2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

One-Class Slab Support Vector Machine

Victor Fragoso[†] Walter Scheirer[‡]Joao Hespanha[∓] Matthew Turk[∓]
[†]West Virginia University [‡]University of Notre Dame [∓]University of California, Santa Barbara
victor.fragoso@mail.wvu.edu walter.scheirer@nd.edu {hespanha@ece, mturk@cs}.ucsb.edu

Abstract—This work introduces the one-class slab SVM (OC SSVM), a one-class classifier that aims at improving the per formance of the one-class SVM. The proposed strategy reduces the false positive rate and increases the accuracy of detecting instances from novel classes. To this end, it uses two parallel hyperplanes to learn the normal region of the decision scores of the target class. OCSSVM extends one-class SVM since it can scale and learn non-linear decision functions via kernel methods. The experiments on two publicly available datasets show that OCSSVM can consistently outperform the one-class SVM and perform comparable to or better than other state-of-the-art one class classifiers.

I. Introduction

Current recognition systems perform well when their train ing phase uses a vast amount of samples from all classes encountered at test time. However, these systems significantly decrease in performance when they face the open-set recogni tion problem [20]: recognition in the presence of samples from unknown or novel classes. This occurs even for already solved datasets (*e.g.*, the Letter dataset [10]) that are recontextualized as open-set recognition problems. The top of the Figure 1 illustrates the general open-set recognition problem.

Recent work has aimed at increasing the robustness of classifiers in this context [1], [19], [20]. However, these ap proaches assume knowledge of at least a few classes during the training phase. Unfortunately, many recognition systems only have a few samples from just the target class. For example, collecting images from the normal state of a retina is easier than collecting those from abnormal retinas [25].

One-class classifiers are useful in applications where col lecting samples from negative classes is challenging, but gath ering instances from a target class is easy. An ensemble of one class classifiers can solve the open-set recognition problem. This is because each one-class classifier can recognize samples of the class it was trained for and detect novel samples; see Figure 1 for an illustration of the ensemble of one-class classifiers. Unlike other solutions to the open-set recognition problem (*e.g.*, the 1-vs-Set SVM [20]), the ensemble offers parallelizable training and easy integration of new categories. These computational advantages follow from the independence of each classifier and allow the ensemble to scale well with the number of target classes.

However, the one-class classification problem is a challeng ing binary categorization task. This is because the classifier is trained with only positive examples from the target class, yet, it must be able to detect novel samples (negative class data). For instance, a one-class classifier trained to detect normal retinas must learn properties from them to recognize other images of normal and abnormal retinas. A vast amount of research has focused on tackling the challenges faced in the one-class classification problem. These strategies include statistical methods [6], [18], neural networks [2], [15], and kernel methods [13], [22], [23].

Despite the advancements, the performance of one-class classifiers falls short for open-set recognition problems. To improve the performance of one-class classifiers, we propose a new algorithm called the one-class slab SVM (OCSSVM), which reduces the rate of classifying instances from a novel class as positive (false positive rate) and increases the rate of detecting instances from a novel class (true negative rate). This work focuses on the one-class SVM classifier as a basis because it can scale well and can learn non-linear decision functions via kernel methods.

The one-class SVM (OCSVM) learns a hyperplane that keeps most of the instances of the target class on its positive side. However, instances from the negative class can also be on the positive side of this hyperplane. The OCSVM does not account for this case, which makes it prone to a high false

Fig. 1. The open-set recognition problem (top) challenges existing recognition systems. This is because classifiers can face instances from novel or unknown classes (images with dashed-frames). These novel classes cause failures during prediction time. Collecting instances from all the possible classes is a challenging task in many applications. For instance, collecting and labeling instances of all existing animals to avoid this problem is impractical. An ideal solution to this open-set recognition problem is an ensemble of one class classifiers (bottom). A single one-class classifier only requires instances of a target positive class to train (illustrated as circles). Such classifiers detect samples from the target classes and identify unknown instances. However, their performance needs improvement in order to solve the open-set recognition problem. The proposed approach improves the performance of the one-class SVM. It is a step towards the solution of the open-set recognition problem with an ensemble of one-class classifiers.

