LORD: Leveraging Open-Set Recognition with Unknown Data
– Supplementary Material –

Tobias Koch[*, †], Christian Riess[†], and Thomas Köhler[*]
[*]e.solutions GmbH, Erlangen, Germany
[†]Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
{tobias.koch, thomas.koehler}@esolutions.de, christian.riess@fau.de

1. Overview

This supplementary material is organized as follows:

- Section 2 provides some further details on the implementation of the strategies in the models.
- Section 3 presents the results from the Clustering-based Extreme Value Machine (C-EVM) and Probability of Inclusion Support Vector Machine (P_I-SVM) [3], exploiting genuine known unknown classes (KUCs) with the learning strategies.
- Section 4 shows visualizations of the decision boundaries on a toy dataset for all models and various mixup-to-known ratios. It also includes additional open-set recognition (OSR) measures for the models in the main work, the C-EVM, and the P_I-SVM, using mixups as KUC surrogates.
- Section 5 contains additional metrics for the models in the main work, the C-EVM, and the P_I-SVM, with constrained mixups.

2. Deployed strategies for open-set models – additional details

In this section, we outline how to deploy the strategies to 6 different OSR models from the following 4 categories.

Open-Set Nearest Neighbor (OSNN). The OSNN [4] exploits the ratio between the two nearest samples from distinct classes as confidence value. Let d_i and d_j be Euclidean distances between a query sample $\boldsymbol{x}$ and its two closest training samples, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, where $y_i \neq y_j$. The ratio is computed as $r = {}^{d_i}/_{d_j}$ with $d_i < d_j$. If $r \leq \delta$, query $\boldsymbol{x}$ is labeled as y_i, otherwise it is labeled as u.

Single Pseudo Label (SPL) includes all KUCs in distance ratio computation and class prediction. Multi Pseudo Label (MPL) treats each KUC as a separate class, impacting the reject option only. For Known *vs*. Rest (KvR), the distance ratio is set to the maximum $r = 1$ if the nearest sample $\boldsymbol{x}_i$ belongs to a KUC. Thus, only known classes (KCs) are used for label prediction, while both KCs and KUCs determine distance ratios. The strategies primarily vary in the confidence computation, leading to nearly identical results.

Deep Neural Network (DNN). Given the vast possibilities of using KUCs in training DNNs by tailoring losses, we opt for the classical cross-entropy loss. The feature extractors mentioned in the dataset-specific paragraphs of the main manuscript are used. A single fully-connected layer with softmax activation is attached and finetuned.

For SPL, we expand the number of output units by one class. We do not evaluate MPL as it does not scale to large KUCs sets. For KvR, we note that this strategy is equivalent to the entropic open-set loss [1]. The objective of entropic open-set is to predict a uniform distribution of unknowns, while KCs are learned according to a cross-entropy loss.

Extreme Value Machines (EVMs). The EVM [5] estimates a Weibull distribution for each sample, considering distances to the nearest samples from other classes. It implements an One *vs*. Rest (OvR) scheme as it uses distances to *rest*-class samples for each *one*-class sample. The τ smallest distances termed *tail* are used to estimate Weibull distributions.

SPL and MPL use 1 or $|\mathcal{T}_u|$ pseudo-classes, respectively. Unlike SPL, MPL allows KUCs in the tail of other KUCs, causing regularization among close KUCs. This effect is visible when comparing SPL and MPL in Tab. 2, where the space between KCs and densely populated KUC areas is prioritized for KCs. KvR uses KUCs only in the tails of KCs, and KUCs never act as one-class themselves, leading to gentle transitions between decision boundaries.

We also explore the C-EVM [2], which employs DBSCAN clustering on a per-class basis before the EVM fitting. Clustering calculates centroids for each cluster, serving as proxies for all samples within the cluster. EVM fitting is exclusively performed on these centroids. This model-agnostic preprocessing technique can be applied to any other method mentioned in this work. One particular aspect of interest is whether clustering can counteract the label noise of the MPL strategy. For SPL and MPL, clustering is applied to the entire KUCs as a unified class. For MPL, the resulting class centroids are treated as independent classes again, replacing redundant KUCs and avoiding label noise. For KvR learning, only the KCs undergo cluster-based reduction while leaving the KUCs unaffected.

Support Vector Machines (SVMs). We deploy the training strategies to two SVM variants. The Weibull SVM (W-SVM) [6] combines a one-class SVM and a binary OvR SVM for each class. Weibull distributions are estimated from both SVMs and probabilities are determined by these Weibull distributions. The Probability of Inclusion SVM (P_I-SVM) [3] predicts unnormalized posterior inclusion probabilities with RBF kernels. Weibull distributions are estimated on samples near decision boundaries.

