[To Reviewer #1] Thanks for your appreciation of this work! We address your concern as follows.

Q1:"...include data sets from corresponding papers (eg 20newsgroup for LACU-SVM)" During rebuttal period, we compared EULAC and LACU-SVM on 20newsgroup for 100 repeated experiments, and their win/tie/loss results are: 44/30/26 for linear kernel and 100/0/0 for Gaussian kernel. This clearly shows the advantage of our method. Note that parameter tuning has been conducted for LACU-SVM, please refer to the response to Q4 of Reviewer #3 if interested.

[To Reviewer #2] Thanks for your detailed review and insightful comments. We answer your main questions as follows.

Q1:"Do we need to commit ourselves to the OVR loss?...considering a loss function such as softmax cross entropy loss apart from the convex formulation..." You are absolutely correct! Proposition 1 (unbiased property) can be generalized to arbitrary multiclass surrogate loss $\Phi(\boldsymbol{f}(\mathbf{x}), y)$. The derivation is straightforward by the condition $(1-\theta) \cdot P_{new} = P_{te} - \theta \cdot P_{tr}$ and we have unbiased risk estimator $R_{LAC} = \theta \cdot \mathbb{E}_{(\mathbf{x},y)\sim P_{tr}}[\Phi(\boldsymbol{f}(\mathbf{x}), y) - \Phi(\boldsymbol{f}(\mathbf{x}), \mathsf{nc})] + \mathbb{E}_{\mathbf{x}\sim P_{te}^X}[\Phi(\boldsymbol{f}(\mathbf{x}, \mathsf{nc})]$. The reason for choosing OVR loss in our paper is to obtain a convex formulation and thus easy to optimize. If convexity is not required (e.g., NN implementation), we can use more flexible multiclass loss and binary loss with above R_{LAC}, so the mentioned cross entropy loss is also applicable. We will make this clear in the revision.

Q2:"How to use the non-negative risk estimator in this problem?" The max operator can be added for each binary classifier to avoid the negative loss. We will add more elaborations about the formulation in the revision.

Q3:"My question is have you tried different loss functions?" Yes, we initially use softmax cross entropy loss for deep models. However, it does not converge in experiments. So we instead use sigmoid loss following Kiryo et al. [24].

Q4:The issue about Theorem 1. Theorem 1 serves as a guide to choose binary loss for OVR scheme. Proposition 1 ($R_{LAC} = R_{te}$) only ensures the unbiasedness for surrogate loss, which does not mean we can optimize R_{LAC} over surrogate loss to obtain a model performing well on 0-1 loss. Thus, a consistency guarantee (Theorem 1) is necessary.

[To Reviewer #3] Thanks for the detailed review and helpful comments. We address your main concerns as follows. For the other minor issues, we will discuss in the paper and revise the paper according to your suggestions.

Q1:About the terminology of LAC and OSR. We would like to revise the terminology in the revision if it is allowed.

Q2:"Some of the claims made about prior work are not accurate." Thanks, previous works [9,10,12] indeed provide theoretical analysis for OSR. Our primary claim is that we are the first to provide the generalization analysis in this line of research. We will revise the paper and appropriately acknowledge theoretical contributions of previous works.

Q3:"...what point is the unbiased risk estimator established?" We use way (2). In addition to labeled training data, there is a set of unlabeled data from an available source reflecting underlying environments. We can then establish an unbiased risk estimator for the underlying distribution in testing time (where new classes emerge now). Moreover, we note that our method can also be extended to *transductive* setting [aka, way (1)]: only labeled data are available in training time. Then, we can use part of testing samples as unlabeled data to establish the unbiased estimator and employ that to the rest (analog to cross-validation procedure). We will clarify the setting and add the remark in the revision.

Q4:"...there isn't a strong expectation for their default settings to be the best" As argued by [3], LACU-SVM is not sensitive to parameters, so we use default ones. In [15], PAC-iForest is somehow sensitive to core parameter q, and we have tuned it in experiments. During rebuttal period, we performed a case study on mnist and validate their claim (see Figure 1 left). We further conducted parameter tuning thoroughly for LACU-SVM (4 parameters) and PAC-iForest (6 parameters) on *each* dataset (see Figure 1 right). Optimal tuning results are slightly better than the default one, while they are still worse than our EULAC.

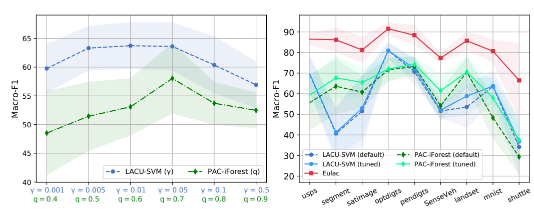

Figure 1: case study (left); parameter study (right).

[To Reviewer #4] Thanks for your review. Below we would like to clarify several serious misunderstandings.

Q1:"the testing distribution were available seems unrealistic." This is a misunderstanding: We do NOT use any testing sample in training time. Instead, we try to approximate the testing distribution by only labeled and unlabeled training data. On the other hand, we are actually studying the *same* setting as previous works [3,15]. It is realistic that in many tasks unlabeled data can be collected for learning with new classes. Examples are the insect recognition [15]: there exist unobserved insects missed by labeled data, and unlabeled data could help to identify them.

Q2:About the novelty. We admit that rewriting the risk estimator has been used in several weakly supervised learning problems, but our paper is the *first* to provide an unbiased risk estimator for learning with augmented classes, which is a unique view in the line of research. By doing so, we can handle even more complex changing environments (like target shift considered in our paper), and we can provide sound and clear theoretical guarantees for the proposed approach.

We hope the reviewer could check comments from other reviews. For example, Reviewer #1 said "The proposed methodology is simple, straightforward and elegant."; Reviewer #2 said "This paper provides theoretical results for learning under this scenario which is quite insightful and quite unique in this line of research."; Reviewer #3 said "The use of unlabeled data for this purpose has not been explored to any great extent in prior work, and this paper provides a convincing path forward." So we believe our work has sufficient novelty and contributions to the community.

Q3:About the testing data. All the methods share the *same* set of testing data in each experiment.