(SVM Score)

x_2

\- +

(w, ρ)

Projection

Hyperplane

\- +

x_1

SVM score

x_2 - +

\-

(w, ρ_2)

(w, ρ_1)

Hyperplanes - +

ρ_1 ρ_2

\-

x_1

SVM score

(a) (b) (c) (d)

Fig. 2. (a) The one-class SVM (OCSVM) learns a hyperplane that softly maximizes the margin between the origin and the data from the positive class. Its decision function projects the data onto the normal vector w to produce the SVM scores. (b) Subsequently, the decision function labels the samples as negative when the SVM scores fall below a threshold ρ, or labels them as positive otherwise. However, the one-class SVM does not account for outliers that can occur on the right tail of the SVM score density. In this case, a high rate of false positives can occur. (c) The proposed strategy considers learning two hyperplanes with the same normal vector but with different offsets. (d) These hyperplanes learn the "normal" region for the SVM scores. This region is called a slab.

positive rate. Unlike the OCSVM, the proposed OCSSVM approach encloses the normal region of the target class in feature space by using two parallel hyperplanes. When an instance falls inside the normal region or the slab created by the hyperplanes, the OCSSVM labels it as a sample from the target class, and negative otherwise. Figure 2 provides an overview of this new algorithm.

Using two parallel hyperplanes has been explored before in visual recognition problems. Cevikalp and Triggs [4] proposed a cascade of classifiers for object detection. Similarly, Scheirer *et al.* [20] proposed the 1-vs-Set SVM, where a greedy algorithm calculates the slab parameters after training a regular linear SVM. However, these methods are not strictly one-class classifiers since they use samples from known negative classes. Parallel hyperplanes have also been used by Giesen *et al.* [11] to compress a set of 3D points and by Glazer *et al.* [12] to estimate level sets from a high dimensional distribution. In contrast to these methods, the OCSSVM targets the open-set recognition problem directly and computes the optimal size of the slab automatically.

This work presents two experiments on two publicly available visual recognition datasets. This is because visual recognition systems encounter novel classes very frequently in natural scenes that contain both target and novel objects. The experiments evaluate the performance of the proposed approach and compare it with other state-of-the-art one-class classifiers. The experiments show that OCSSVM consistently outperforms the one-class SVM and performs comparable to or better than other one-class classifiers.

The OCSSVM represents a step towards the ideal robust recognition system based on an ensemble of one-class classifiers. The proposed OCSSVM can also improve the performance of other applications such as the identification

Their strategy consists of mapping the data to a feature space via kernel methods. Subsequently, it finds a hyperplane in this new feature space that maximizes the margin between the origin and the data.

To find this hyperplane, Schölkopf *et al.* proposed the following optimization problem:

$$\begin{aligned} \underset{w,\rho,\xi}{\text{minimize}} \quad & \frac{1}{2}\|w\|^2 + \frac{1}{\nu m}\sum_{i=1}^{m}\xi_i - \rho \\ \text{subject to} \quad & \langle w, \Phi(x_i)\rangle \geq \rho - \xi_i, \\ & \xi_i \geq 0,\ i = 1, \ldots, m, \end{aligned} \tag{1}$$

where m is the number of total training samples from the target class; ν is an upper-bound on the fraction of outliers and a lower bound on the fraction of support vectors (SV); x_i is the i-th training sample feature vector; w is the hyperplane normal vector; $\Phi()$ is a feature map; ξ are slack variables; and ρ is the offset (or threshold). Solvers for this problem compute the dot product $\langle\Phi(x_i), \Phi(x_j)\rangle$ via a kernel function $k(x_i, x_j) = \langle\Phi(x_i), \Phi(x_j)\rangle$.