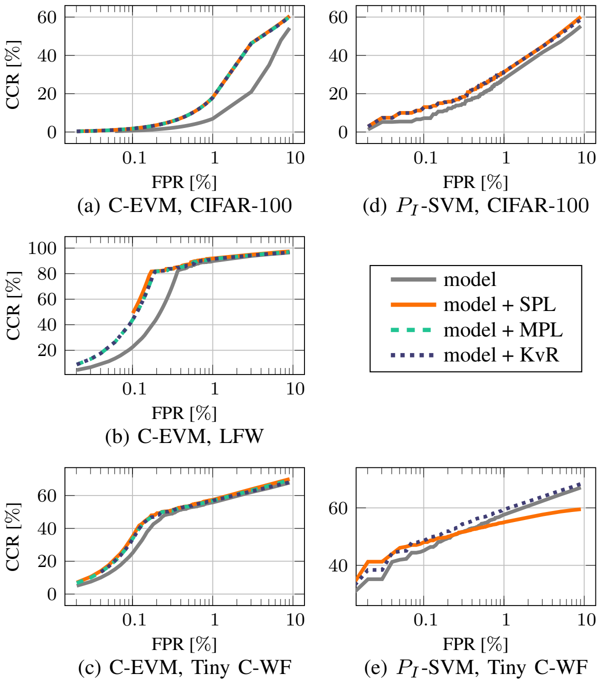
The training strategies SPL and MPL differ only in the number of pseudo-classes. We omit MPL due to its limited scalability to large number of classes. Also, MPL is makes it challenging to serve the SVMs' requirement of at least three samples per class (preferably more). For the same reason, we do not evaluate the SVMs on Labeled Faces in the Wild (LFW). The KvR strategy can be deployed to the SVMs by not representing the KUCs as a positive one-class. Instead, they are always considered part of the rest-class during training.

3. Performance of training strategies – additional results

This section complements the experiments in which genuine KUCs are exploited by the three learning strategies: 1) SPL, 2) MPL, and 3) KvR.

Figures 1a – 1c show the biased Open-Set Classification Rate (OSCR) of the C-EVM within the open-set relevant false positive rate (FPR) range on CIFAR-100, LFW, and Tiny CASIA-WebFace (C-WF). The C-EVM demonstrates similar behavior to the vanilla EVM [5]. All strategies perform comparably well and outperform the baseline. However, it remains inconclusive whether the prior clustering of background data in SPL and MPL prevents potential label noise, leading to improved detection. Two possible conclusions arise: 1) In all three datasets, there is no noise within the genuine KUCs. 2) The C-EVM performs well even without prior clustering of background data, as in KvR.

Figures 1d and 1e show the biased results of the P_I-SVM. For CIFAR-100, consistent improvement over the baseline is evident. However, for Tiny C-WF, SPL at FPRs greater than 1 % results in a degradation of the correct classification rate (CCR). The P_I-SVM appears to encounter challenges in modeling all KUCs within a single class, whereas the W-SVM in the main work performs well. This discrepancy might be attributed to the W-SVM's use of a one-class and a binary SVM for each class, while the P_I-SVM deals solely with a binary SVM.



(a) C-EVM, CIFAR-100

(d) P_I-SVM, CIFAR-100

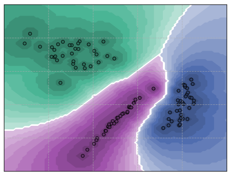
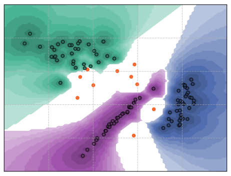
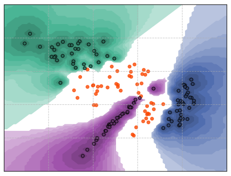
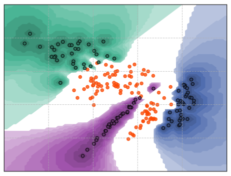
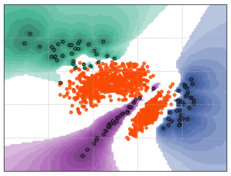
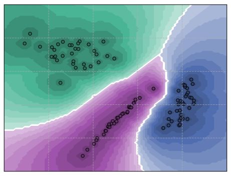
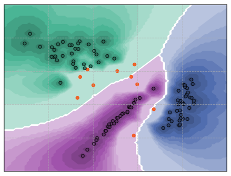
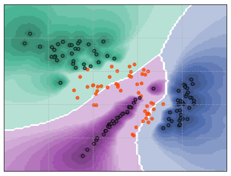
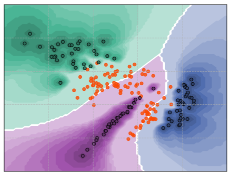
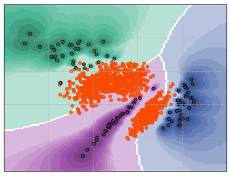
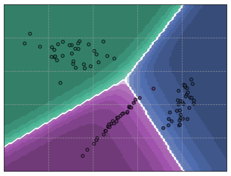
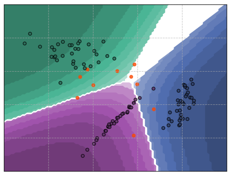
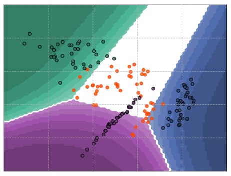
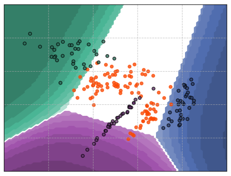
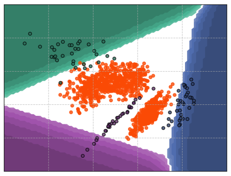
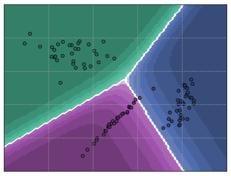
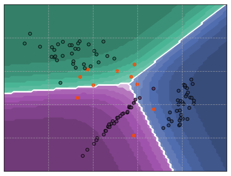
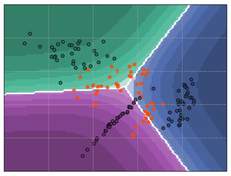
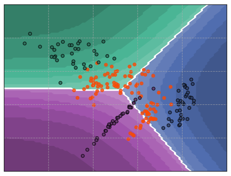
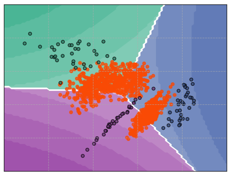
(b) C-EVM, LFW

