the 200 available classes as known ones. From the remaining 100 classes, 50 are randomly chosen and treated as unknown. As done in [36], we extract features from patches around ground truth part locations but restrict feature representations as in [13] to histograms over both quantized OpponentSIFT descriptors [31] and color name descriptors [38]. Compared to [36], we do not use geometric blur features and attribute descriptions.

Since we have distinct features for individual parts of the birds, we propose learning an individual local novelty detection model for each feature representation of each part and combine the obtained novelty scores to form a single novelty measure from those *part-specific novelty detection models*. We have tested several combination rules and found that simple averaging yields the best performances. Note that we only consider those parts of a bird, that are actually visible in the image. Therefore, some training images do not contribute to specific part models and not every part model necessarily contributes to the score of a specific test image.

We compare our local novelty detection approach using 100, 200, 300, and 400 neighbors with the full KNFST model as well as the methods presented in [36]. The results have been averaged over five random class selections and are summarized in Table 1. Our local novelty detection approach with 200 neighbors achieves the best result (59.7 % AUC) outperforming the full KNFST model [5] (51.9 %) and unfamiliar class detection (UCD) based on attributes proposed by [36] (57.2 %). Although each part-specific model alone would on average only yield values between 50.3 % and 57.9 % AUC, model combination leads to a performance boost showing that a single part alone is not sufficient to distinguish between known and unknown bird species. A visualization of the novelty scores obtained from the part-specific novelty detection models is given in Fig-



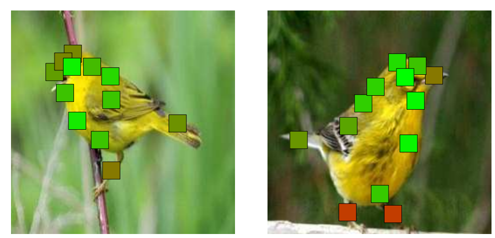
(a) Labeled-Faces-in-the-Wild [17] (b) Caltech-256 [14] (c) ImageNet [8]

Figure 5. Example images from some datasets used in our experiments. While the intra-class variability increases from (a) face recognition as a classification problem at an instance level to (b) general object category recognition with (c) increasing variability of object size and pose as well as changing background and context, the novelty detection performance naturally decreases in that direction.

Method	AUC
UCD (Wah and Belongie, CVPR 2013 [36])	57.2 %
One-class SVM (result presented in [36])	51.8 %
Two-class SVM (result presented in [36])	51.4 %
Full KNFST (Bodesheim *et al.*, CVPR 2013 [5])	51.9 %
Local KNFST models, k=100 (this paper)	57.3 %
Local KNFST models, k=200 (this paper)	**59.7** %
Local KNFST models, k=300 (this paper)	59.5 %
Local KNFST models, k=400 (this paper)	59.0 %

Table 1. Results of our novelty detection experiments in fine-grained recognition scenarios using the CUB-200-2011 dataset.

ure 6. It allows for identifying those parts that contribute most to the final decision and we clearly observe the distinction between known and unknown bird categories. Whereas only the legs of the known bird in the right image of Figure 6(a) deviate significantly from already observed ones, also parts at the head of the unknown bird in the right image of Figure 6(b) indicate novelty and make the difference.



(a) Birds of known categories

(b) Birds of unknown categories

Figure 6. Visualizing the novelty scores for different birds of the CUB-200-2011 dataset [37]. Individual estimates are obtained from part-specific novelty detection models of our local learning approach. Colored squares at part positions indicate corresponding novelty scores ranging from known (green) to novel (red).

5. Conclusions

In this paper, we have shown that the concept of local learning is beneficial for multi-class novelty detection tasks. Besides being intuitive, using only the nearest neighbors of a test sample in the training set to learn a local model yields state-of-the-art performance on various benchmark datasets. These results suggest that it is sufficient to compute novelty scores only with respect to the boundaries of the known class distributions represented by a model built from training samples closest to the test sample. Additionally, we have shown that our local novelty detection framework is suitable for unknown face detection and thus helpful for many surveillance applications that involve person identification and face recognition. Last but not least, our experiments on the birds dataset reveal the wide applicability of our local novelty detection approach even for the challenging fine-grained recognition task.

As stated in [5], novelty detection in large-scale scenarios with hundreds or thousands of categories is an important problem and we are far from having a satisfactory solution. However, our local learning approach is a first step towards large-scale novelty detection, since we are able to even apply kernel-based methods for an increasing number of known categories and control the amount of computation time by selecting an appropriate size of the neighborhood.

We encourage other researchers in the field of novelty detection to use our local novelty detection framework with the algorithms they developed, because different methods can easily be incorporated.

Future work will be focused on a combined treatment of feature extraction and novelty detection. So far, the developed novelty detection techniques can be applied to almost any type of feature representation as long as an appropriate kernel can be defined. However, a more vision-based approach exploiting all the visual information given in the image, *e.g.*, local structures and context, should be taken into account explicitly. Thus, it would be interesting to study whether deep convolutional approaches for image categorization [21] can be applied for visual novelty detection. Furthermore, applying local learning to open set recognition in order to compare with the recent advances in this field of research [19, 32] seems promising as well.