Schölkopf *et al.* [22] proposed to solve the problem shown in Eq. (1) via its dual problem:

$$\begin{aligned} \underset{\alpha}{\text{minimize}} \quad & \frac{1}{2}\alpha^T K\alpha \\ \text{subject to} \quad & \|\alpha\|_1 = 1, \\ & 0 \leq \alpha_i \leq \frac{1}{\nu m},\ i = 1, \ldots, m, \end{aligned} \tag{2}$$

where K is the kernel matrix calculated using a kernel function, *i.e.*, $K_{ij} = k(x_i, x_j)$, and α are the dual variables. This optimization problem is a constrained quadratic program which is convex. Thus, solvers can use Newton like methods [3], [21] or a variant of the sequential-minimal optimization (SMO) technique [16].

of abnormal episodes in gas-turbines [7]; the detection of ab normal medical states from vital signs [14]; and the detection of impostor patterns in a biometric system [14]. *A. Brief Review of the One-class SVM*

Scholkopf ¨ *et al.* [22] proposed the one-class support vector machine (OCSVM) to detect novel or outlier samples. Their goal was to find a function that returns +1 in a "small" region capturing most of the target data points, and -1 elsewhere.

| {z }

The SVM decision function is calculated as follows:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{m} \alpha_i k(x_i, x) - \rho\right), \quad (3)$$

$\langle w, \Phi(x) \rangle$

where the offset ρ can be recovered from the support vectors that lie exactly on the hyperplane, *i.e.*, the training feature

vectors whose dual variables satisfy $0 < \alpha_i < \frac{1}{\nu m}$. In this work, the projection $\langle w, \Phi(x) \rangle$ of a sample x onto the normal vector w is called the SVM score.

B. Discussion

An interpretation of the solution ($w^?$, $\rho^?$) for the problem stated in Eq. (1) is a hyperplane that bounds the SVM scores from below; see the inequality constraints in Eq. (1). This interpretation also considers that the SVM score is a random variable. In this context, $\rho^?$is a threshold that discards outliers falling on the left tail of the SVM score density. Figures 2(a) and 2(b) illustrate this rationale.

However, the one-class SVM does not account for outliers that occur on the right tail of the SVM-score density. It needs to account for them to reduce false positives. Its decision rule considers these outliers as target samples yielding undesired false positives and decrease of performance.

The proposed strategy does account for these outliers. It learns two hyperplanes that tightly enclose the normal support of the SVM score density from the positive class. These hy perplanes bound the density from "below" and from "above." The proposed strategy considers samples falling in between these hyperplanes the "normal" state of the positive class SVM scores. It considers samples falling outside these hyperplanes outliers: novel or abnormal samples. The region in between the hyperplanes is called a "slab." In contrast with the SVM's default strategy, the proposed strategy assumes that samples from the negative class can have both negative and positive SVM scores; Figures 2(c) and 2(d) illustrate the proposed strategy.

II. One-Class Slab Support Vector Machine

This section describes the proposed one-class slab support vector machine. OCSSVM requires two hyperplanes to clas sify instances as negative (novel or abnormal samples) or posi tive (target class samples). Both hyperplanes are characterized

respectively; $\Phi()$ is the implicit feature map in the kernel function; and ν_1, ν_2, and ε are parameters. The parameter ε controls the contribution of the slack variables ξ and the offset ρ_2 to the objective function. The parameters ν_1 and ν_2 control the size of the slab.

This proposed optimization problem extends the formula tion introduced by Scholkopf ¨ *et al.* [22]. It adds two new linear inequality constraints per training sample, which are the constraints for the hyperplane f_2, and penalty terms in the objective function of the optimization problem shown in Eq. (1). This extension is mainly composed of linear terms and constraints. Consequently, it preserves convexity.