(c) C-EVM, Tiny C-WF

(e) P_I-SVM, Tiny C-WF

Figure 1. Results of the biased evaluation of 2 models (column-wise) exploiting genuine KUCs with the strategies and the baseline on 3 datasets (row-wise). The models are C-EVM in (a) – (c) and P_I-SVM in (d) and (e). Shown is the biased Open-Set Classification Rate (OSCR) in the open-set relevant FPR range up to 10 %.

Table 1. OSNN (top) and DNN (bottom) class boundaries with applied strategies and a column-wise increase in mixup samples. This toy dataset contains dark-edged dots from 3 known classes (KCs) and orange dots as mixups. Colored areas display class assignment, with opacity indicating confidence, where white is zero confidence or, conversely, high confidence for open space.

Strategy	**# Mixup-to-Known Ratio**				
	0	0.1	0.5	1	10
OSNN (SPL) **&** **OSNN(MPL)**					
OSNN (KvR)					
DNN (SPL)					
DNN (KvR)					

4. Augmenting models by manifold mixup – toy examples and additional results

In this section, we present additional decision boundary illustrations using a toy example and various mixup-to-known ratios. The second paragraph contains additional open-set measures for the experiment involving naïve mixup samples as KUCs surrogates.

Toy example visualizations. Tab. 1 shows the behavior of the OSNN and DNN with the deployed strategies. As observed in the main manuscript, the strategies exhibit minimal differences in OSNN. The model only considers two nearest neighbors, leaving limited scope for variation.

For the DNN, the open space expands as the number of mixups increases, consequently pushing the decision boundaries closer to the known classes. In instances where SPL encounters unfavorably located classes, such as the purple one, the outcome can be very unfavorable. In contrast, the KvR approach is more lenient and consistently offers more flexibility by employing an appropriate threshold.

Tab. 2 shows both EVM variants with the toy example. As the vanilla EVM has already been discussed in the main manuscript, this section displays the remaining mixup-to-known ratios.

Unlike other methods, the training data for the C-EVM is shown here after the cluster-based reduction. As described in Section 2, SPL and MPL treat KUCs as a single class during clustering, potentially reducing label noise in MPL. In KvR, the background data is not reduced. Notably, we observe a concave-like function when combining oversampling mixups with cluster-based reduction. Oversampling generates larger coherent clusters, leading to a decrease in the number of clusters beyond a certain point. Each cluster is reduced to a centroid, resulting in the observation that more mixups lead to fewer mixups in this type of reduction.

Tab. 3 displays the toy example alongside the W-SVM and P_I-SVM. The decision boundaries of both SVMs are generally comparable with only slight variations. In the low-confidence range, there are occasional abrupt changes. However, since the nearly transparent areas correspond to very low confidence values, a suitable threshold would usually consider these areas as open space.

Table 2. EVM (top) and C-EVM (bottom) class boundaries with applied strategies and a column-wise increase in mixup samples. This toy dataset contains dark-edged dots from 3 KCs and orange dots as mixups. Note that for the C-EVM the visible training and mixup dots are the remaining samples *after* the cluster-based reduction. Only in KvR are the mixup samples not reduced. Colored areas display class assignment, with opacity indicating confidence, where white is zero confidence or, conversely, high confidence for open space.

