Thanks all reviewers for their thoughtful comments. Below, we conduct a point-point response to each comment.

R2.1 Missing related works: We will discuss them in the revised manuscript. Briefly speaking, [Tsai et al. TMM19] and [Hsu et al. ECCV18] apply the SPL processes to refine the saliency maps obtained in co-saliency detection. [Zhang et al. CVPR20] and [Nguyen et al. NIPS19] are two latest works on weakly and unsupervised saliency detection and we will report their performance in the revised manuscript. [Zhang et al, PAMI20] and [Dizaji et al. CVPR20] integrate SPL and AL for **domain adaption** and **clustering**, respectively. They adopt **heuristically designed** self-paced regularizers to guide the AL processes. In contrast, the self-paced regularizer in our model is **implicitly learned in a data-driven manner** rather than being heuristically designed and it guides an FCL process rather than an AL process.

R2.2 Discussion with [23] and [27]: As we mentioned in the paper, these works share a similar spirit with our work. As for the difference, Yan et al. [23] define their problem on sparsely annotated video frames. The learning process is based on the dependencies among adjacent video frames. In contrast, our task does not have such dependency to explore. [27] presents an IoT-oriented saliency learning framework. The intention is to leverage both labeled and unlabeled data from different problem domains. It requires to generate image data for different problem domains. In contrast, our task works under a single and common problem domain, thus needing no such data generation process.

R2.3 Relation between semi-supervised segmentation (SSS) and FCSOD: Besides the superficial difference in the number of classes, the challenges met by them are also different. Specifically, as the salient class that needs to be segmented in FCSOD would usually cover a number of different semantics rather than forming by a specific semantic, FCSOD is encountered with heavier **intra-class variance** than SSS. This would bring the challenging learning ambiguity issue in FCSOD. Besides, the current SSS methods usually leverage the semantic vector as an informative attribute to guide the GAN-based semi-supervised learning process, which, however, is **absent** from the FCSOD task. Consequently, such SSS algorithms could not be easily applied to FCSOD.

R2.4 Ablation study: We will add more ablation studies into the revised paper. Specifically, our method uses in total 24.5K iterations and the loss and performance curves are shown in Fig. 1. For [3], we will report its results in the revised paper. We have also carried out experiments when varying ratios of labeled data. As reported in Table 1, our approach can learn with different ratios of labeled data robustly.

Table 1. Experiments under different ratios of labeled data.

	1%	5%	10%	30%	Full
maxF	.824	.840	.846	.846	.863
MAE	.054	.049	.045	.044	.044
S	.793	.814	.822	.831	.854

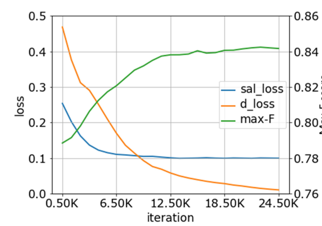

Fig. 1. Loss and performance curves.

R3.1 Contribution and novelty: As recognized by recent works in the SOD community [20-23], one important issue is that the annotation of salient regions is very tedious and training samples with accurate annotations remain scarce and expensive. To solve this problem, we study an under-explored yet meaningful learning scenario and propose a novel learning framework. Thus, we do believe that we contribute to the saliency detection community, as evidenced by the statements of the other two reviewers, e.g., “valuable research direction to the field of SOD", “worth being explored", etc. Besides, we kindly argue that we build a novel APL framework rather than “applying a well-explored gradually learning approach". Featured by the capacity to infer learning pace under a data-driven adversarial learning manner, the proposed APL is beyond the exploration of any existing work.

R3.2 Argument of FCL against SSL and WSL: In the context of SOD, we usually treat each image as a sample. As the learning problem that uses a few samples to train a target model is named as few-shot learning, we name our problem as FCL as it requires annotation cost only on **a few** samples. Different from FCSOD, SSSOD provides **incomplete annotation for every sample**, while WSSOD provides **weak (image-level) annotation for every sample**. Under this circumstance, we think our argument of FCL against SSL and WSL is correct.

R3.3 Cost study: We kindly remind the reviewer that we have actually provided the number of labeled images in L16&245, which is 1K. Compared to the traditional SOD methods that train on 10K labeled images, our approach only requires 10% annotation cost, thus called few-cost SOD. In addition, we have conducted more experiments to study the performance of our approach under various annotation costs as reported in Table 1.

R3.4 More fit to vision conferences: Our core innovation is the novel adversarial-paced learning scheme, which we believe fits NeurIPS well. Besides, we kindly remind the reviewer that many vision task-related works have been published in NeurIPS, such as the famous Faster R-CNN [Ren et al. NeurIPS15] and Object bank [Li et al. NeurIPS10], and the recent SOD methods [Nguyen et al. NeurIPS19] and [Zhang et al. NeurIPS19].

R4.1 Upper bound discussion: This is an interesting point. To our best knowledge, the upper bound of FCSOD should be higher than unsupervised SOD as unsupervised SOD does not use any human annotation. With the same label cost, the upper bound of FCSOD and WSSOD should be close. However, as FCSOD leverages a small number of strong annotations while WSSOD leverages a lot of weak annotations, FCSOD should theoretically work better when dealing with data with small domain shifts.

R4.2 Fully supervised performance of the used saliency model: Please see the last column of Table 1.

R4.3 Whether SSS methods can be directly adopted to FCSOD: As we discussed in **R2.3**, they cannot.