The offsets ρ_1 and ρ_2 have the following interpretation: they are thresholds that bound the SVM scores from the positive class (*i.e.*, $\langle w, \Phi(x_i) \rangle$) from below and above, respectively. This new interpretation motivates the names for the lower and upper hyperplanes mentioned earlier. The region in between these bounds is the "slab," and its size can be controlled by ν_1 and ν_2. The slack variables ξ and $\bar{\xi}$ allow the OCSSVM to exclude some SVM scores that deviate from the slab region: the normal region of the SVM score density from the positive class.

The decision function of the OCSSVM,

$$f(x) = \mathrm{sgn}\,\{(\langle w, \Phi(x) \rangle - \rho_1)(\rho_2 - \langle w, \Phi(x) \rangle)\}, \quad (5)$$

is positive when SVM scores fall inside the slab region, and negative otherwise.

Solving the primal problem (shown in Eq. (4)) is challeng ing – especially when a non-linear kernel function is used. However, the dual problem of several SVMs often yields a simpler-to-solve optimization problem. The dual problem for the OCSSVM (derived in supp. material) is

1

minimize $_{\alpha,\bar{\alpha}}$

$2(\alpha - \bar{\alpha})^T K (\alpha - \bar{\alpha})$ $v_1 m$, $\sum^m_i \alpha_i = 1$,

by the same normal vector w, and two offsets ρ_1 and ρ_2. The goal of OCSSVM is to find two hyperplanes that tightly enclose the region in feature space of the SVM-score density for the positive class. The positive side of each hyperplane

subject to $0 \leq \alpha_i \leq 1$

(6)

$v_2 m$, $\sum^m_i \bar{\alpha}_i = \varepsilon$, $i = 1, \ldots, m$,

$0 \leq \bar{\alpha}_i \leq \varepsilon$

coincides with the slab region and their negative side indicates the area where novel or abnormal samples occur; Figs. 2(c) and 2(d) illustrate the proposed configuration of the hyperplanes and decision process.

OCSSVM solves a convex optimization problem to find the hyperplane parameters (w, ρ_1, ρ_2). This problem is stated as follows:

$$\underset{\mathrm{w},\rho_1,\rho_2,\xi,\bar{\xi}}{\text{minimize}} \quad \frac{1}{2}\|\mathrm{w}\|^2 + \frac{1}{v_1 m}\sum_{i=1}^m \xi_i - \rho_1 + \frac{\varepsilon}{v_2 m}\sum_{i=1}^m \bar{\xi}_i + \varepsilon\rho_2$$

subject to $\langle \mathrm{w}, \Phi(\mathrm{x}_i)\rangle \geq \rho_1 - \xi_i$, $\xi_i \geq 0$,

$\langle \mathrm{w}, \Phi(\mathrm{x}_i)\rangle \leq \rho_2 + \bar{\xi}_i$, $\bar{\xi}_i \geq 0$, $i = 1, \ldots, m$, (4)

where (w, ρ_1) are the parameters for the "lower" hyperplane f_1; (w, ρ_2) are the parameters of the "upper" hyperplane f_2, ξ and $\bar{\xi}$ are slack variables for the lower and upper hyperplanes,

where and

where K is the kernel matrix; α_i and $\bar{\alpha}_i$ are the i-th entries for the dual vectors α and $\bar{\alpha}$, respectively; and $0 \leq v_1 \leq 1$, $0 \leq v_2 \leq 1$, and $0 \leq \varepsilon$ are parameters. This dual problem is a constrained quadratic program that can be solved with convex solvers. This work considers only positive definite kernels, *i.e.*, K is positive definite [21]. Therefore, $\varepsilon \neq 1$ must hold to avoid the trivial solution: $\alpha = \bar{\alpha}$.