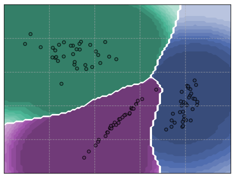
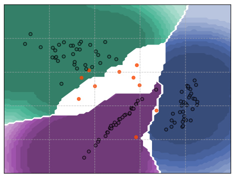
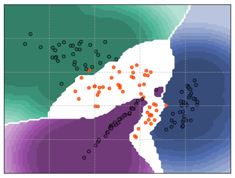
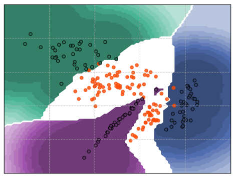
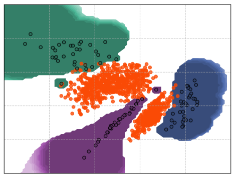
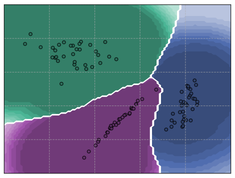
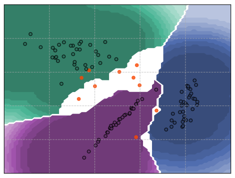
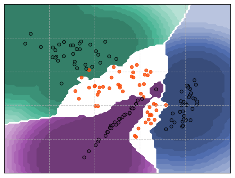
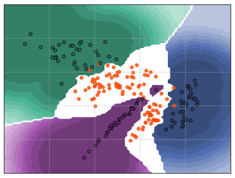
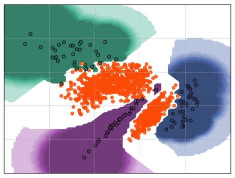
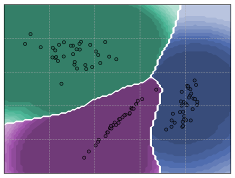
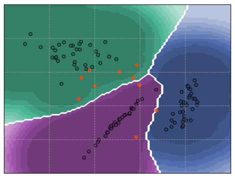
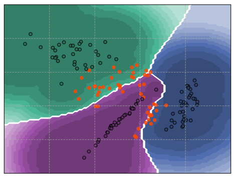
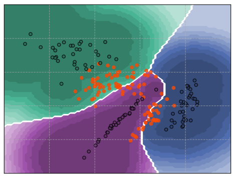
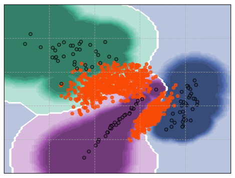
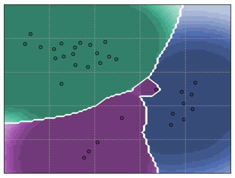
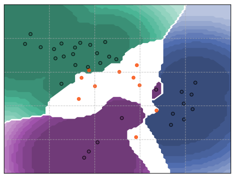
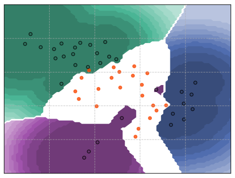
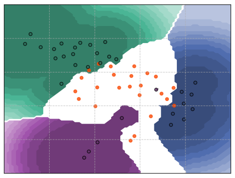
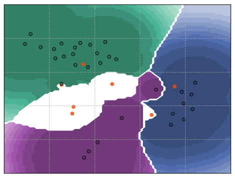
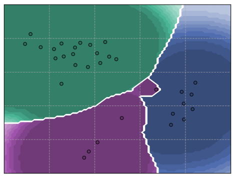
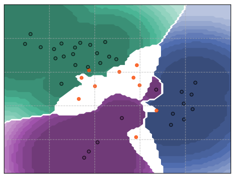
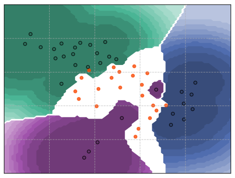
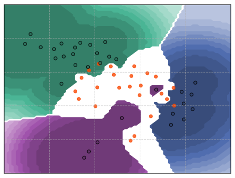
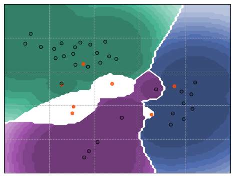
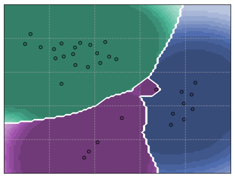
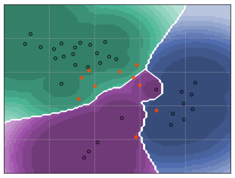
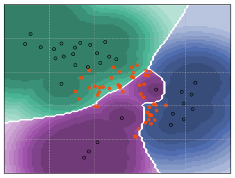
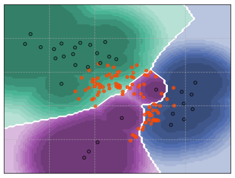
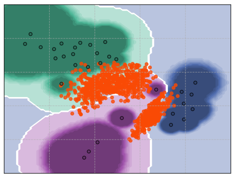
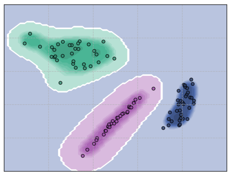
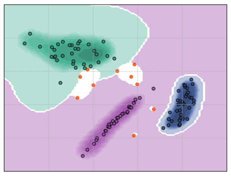
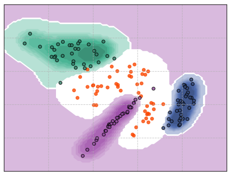
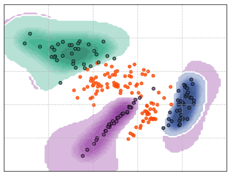
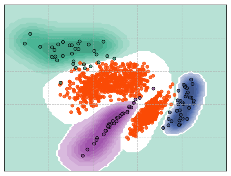
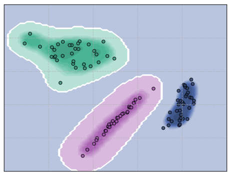
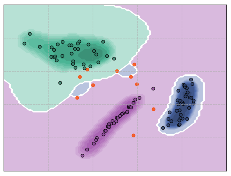
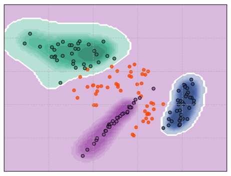
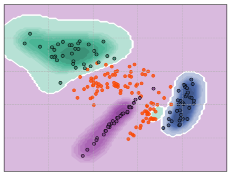
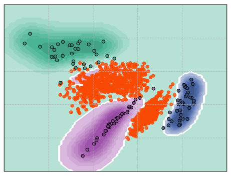
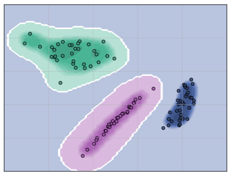
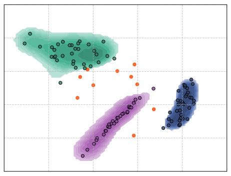
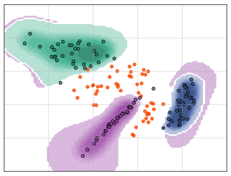
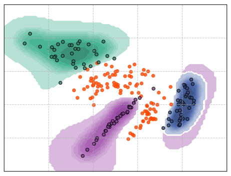
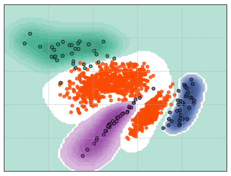
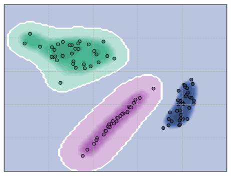
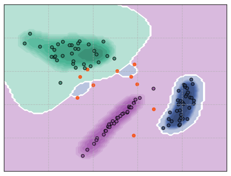
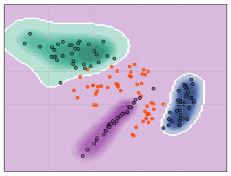
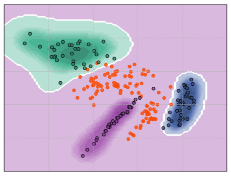
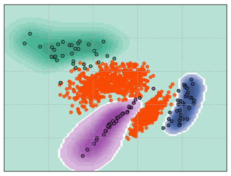
Strategy	**# Mixup-to-Known Ratio**				
	0	0.1	0.5	1	10
EVM (SPL)					
EVM (MPL)					
EVM (KvR)					
C-EVM (SPL)					
C-EVM (MPL)					
C-EVM (KvR)					