References

[1] F. Angiulli. Condensed nearest neighbor data domain description. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(10):1746–1758, 2007.

[2] F. Angiulli. Prototype-based domain description for one-class classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(6):1131–1144, 2012.

[3] A. Barla, F. Odone, and A. Verri. Histogram intersection kernel for image classification. In *Proc. ICIP*, pages 513–516. IEEE, 2003.

[4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[5] P. Bodesheim, A. Freytag, E. Rodner, M. Kemmler, and J. Denzler. Kernel null space methods for novelty detection. In *Proc. CVPR*, pages 3374–3381. IEEE, 2013.

[6] L. Bottou and V. Vapnik. Local learning algorithms. *Neural Computation*, 4(6):888–900, 1992.

[7] P. Casale, O. Pujol, and P. Radeva. Approximate convex hulls family for one-class classification. In *Proc. Int. Workshop on Multiple Classifier Syst.*, pages 106–115. Springer, 2011.

[8] 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*, pages 248–255. IEEE, 2009.

[9] C. Desir, S. Bernard, C. Petitjean, and L. Heutte. One class random forests. *Pattern Recognition*, 46(12):3490–3506, 2013.

[10] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Lett.*, 27(8):861–874, 2006.

[11] A. Freytag, E. Rodner, P. Bodesheim, and J. Denzler. Rapid uncertainty computation with Gaussian processes and histogram intersection kernels. In *Proc. ACCV*, pages 511–524. Springer, 2012.

[12] A. Freytag, E. Rodner, T. Darrell, and J. Denzler. Exemplar-specific patch features for fine-grained recognition. In *Proc. GCPR*, pages 144–156. Springer, 2014.

[13] C. Göring, E. Rodner, A. Freytag, and J. Denzler. Nonparametric part transfer for fine-grained recognition. In *Proc. CVPR*, pages 2489–2496. IEEE, 2014.

[14] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical report, California Inst. of Technology, 2007.

[15] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *Proc. ICCV*, pages 498–505. IEEE, 2009.

[16] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.

[17] 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. Technical Report 07-49, Univ. of Massachusetts, Amherst, Oct. 2007.

[18] R. T. Ionescu, M. Popescu, and C. Grozea. Local learning to improve bag of visual words model for facial expression recognition. In *ICML Workshop on Representation Learning*, 2013.

[19] L. P. Jain, W. J. Scheirer, and T. E. Boult. Multi-class open set recognition using probability of inclusion. In *Proc. ECCV*, pages 393–409. Springer, 2014.

[20] M. Kemmler, E. Rodner, and J. Denzler. One-class classification with Gaussian processes. In *Proc. ACCV*, pages 489–500. Springer, 2010.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proc. NIPS*, pages 1097–1105, 2012.

[22] T. Landgrebe, P. Paclik, D. M. J. Tax, and R. P. W. Duin. Optimising two-stage recognition systems. In *Proc. Int. Workshop on Multiple Classifier Syst.*, pages 206–215. Springer, 2005.

[23] S. Maji, A. C. Berg, and J. Malik. Classification using intersection kernel support vector machines is efficient. In *Proc. CVPR*, pages 1–8. IEEE, 2008.

[24] M. Markou and S. Singh. Novelty detection: A review–part 1: Statistical approaches. *Signal Process.*, 83(12):2481–2497, 2003.

[25] M. Markou and S. Singh. Novelty detection: A review–part 2: Neural network based approaches. *Signal Process.*, 83(12):2499–2521, 2003.

[26] S. Marsland. Novelty detection in learning systems. *Neural Computing Surveys*, 3:157–195, 2003.

[27] E. G. Ortiz and B. C. Becker. Face recognition for web-scale datasets. *Comput. Vision and Image Understanding*, 118:153–170, 2014.

[28] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Process.*, 99:215–249, 2014.

[29] E. Rodner, A. Freytag, P. Bodesheim, and J. Denzler. Large-scale Gaussian process classification with flexible adaptive histogram kernels. In *Proc. ECCV*, pages 85–98. Springer, 2012.

[30] V. Roth. Kernel Fisher discriminants for outlier detection. *Neural Computation*, 18(4):942–960, 2006.

[31] K. E. A. v. d. Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1582–1596, 2010.

[32] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult. Towards open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(7):1757–1772, 2013.

[33] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[34] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Mach. Learning*, 54(1):45–66, 2004.

[35] D. M. J. Tax and R. P. W. Duin. Growing a multi-class classifier with a reject option. *Pattern Recognition Lett.*, 29(10):1565–1570, 2008.

[36] C. Wah and S. Belongie. Attribute-based detection of unfamiliar classes with humans in the loop. In *Proc. CVPR*, pages 779–786. IEEE, 2013.

[37] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Inst. of Technology, 2011.

[38] J. v. d. Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE Trans. Image Process.*, 18(7):1512–1523, 2009.

[39] A. Yu and K. Grauman. Fine-grained visual comparisons with local learning. In *Proc. CVPR*, pages 192–199. IEEE, 2014.

[40] H. Zhang, A. C. Berg, M. Maire, and J. Malik. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In *Proc. CVPR*, pages 2126–2136. IEEE, 2006.