The decision function can be re-written in terms of only the dual variables α, $\bar{\alpha}$ as follows:

$$f(\mathrm{x}) = \operatorname{sgn}\{(s_\mathrm{w} - \rho_1)(\rho_2 - s_\mathrm{w})\}, \quad (7)$$

$$s_\mathrm{w} = \langle \mathrm{w}, \Phi(\mathrm{x})\rangle = \sum_{i=1}^m (\alpha_i - \bar{\alpha}_i)\, k(\mathrm{x}, \mathrm{x}_i); \quad (8)$$

$$\rho_1 = \frac{1}{N_{\mathrm{SV}_1}} \sum_{i:0<\alpha_i<\frac{1}{v_1 m}}^{N_{\mathrm{SV}_1}} \sum_j^m (\alpha_j - \bar{\alpha}_j)\, k(\mathrm{x}_i, \mathrm{x}_j) \quad (9)$$

$$\rho_2 = \frac{1}{N_{\mathrm{SV}_2}} \sum_{i:0<\bar{\alpha}_i<\frac{\varepsilon}{v_2 m}}^{N_{\mathrm{SV}_2}} \sum_j^m (\alpha_j - \bar{\alpha}_j)\, k(\mathrm{x}_i, \mathrm{x}_j). \quad (10)$$

The SVM score s_w is obtained from Eq. (8) and re-writing dot products with the kernel function. On the other hand, the offsets require analysis from the KKT conditions (see supplemental material) to establish their relationship with the dual variables. The offset computation requires knowledge of the support vectors that lie exactly on the lower and upper hyperplanes. These support vectors are detected by evaluating if their dual variables satisfy $0 < \alpha_i < \frac{1}{v_1 m}$ and $0 < \bar{\alpha}_i < \frac{\varepsilon}{v_2 m}$ for the lower and upper hyperplane, respectively. Equations (9) and (10) require the number of support vectors N_{SV_1}, N_{SV_2} that exactly lie on the lower and upper hyperplanes, respectively. Moreover, it can be shown via the KKT conditions that if $\alpha_i > 0$, then $\bar{\alpha}_i = 0$, and that if $\bar{\alpha}_i > 0$, then $\alpha_i = 0$. This means that each

TABLE I

MEDIAN MATTHEWS CORRELATION COEFFICIENTS OVER THE 26 LETTERS. BOLD NUMBERS INDICATE THE HIGHEST SCORE FOR A KERNEL (ROW). OCSSVM CONSISTENTLY OUTPERFORMED OCSVM AND PERFORMED COMPARABLE TO OR BETTER THAN THE REMAINING ONE-CLASS CLASSIFIERS.

Kernel	KDE	KPCA	SVDD	OCSVM	OCSSVM
Linear	-	0.01	0.09	0.02	0.14
RBF	0.18	0.17	0.11	0.07	0.39
Intersection	-	0.18	0.01	0.04	0.26
Hellinger	-	0.01	0.02	0.02	0.13
χ^2	-	0.18	0.02	0.02	0.18

The experiments trained a one-class classifier for each class in the datasets. Recall that one-class classifiers only use positive samples for training. To evaluate the performance of the one-class classifiers, the experiments used the remaining hyperplane has its own set of support vectors; the reader is referred to the supplemental material for a more detailed analysis of the KKT conditions.

III. Experiments

This section presents two experiments (described in Sec tions III-A and III-B) that assess the performance of the proposed OCSSVM. These experiments use two different publicly available datasets: the letter dataset [10] and the PascalVOC 2012 [9] dataset.

We implemented a primal-dual interior point method solver in C++[1]to find the hyperplane parameters of the proposed OCSSVM. The experiments on the letter dataset were carried out on a MacBook Pro with 16BG of RAM and an Intel core i7 CPU. The experiments on the PascalVOC 2012 dataset were executed on a machine with 32GB of RAM and an Intel core i7 CPU.

The experiments compared the proposed approach to other state-of-the-art one-class classifiers: support vector data de scription (SVDD) [23], one-class kernel PCA (KPCA) [13], kernel density estimation (KDE), and the one-class support vector machine (OCSVM) [22] – the main baseline. The exper iments used the implementations from LibSVM [5] for SVDD and SVM; and a publicly available Matlab implementation we created for the one-class kernel PCA algorithm to apply to the letter dataset. However, the experiments used a C++ KPCA implementation (also developed in house) for the PascalVOC 2012 dataset, since the Matlab implementation struggled with the high dimensionality of the feature vectors and large number of samples in the dataset. For the multivariate kernel density estimation, we used Ihler's publicly available Matlab toolkit [2]. However, the KDE method did not run on the PascalVOC 2012 dataset due to the large volume of data. Thus, the experiments omit KDE results for that dataset.