Table 3. W-SVM (top) and P_I-SVM (bottom) class boundaries with applied strategies and a column-wise increase in mixup samples. This toy dataset contains dark-edged dots from 3 known classes (KCs) and orange dots as mixups. Colored areas display class assignment, with opacity indicating confidence, where white is zero confidence or, conversely, high confidence for open space.

Strategy	**# Mixup-to-Known Ratio**				
	0	0.1	0.5	1	10
W-SVM (SPL)					
W-SVM (KvR)					
P_I-SVM (SPL)					
P_I-SVM (KvR)					

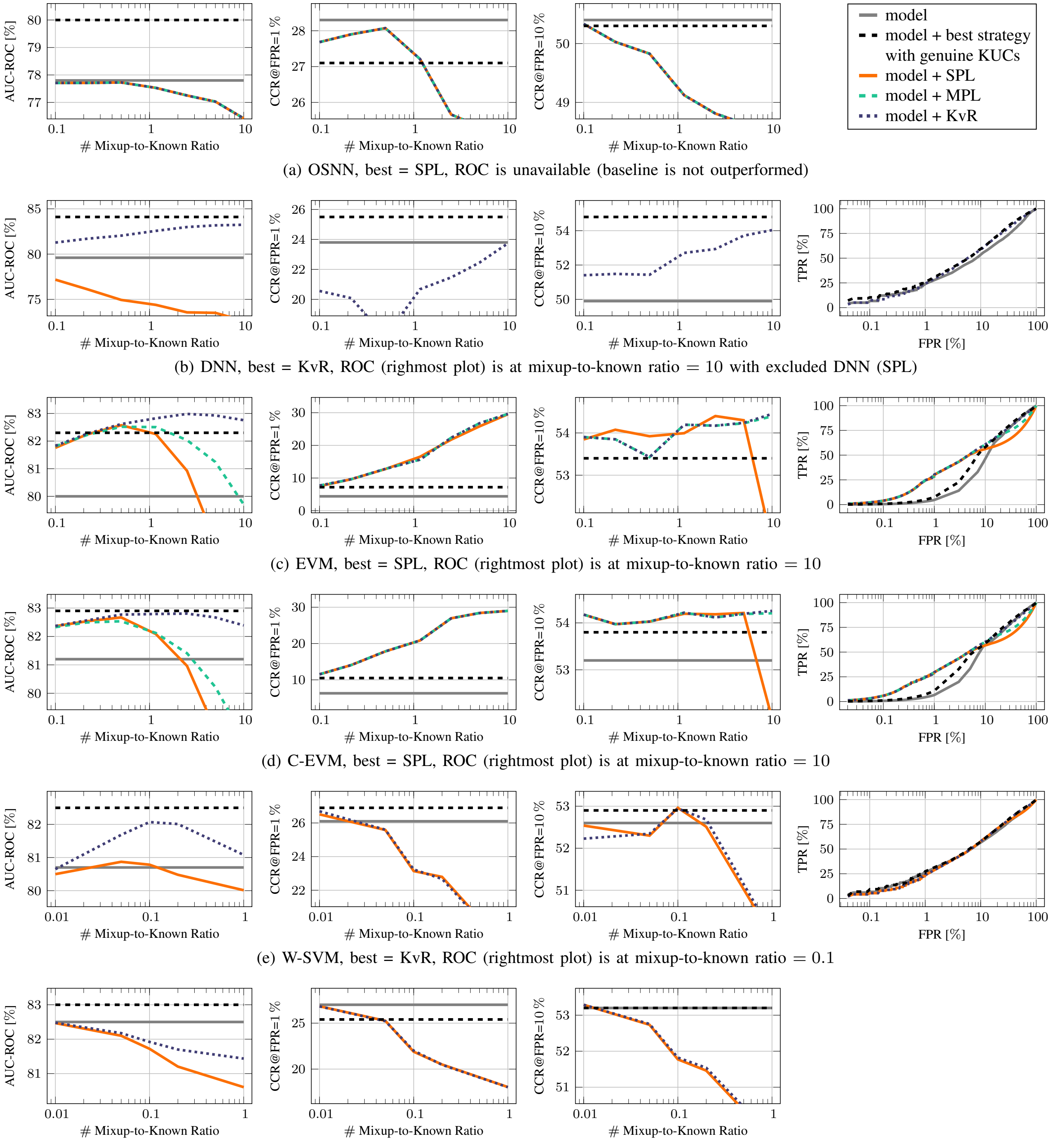
Additional results. This paragraph completes the experiments involving the replacement of genuine KUCs with mixup samples. Figure 2 combines additional open-set measures for the methods discussed in the main manuscript, along with the remaining results for the C-EVM and P_I-SVM. The extended evaluation includes the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), CCR@FPR=1 %, CCR@FPR=10 %, and the Receiver Operating Characteristic (ROC) at a specific mixup-to-known ratio. The latter represents the point of optimal performance while exploiting mixups. The optimal point is indicated in the respective subtitle of each model.

In Fig. 2a, OSNN demonstrates a failure case, exhibiting a degradation in performance across all metrics. However, we demonstrate that resolving the occupation problem enhances OSRs performance for this classifier.

In Fig. 2b, KvR demonstrates an improvement in the AUC-ROC, but this improvement comes partly at the expense of the CCR@FPR=1 %, which decreases to the mixup-to-known ratio of 0.3 and then starts to increases again. In comparison, CCR@FPR=10 % steadily increases and reaches 54 % at a mixup-to-known ratio of 10, while the baseline achieves 50 %. This gain is also evident in the ROC where it extends to very high FPRs. In conclusion, mixup proves to be a suitable approach for improving DNN (KvR) in applications with medium security requirements.

In Figs. 2c and 2d, the EVM and C-EVM outperform the baseline by exploiting mixups. Both variants show similar behavior, with KvR outperforming the other strategies in this experiment. The most significant improvement over the baseline is observed at the CCR@FPR=1 %, reaching over 25 % at the highest evaluated mixup-to-known ratio. This advantage also extends to the CCR@FPR=10 %, except for SPL, which experiences a sharp drop with many mixups, as previously observed in the AUC-ROC. The ROC shows that for SPL and MPL, the true positive rate (TPR) deteriorates at FPRs greater than 10 %. This suggests that mixups have no positive effect on the closed-set case. However, there is a tremendous gain at FPRs between 0.1 to 10 %.

While the W-SVM in Fig. 2e temporarily outperforms the baseline, the P_I-SVM in Fig. 2f does not benefit from the mixup samples. It is noteworthy that training with genuine KUCs also did not results in any enhancement. A reasonable attempt at improvement would involve conducting a hyperparameter search for training with KUCs as well.



(a) OSNN, best = SPL, ROC is unavailable (baseline is not outperformed)

(b) DNN, best = KvR, ROC (righmost plot) is at mixup-to-known ratio = 10 with excluded DNN (SPL)

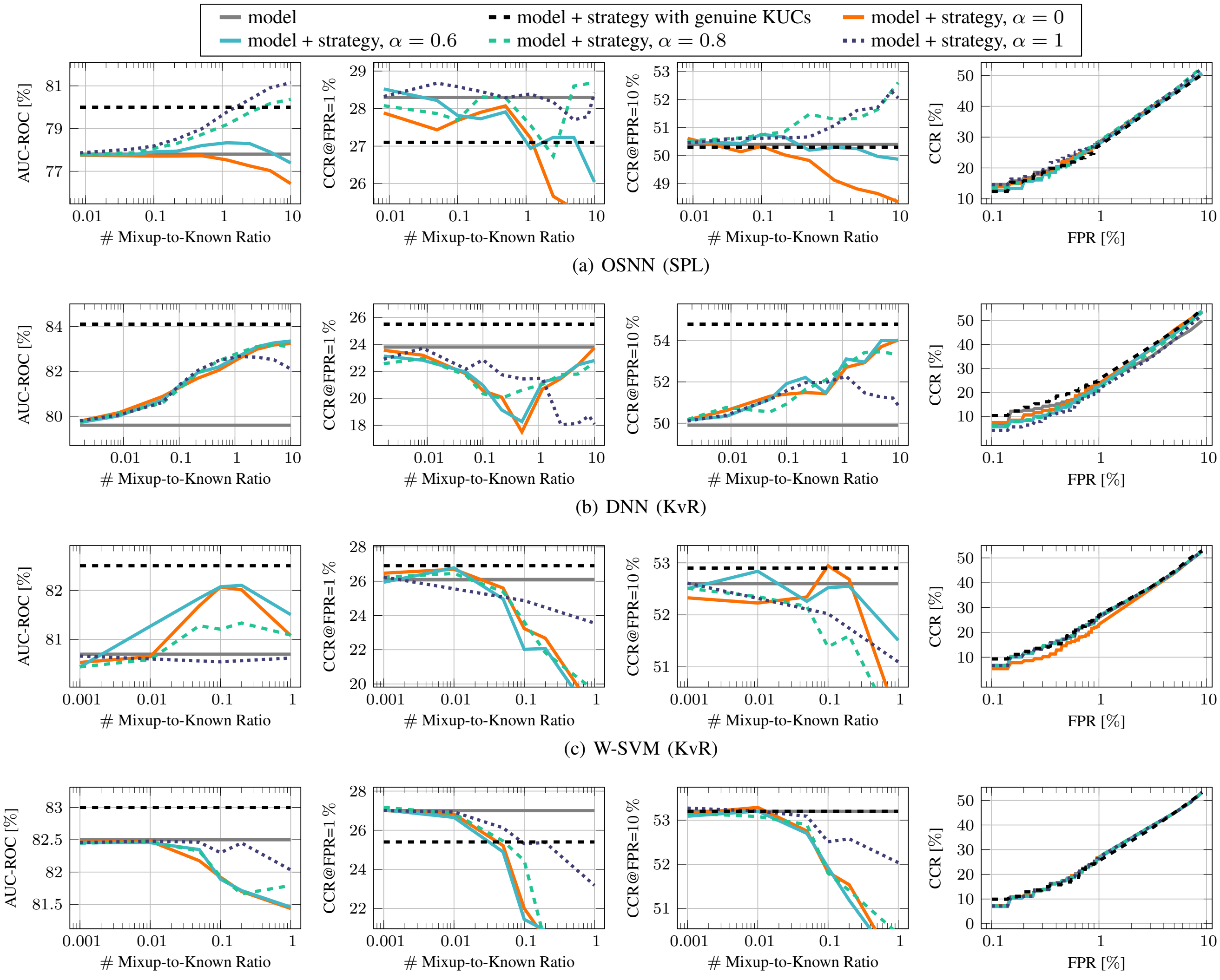
(c) EVM, best = SPL, ROC (rightmost plot) is at mixup-to-known ratio = 10