[1]http://vfragoso.com

[2]Multivariate KDE: http://www.ics.uci.edu/˜ihler/code/kde.html

classes as negative samples (*i.e.*, novel class instances). The tested datasets are unbalanced in this setting since there are more instances from the negative class compared to the positive class. Note that common metrics such as precision, recall, and f1-measure are sensitive to unbalanced datasets. This is because they depend on the counts of true positives, false positives, and false negatives.

Fortunately, the Matthews correlation coefficient (MCC) [17] is known to be robust to unbalanced datasets. The MCC ranges between −1 and +1. A coefficient of +1 corresponds to perfect prediction, 0 corresponds to an equivalent performance of random classification, and −1 corresponds to a perfect disagreement between predictions and ground truth labels; see supp. material for more details about MCC.

The experiment used common kernels (*e.g.*, linear and radial basis function (RBF)) as well as efficient additive kernels [24] (*e.g.*, intersection, Hellinger, and χ^2). Among these kernels, only the RBF kernel requires setting a free parameter: γ. Also, the experiment used a Gaussian kernel for the KDE method. Its bandwidth was determined by the rule-of-thumb method, an automatic algorithm for kernel bandwidth estimation included in the used Matlab KDE toolbox. The experiments compare the KDE method only with the remaining one-class classifiers using an RBF kernel since the Gaussian kernel belongs to that family.

The experiments ran a grid-search over various kernel and classifier parameters, such as γ for the RBF kernel, C parameter for SVDD, v_1, v_2, v for the one-class SVMs, and number of components for KPCA, using a validation set for every class in every dataset; the reader is referred to the supplemental material where these parameters are shown.

To determine the ε parameters for training the proposed OCSSVM, the experiments used a toy dataset where samples from a bivariate Normal distribution were used. It was ob served that $\varepsilon =^{2}{}_{3}$produced good results; see supp. material for more details of this process.

A. Evaluation on Letter Dataset

This experiment aims at evaluating the performance of the OCSSVM. The tested dataset is letter [10], which contains 20,000 feature vectors of the 26 capital letters in the English

NA NA NA NA

KPCA SVDD OCSVM OCSSVM
KDE KPCA SVDD OCSVM OCSSVM

KPCA SVDD OCSVM OCSSVM
KDE KDE KPCA SVDD OCSVM OCSSVM KDE KDE KPCA SVDD OCSVM OCSSVM

(a) Linear (b) RBF (c) Hellinger (d) Intersection (e) χ^2

Student Version of MATLAB

Fig. 3. Matthews correlation coefficient on the letter dataset across different kernels (a-e); brighter indicates better performance. The proposed OCSSVM performed comparable or better than one-class kernel PCA (KPCA), kernel density estimation (KDE), support vector data description (SVDD), and one-class SVM (OCSVM). A comparison with the KDE method is only valid when using the RBF kernel.

alphabet. Each feature vector is a 16-dimensional vector capturing statistics of a single character. The dataset provides 16,000 samples for training and 4,000 for testing. The one class classification problem consists of training the classifier with instances of a single character (the positive class), and detecting instances of that character in the presence of novel classes – instances of the remaining 25 characters.