(d) C-EVM, best = SPL, ROC (rightmost plot) is at mixup-to-known ratio = 10

(e) W-SVM, best = KvR, ROC (rightmost plot) is at mixup-to-known ratio = 0.1

(f) P_I-SVM, best = KvR, ROC is unavailable (baseline is not outperformed)

Figure 2. Unbiased results of all models (row-wise) trained with the different strategies and mixup samples on CIFAR-100. The metrics are (left to right): AUC-ROC, CCR@FPR=1 %, CCR@FPR=10 %, and the ROC. The first 3 metrics are shown w. r. t. the mixup-to-known ratio. The ROC is depicted at a specific mixup-to-known ratio. The baseline model without KUCs (—) and the best strategy of each model trained with genuine KUCs (- -) from the first unbiased experiment, *cf*. Fig. 4 in the main work, serve as reference. This best strategy and the mixup-to-known ratio of the ROC are indicated in the respective subtitles.



(a) OSNN (SPL)

(b) DNN (KvR)

(c) W-SVM (KvR)

(d) P_I-SVM (KvR)

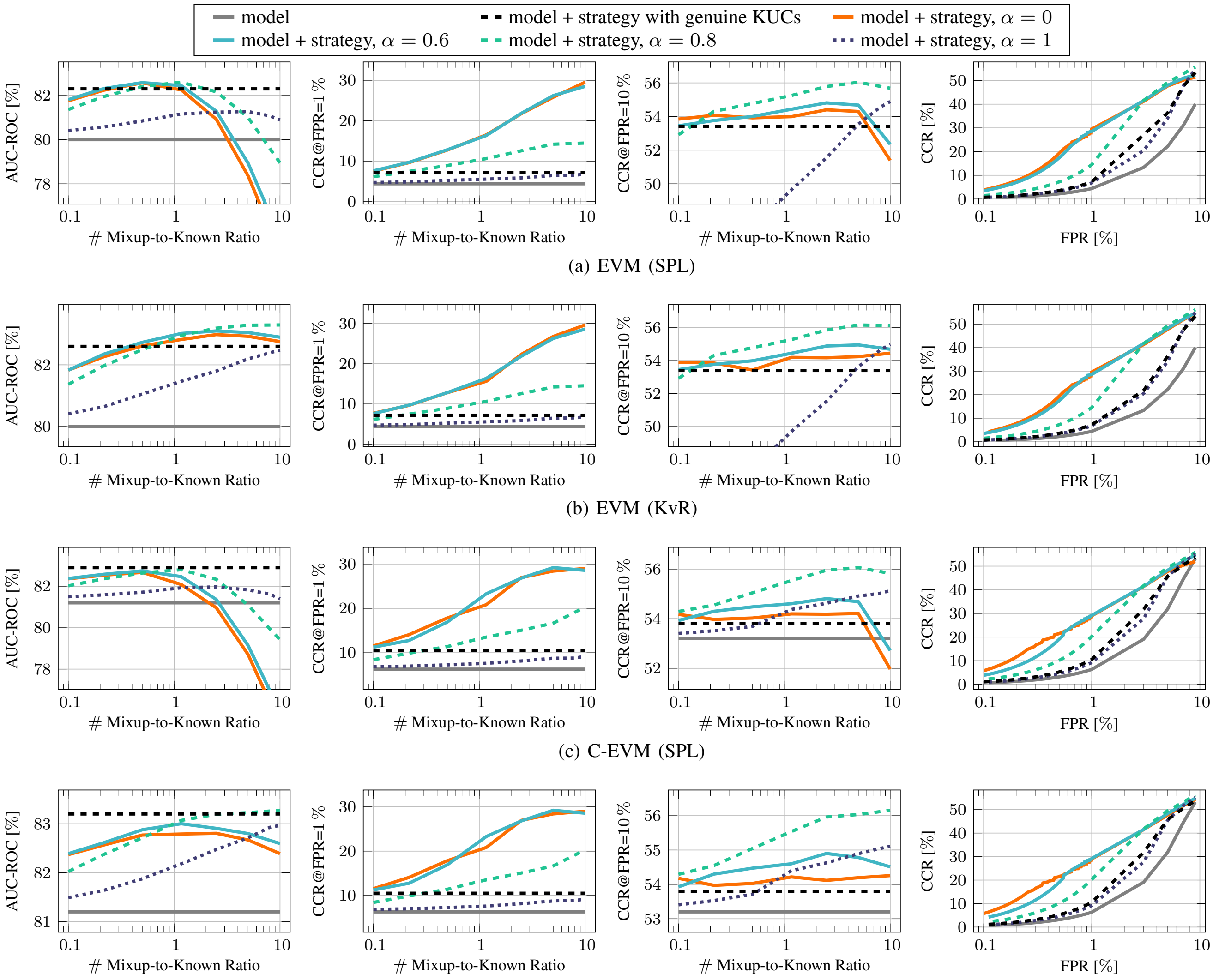
Figure 3. Unbiased results of 4 models (row-wise) exploiting constrained mixups on CIFAR-100. The metrics are (left to right): AUC-ROC, CCR@FPR=1 %, CCR@FPR=10 %, and the OSCR at a certain mixup-to-known ratio. The OSCR corresponds to the maximum of each curve in the CCR@FPR=10 %. For example, in (a), for $\alpha = 0.8$ it is the mixup-to-known ratio of 10 and for $\alpha = 1$ the ratio of 8.