Figure 3 shows the results of this experiment. It visualizes the performance of the tested classifiers across classes for different kernels. Table I presents a performance summary per kernel and per method. The results shown in Figure 3 and Table I only include a comparison of the KDE method and the one-class classifiers with an RBF kernel since the KDE method uses a Gaussian kernel, which belongs to the RBF family. Because the experiment uses Matthews correlation coefficient (MCC), higher scores imply better performance. Thus, a con sistent bright vertical stripe in a visualization indicates good performance across all the classes in the dataset for a particular kernel. The figure shows that the proposed OCSSVM tends to have a consistent bright vertical stripe across different kernels and classes. This can be confirmed in Table I where OCSSVM achieves the highest median MCC for all of the kernels. The visualizations also show that the proposed OCSSVM outperformed the SVM method consistently. Comparing the OCSSVM and the SVM columns in Table I confirms the better performance of the proposed method. Table I also shows that OCSSVM performed comparable or better than one-class kernel PCA (KPCA), kernel density estimation (KDE), and support vector data description (SVDD).

B. Evaluation on PascalVOC 2012 Dataset

The goal of this experiment is to assess the performance of the OCSSVM on a more complex dataset: PascalVOC 2012 [9]. This dataset contains 20 different visual classes (objects) and provides about 1,000 samples per class. It has been used mainly for object detection. The experiment used HOG [8] features for every object class. To mimic novel classes that an object detector encounters, the experiment randomly picked 10,000 background regions for which HOG features were computed. The dimensionality of these features per class ranges from 2,304 to 36,864. This experiment used high-dimensional feature vectors and a large number of samples. Consequently, the kernel density estimation (KDE) MATLAB toolkit struggled and did not run properly on this dataset. Hence, the experiment omits the result for this method. The experiment trained one-class classifiers for each object using a 3-fold cross-validation procedure. The testing set for a fold was composed of object samples and all background features. Figure 4 shows the visualizations of the average Matthews correlation coefficients (MCC) for this experiment. In addition, Table II presents a summary of this experiment. Table II shows that OCSSVM tended to outperform the one class SVM across kernels. Moreover, it performed comparable to or better than one-class KPCA and SVDD across kernels. Figure 4 shows that the OCSSVM tended to outperform the SVM method across classes and kernels.

TABLE II

MEDIAN OF THE 3-FOLD MATTHEWS CORRELATION COEFFICIENTS OVER THE 20 CLASSES IN THE PASCALVOC 2012 DATASET PER KERNEL. BOLD NUMBERS INDICATE THE HIGHEST SCORE FOR A KERNEL (ROW). OCSSVM OUTPERFORMED THE OCSVM IN MOST OF THE CASES, WITH THE EXCEPTION OF THE RBF KERNEL CASE. IT PERFORMED COMPARABLE TO OR BETTER THAN ONE-CLASS KPCA AND SVDD.

Kernel	KPCA	SVDD	OCSVM	OCSSVM
Linear	0.02	0.09	0.01	0.07
RBF	0.05	0.07	0.14	0.09
Intersection	0.18	0.01	0.04	0.26
Hellinger	0.01	0.02	0.02	0.13
χ^2	0.18	0.02	0.02	0.18

IV. CONCLUSIONS AND FUTURE DIRECTIONS

This work presented the one-class slab support vector machine as a step towards the idealized one-class solution for open-set recognition. In contrast to the regular one-class SVM, which learns a single hyperplane for identifying target samples, instances from the positive class, the proposed clas sifier uses two parallel hyperplanes learned in feature space to enclose a portion of the target samples. However, each plane has an offset with respect to the origin that places them at different locations in feature space, creating a “slab.” The proposed approach to train the OCSSVM is a quadratic program (QP) that estimates the hyperplane normal vector and the two offsets.

The proposed OCSSVM showed consistent performance improvement over the regular one-class SVM on two different datasets: letter [10] and the PascalVOC 2012 [9]. The proposed strategy performed comparable or better than other state of

KPCA SVDD OCSVM **OCSSVM** KPCA SVDD OCSVM **OCSSVM** KPCA SVDD OCSVM **OCSSVM** KPCA SVDD OCSVM **OCSSVM** KPCA SVDD OCSVM **OCSSVM** (a) Linear (b) RBF (c) Hellinger (d) Intersection (e) 2

Fig. 4. Average of the 3-fold Matthews correlation coefficient scores per class; brighter indicates better performance. The proposed OCSSVM outperformed the SVM using efficient additive kernels (Hellinger, Intersection, and χ^2). It performed comparable or better than the one-class kernel PCA (KPCA), the support vector data description (SVDD), and the one-class SVM (OCSVM).

the art one-class classifiers, such as support vector data de scription [23], one-class kernel PCA [13], and kernel density estimation.