5. Solving the occupation problem – additional results

This section contains additional results of the assessment with constrained mixups to solve the occupation problem. While the main manuscript focuses on the AUC-ROC, here we provide the other open-set measures as well. Figure 3 displays the results for the OSNN (SPL), DNN (KvR), W-SVM (KvR), and P_I-SVM. Figure 4 contains the results for the EVM and C-EVM, both with SPL and KvR.

In general, the OSNN in Fig. 3a benefits from more constrained mixups. While the CCR@FPR=1 % varies unstably, the CCR@FPR=10 % shows improvement. In contrast, the DNN (KvR) in Fig. 3b does not show any improvements with constrained mixups. Stronger constraints can reduce the performance drop in the CCR@FPR=1 % by 3 %, but with $\alpha = 1$ it appears merely shifted. Both SVM variants in Figs. 3c and 3d marginally benefit from stronger constraints. Their overall downward trend is reduced with $\alpha = 1$ and could potentially be further improved with even stronger constraints. However, based on the current results, the exploitation of mixup, or KUCs in general, in combination with SVMs is limited for the detection of unknown unknown classes (UUCs).

The EVM variants in Fig. 4 exhibit minimal variation. Lower constraints enhance the CCR@FPR=1 % while a constraint with $\alpha = 0.8$ promotes the CCR@FPR=10 %. This difference in behavior can be leveraged when considering different safety requirements focused on different FPRs.



(a) EVM (SPL)

(b) EVM (KvR)

(c) C-EVM (SPL)

(d) C-EVM (KvR)

Figure 4. Unbiased results of the EVM variants with SPL and KvR (row-wise) exploiting constrained mixups on CIFAR-100. The metrics are (left to right): AUC-ROC, CCR@FPR=1 %, CCR@FPR=10 %, and the OSCR at a specific mixup-to-known ratio. The OSCR corresponds always to the maximum of each curve at the CCR@FPR=1 %.

References

[1] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing Network Agnostophobia. *Advances in Neural Information Processing Systems (NIPS)*, 31, 2018. 1

[2] James Henrydoss, Steve Cruz, Chunchun Li, Manuel Günther, and Terrance E. Boult. Enhancing Open-Set Recognition Using Clustering-Based Extreme Value Machine (C-EVM). In *International Conference on Big Data (BigData)*, pages 441–448. IEEE, 2020. 1

[3] Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. Multi-Class Open Set Recognition Using Probability of Inclusion. In *European Conference on Computer Vision (ECCV)*, pages 393–409. Springer, 2014. 1, 2

[4] Pedro R. Mendes Júnior, Roberto M. De Souza, Rafael de O. Werneck, Bernardo V. Stein, Daniel V. Pazinato, Waldir R. de Almeida, Otávio A. B. Penatti, Ricardo da S. Torres, and Anderson Rocha. Nearest Neighbors Distance Ratio Open-Set Classifier. *Springer Machine Learning (ML)*, 106(3):359–386, 2017. 1

[5] Ethan M. Rudd, Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. The Extreme Value Machine. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 40(3):762–768, 2017. 1, 2

[6] Walter J. Scheirer, Lalit P. Jain, and Terrance E. Boult. Probability Models for Open Set Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(11):2317–2324, 2014. 2