The approach used a Newton-based QP solver to train the OCSSVM. However, this solver is not efficient and a derivation of a sequential-minimal-optimization (SMO) [16] is planned for future work. The plan includes the adaptation of the SMO solver to deal with an extra inequality constraint that the QP of the OCSSVM includes.

Acknolwedgments: This material is based upon work supported by the National Science Foundation under Grant No. 1423676.

REFERENCES

[1] A. Bendale and T. E. Boult. Towards open world recognition. In

Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, June 2015.
[2] C. M. Bishop. Novelty detection and neural network validation. In *IEE Proc. Vision, Image and Signal Processing*, 1994.
[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
[4] H. Cevikalp and B. Triggs. Efficient object detection using cascades of nearest convex model classifiers. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.
[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.
[6] D. A. Clifton, S. Hugueny, and L. Tarassenko. Novelty detection with multivariate extreme value statistics. *Journal of signal processing systems*, 65(3):371–389, 2011.
[7] D. A. Clifton, L. Tarassenko, N. McGrogan, D. King, S. King, and P. Anuzis. Bayesian extreme value statistics for novelty detection in gas-turbine engines. In *Proc. of the IEEE Aerospace Conf.*, 2008.
[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition,*, 2005.
[9] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisser man. The pascal visual object classes (voc) challenge. *Intl. Journal of Computer Vision*, 88(2):303–338, 2010.
[10] P. W. Frey and D. J. Slate. Letter Recognition Using Holland-Style Adaptive Classifiers. *Machine Learning*, 6(2):161–182, 1991. [11] J. Giesen, S. Spalinger, and B. Scholkopf. Kernel methods for implicit ¨ surface modeling. In *Advances in Neural Information Processing Systems*, pages 1193–1200, 2004.
[12] A. Glazer, M. Lindenbaum, and S. Markovitch. q-ocsvm: A q-quantile estimator for high-dimensional distributions. In *Advances in Neural Information Processing Systems*, pages 503–511, 2013.
[13] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.
[14] H.-j. Lee and S. Cho. Retraining a Novelty Detector with Impostor Patterns for Keystroke Dynamics-Based Authentication. In *Advances in*

Biometrics, volume 3832 of *Lecture Notes in Computer Science*, pages 633–639. Springer Berlin Heidelberg, 2005.
[15] M. M. Moya and D. R. Hush. Network constraints and multi-objective optimization for one-class classification. *Neural Networks*, 9(3):463 – 474, 1996.
[16] J. Platt. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Technical Report MSR-TR-98-14, Microsoft Research, April 1998.
[17] D. M. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, 2011.
[18] G. Ritter and M. T. Gallegos. Outliers in statistical pattern recognition and an application to automatic chromosome classification. *Pattern Recognition Letters*, 18(6):525–539, 1997.
[19] W. J. Scheirer, L. P. Jain, and T. E. Boult. Probability Models for Open Set Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36, November 2014.
[20] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boult. Towards Open Set Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 36, July 2013.
[21] B. Scholkopf and A. J. Smola. ¨ *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001. [22] B. Scholkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. ¨ Platt. Support Vector Method for Novelty Detection. In *Advances in Neural Information Processing Systems 12*, pages 582–588. MIT Press, 2000.
[23] D. M. Tax and R. P. Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.
[24] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit

feature maps. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(3):480–492, 2012.
[25] T. Yamuna and S. Maheswari. Detection of abnormalities in retinal im ages. In *Intl. Conf. on Emerging Trends in Computing, Communication and Nanotechnology*, 2013.